



Search



Home



Library



Our Music AI

Lyrics generation with lyrics conditioned melody

제 15회 투빅스 컨퍼런스

강효은, 고나경, 국주현, 김상윤, 김성훈, 이예진, 이정우, 지윤혁, 현승현, 홍종현





Table of contents



01

Motivation



Help songwriting

03

Data
Preprocessing



MIDI & Lyrics

05

Result &
Limits



Original &
Generated lyrics

02

Data



Explain MIDI data

04

Model



SeqGAN

06

Thanks



Team
member



Search



Home



Library



01



Motivation

Help songwriting



Motivation



Original songwriting



작사가의 경험에 의존해야 함
본인이 모르는 표현이나 단어 등은
쓰일 수 없음



Model based songwriting



새로운 표현, 단어가
사용된 가사를 참고해
다채로운 작사 가능!



Search



Home



Library



02



Data

MP3-data & MIDI-data & Lyrics-data



Data



MP3

1867개

- 90년대 ~ 현재
TOP 인기곡



Lyrics

1867개

- 음원사이트에서
가사 크롤링

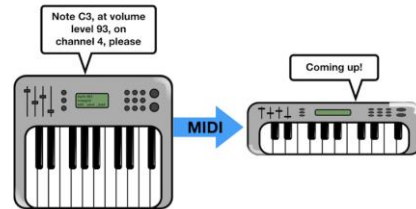


MIDI

1867개

- 소리의 파형을 녹음한 MP3와 달리,
MIDI는 전자 악기를 연주하기 위한 악보 그 자체

HOW MIDI WORKS...





Search



Home



Library



03



Data Preprocessing

Preprocess MIDI-data & Lyrics-data



Data Preprocessing



MIDI

- Spleeter 라이브러리를 활용한 MR제거
- Audio-Convert API를 통해 MIDI 변환



Lyrics

- MP3와 동일한 제목의 가사 크롤링
- GoogleTrans API를 통해 외국어 번역



Data Preprocessing – MIDI Data



Voice Seperation

Spleeter

- 가사와 멜로디의 notes의 alignmnet를 맞추기 위해 instruments 제거



- Spleeter Model 활용



Convert to MIDI

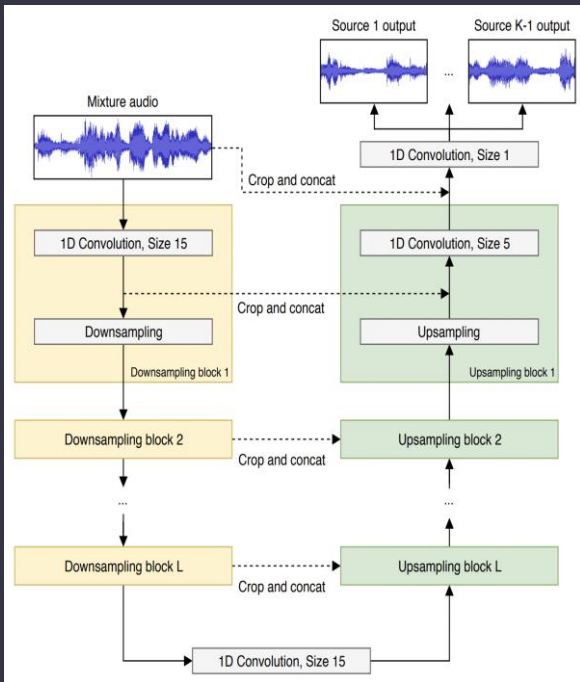
Audio-Convert

- 보컬 사운드만 남은 MP3 음원 파일을 MIDI 파일로 변환

- Audio-Convert API을 활용



Spleeter - Music Source Separation Library



Architecture

Romain Hennequin(2019).

SINGING VOICE SEPARATION WITH DEEP U-NET CONVOLUTIONAL NETWORKS

- Skip connection이 포함된 Encoder - Decoder CNN U-Net (12 layers)
- Encoder는 vocal과 instruments(drums, bass, other) representation 학습
- Decoder는 components를 masking처리하여 waveform 생성

Example

아이유 - Blueming (Original)



아이유 - Blueming (Vocal)





Data Preprocessing - Lyrics Data



Crawling

Selenium



- mp3와 동일한 가사 데이터를 생성하기 위해 Melon, Bugs 등 여러 음악 사이트의 가사 데이터 크롤링 및 수집



Text Preprocessing

Re



- 정규표현식을 통해 특수문자, 숫자, 등 제거하여 전처리 진행



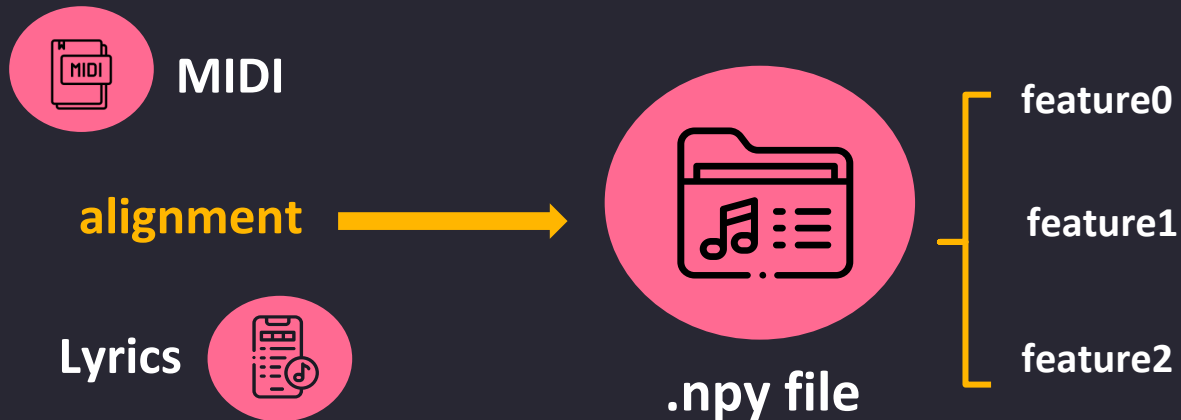
Translation

GoogleTrans
Hangulize



- GoogleTrans를 통해 외국어를 한국어로 번역
- 추가적으로 번역이 안된 영어는 hangulize 모듈을 통해 발음을 한글로 변환

Data Preprocessing - .npy file



* .npy? : python의 numpy 배열을 담은 file

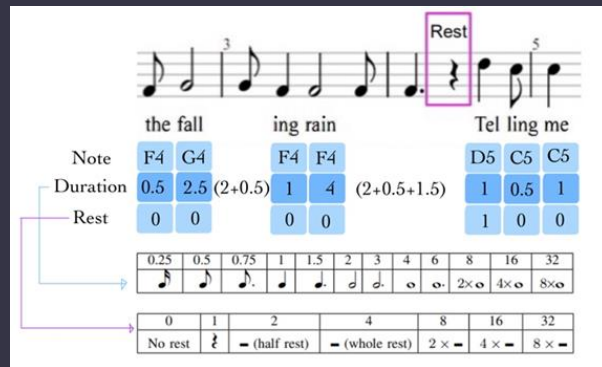
Data Preprocessing - .npy file



feature1 from MIDI

[pitch, duration, rest]

* pitch : 음높이, *duration : 음표의 길이, *rest : 음표가 나오기 전 쉼표의 길이



feature0 from MIDI

[start, length, freq, velocity]

* start : 음표의 시작 시간, *length : 음표의 길이(초단위), *freq : 음표의 주파수, *velocity : 음표를 치는 힘



mido



library for working with MIDI messages and ports
convenient message objects



```
for msg in midi:  
    print(msg)
```



```
note_off channel=0 note=97 velocity=64 time=0  
note_on channel=0 note=58 velocity=70 time=0.011627906976744186  
note_off channel=0 note=59 velocity=64 time=0  
note_off channel=0 note=68 velocity=64 time=0  
note_off channel=0 note=75 velocity=64 time=0  
note_on channel=0 note=77 velocity=37 time=0  
note_off channel=0 note=78 velocity=64 time=0  
note_off channel=0 note=80 velocity=64 time=0  
note_off channel=0 note=93 velocity=64 time=0  
note_on channel=0 note=78 velocity=32 time=0.011627906976744186  
note_on channel=0 note=85 velocity=43 time=0  
note_on channel=0 note=93 velocity=44 time=0  
note_off channel=0 note=96 velocity=64 time=0  
note_off channel=0 note=77 velocity=64 time=0.011627906976744186  
note_on channel=0 note=86 velocity=30 time=0
```



note on / off

channel

note

velocity

time



Data Preprocessing - .npy file



feature0&1 from MIDI

midi file의 모든 note들을 midi token으로 사용하는 것은 불가능

가사 token과의 alignment를 위해 midi 파일을 100 ~ 400 토큰으로 추려내야 함

`note_on & msg.time >= limit`

pick!

`note_on & msg.time < limit`

`note_off`

개별 곡 마다

limit length를 0.03, 0.04, 0.05,

0.07, 0.08, 0.09, 0.1, 0.12, 0.15, 0.2

로 설정 후 추출되는 토큰의 개수가

100~400 인 limit를 이용



Data Preprocessing - .npy file



feature0 from MIDI

start : sum (picked notes' length)

length : msg.time

freq : $400 * 2^{((msg.note-69)/12)}$

velocity : msg.velocity



feature1 from MIDI

pitch : msg.note

duration : 0 ~ 0.1: duration 0.25, 0.5, 0.75

0.1 ~ 0.4 : duration 1, 1.5

0.4 ~ 0.6 : duration 2, 2.5

0.6 ~ 0.8 : duration 3, 3.5

0.8 ~ 1.1 : duration 4, 4.5

rest : sum (unpicked notes' length)

* 현재 note 직전에 선택된 note가 끝난 후,
현재 노트가 시작되기까지의 시간



Data Preprocessing - .npy file



feature2 from lyric

[feature0, feature1과 alignment된 가사]

$\text{♩} = 120$

List en to the rhy thm of the fall ing rain Tel ling me

Lyrics	List	en	to	the	rhy	thm	of	the	fall	ing	rain	Tel	ing	me
Note	D5	C5	C5	A4	A4	G4	G4	F4	G4	F4	F4	D5	C5	C5
Duration	0.5	1	0.5	1	0.5	1	1	0.5	2.5	1	4	1	0.5	1
Rest	0	0	0	0	0	0	0	0	0	0	0	1	0	0

-text cleaning: 특수문자 제거

-okt.morphs : 형태소 추출



Data Preprocessing - .npy file



▶ 1. $\text{len}(\text{feature0}) > \text{len}(\text{feature2})$: midi가 가사보다 길 때

$$n = \text{len}(\text{feature0}) - \text{len}(\text{feature2}), K = \text{len}(\text{feature0}) / n$$

[0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]

↘ 0 + 1 인덱스 제거 (l + 1)

[0, , 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]

↘ 0 + 4 인덱스 제거 (l + k)

[0, , 2, 3, 4, , 6, 7, 8, 9, 10, 11, 12, 13]

↘ 4 + 4 인덱스 제거 (l + 2k)

[0, , 2, 3, 4, , 6, 7, 8, , 10, 11, 12, 13]

Ex) $k = 4, n = 3$

▶ 2. $\text{len}(\text{feature0}) < \text{len}(\text{feature2})$: 가사가 midi보다 길 때

$$n = \text{len}(\text{feature2}) - \text{len}(\text{feature1}), K = \text{len}(\text{feature2}) / n$$

lyric: [['l'], ['m'], ['23'], ['난'], ['수수께끼'], ['Question'], ['뭐'], ['게'], ['요'], ['맞혀'], ['봐요']]

step_1: [['lm'], ['23'], ['난'], ['수수께끼'], ['Question'], ['뭐'], ['게'], ['요'], ['맞혀'], ['봐요']]

step_2: [['lm'], ['23'], ['난수수께끼'], ['Question'], ['뭐'], ['게'], ['요'], ['맞혀'], ['봐요']]

step_3: [['lm'], ['23'], ['난수수께끼'], ['Question'], ['뭐게'], ['요'], ['맞혀'], ['봐요']]

step_4: [['lm'], ['23'], ['난수수께끼'], ['Question'], ['뭐게'], ['요'], ['맞혀봐요']]

Ex) $k = 2, n = 4$



Search



Home



Library



04



Model

SeqGAN



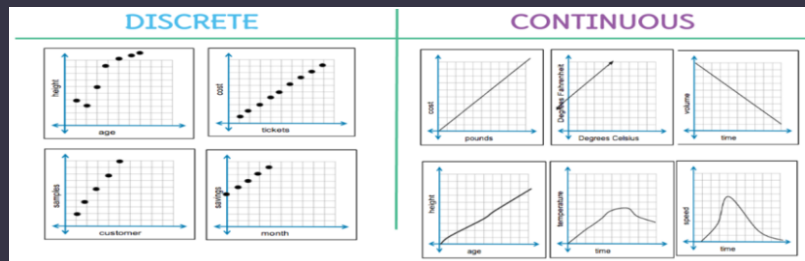
SeqGAN

: 기존 GAN의 한계점

GAN의 경우

Real-Value Continuous Data 생성하지만,

Text Token은 Discrete함



기존 GAN에서

Discriminator 는 문장이 다 생성되고 나서야

score / loss를 매길 수 있음

우리의

네모

칸은

X

우리의

네모

칸은

Bloom

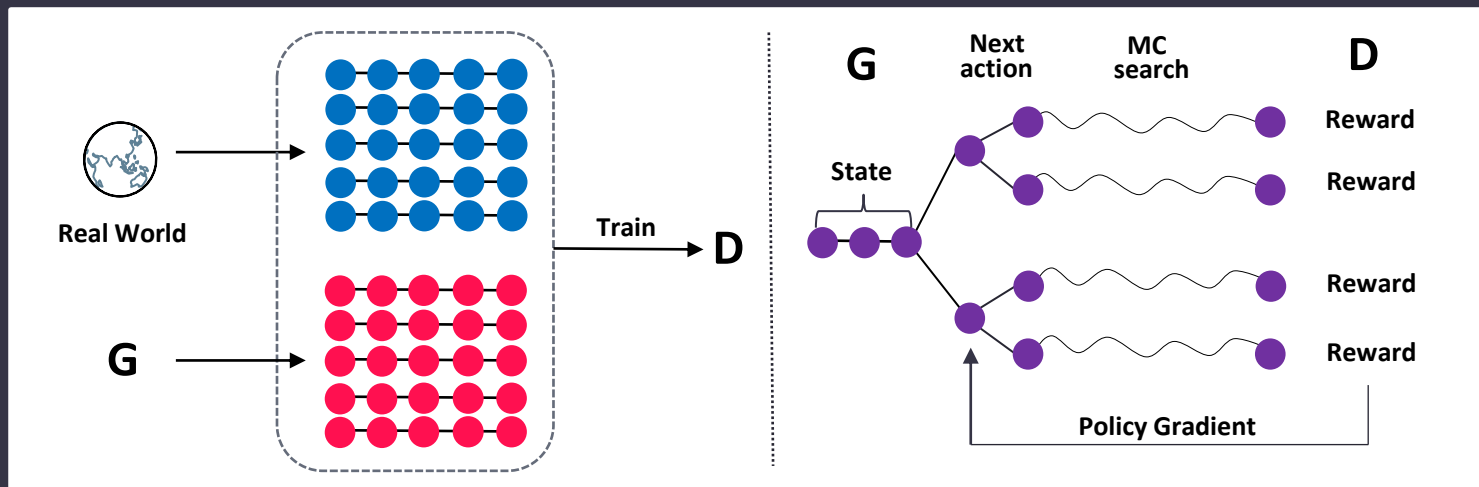
O



SeqGAN

Sequence Generative Adversarial Nets

: discrete한 text data에 적용하기 위해 강화학습을 접목한
text generation GAN





SeqGAN — 몬테카를로 서치 (Monte Carlo search)

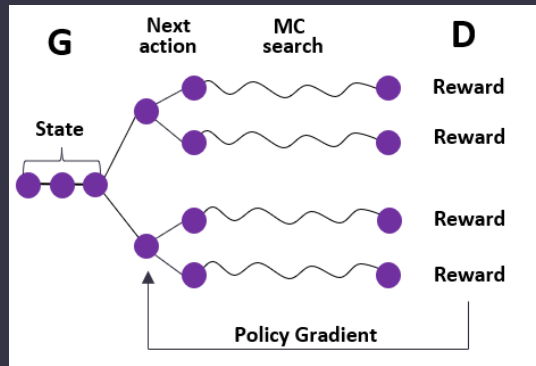


▶ Monte carlo란?

- 강화학습에서 MDP(Markov Decision Process) 를 해결하는 방법의 한 종류(알파고)
 - * MDP: 다음상태는 현재상태와 의사결정자의 행동에만 영향을 받으며 이전의 모든 상태와는 확률적으로 독립적인 확률과정
- 확률 분포를 구하기 어려울 때 무작위 표본 추출을 통해 확률 분포를 도출하는 방식
- 본 프로젝트에서는 'rollout policy' 방법을 사용하여 몬테카를로 서치 진행

ROLLOUT POLICY

16 ROLLOUT_lenth 를 바탕으로 텍스트 토큰 16개에 대해 무작위로 샘플링하여 문장 생성 -> 생성된 문장에 대해 미리 reward를 계산하고 gredient 업그레이드!





SeqGAN — Policy Gradient OBJ FUNC

▶ OBJ FUNC

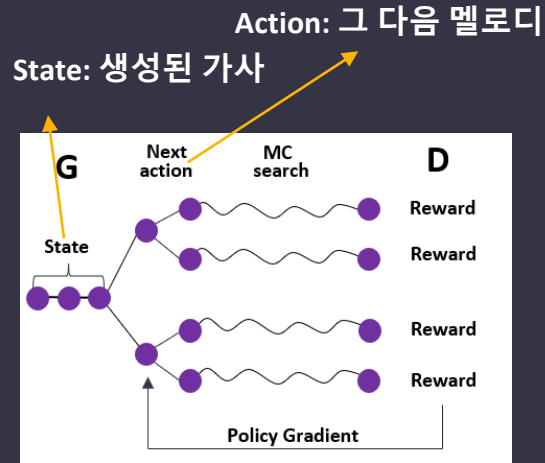
$$\begin{aligned}
 J(\theta) &= \mathbf{E}[R_T | s_0, \theta] \quad \text{State: 생성된 가사} \\
 &= \sum_{m_1 \in \mathbf{M}_{\text{vocab}}} G_{\theta}(m_1 | s_0) \cdot Q_{D_{\phi}}^{G_{\theta}}(s_0, m_1), \\
 &\quad \text{Action: 그 다음 멜로디}
 \end{aligned}$$

G_{θ} : generator;
generative model G는 conditioning information m_i 를 바탕으로 가사 생성

D_{ϕ} : discriminator

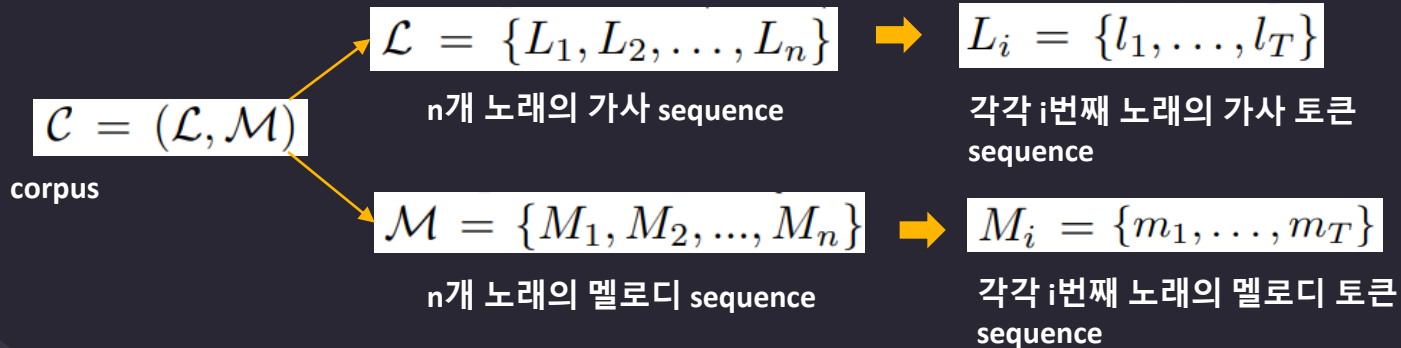
R_T : the reward for a complete sequence

$Q_{D_{\phi}}^{G_{\theta}}(s_0, m_1)$: action-value function;
Start state에서부터 reward의 축적 값; D로부터 발생; 최대가 되도록 학습





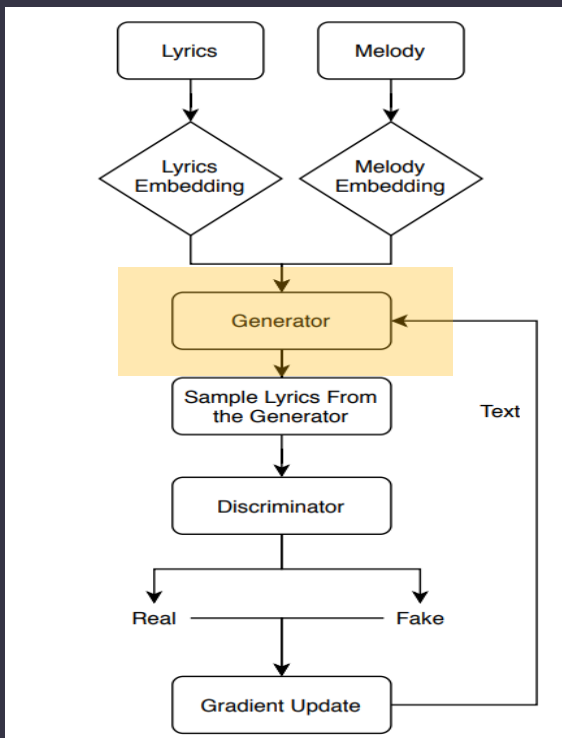
SeqGAN - Input



* 이때, 멜로디 m_i 는 note, duration, rest로 이루어짐

* 멜로디와 가사 각각 토큰화 진행 및 임베딩 → 가사 - note - duration - rest 순서로 CONCAT!

SeqGAN - structure



Generator 구조

- LSTM을 backbone으로 사용
- Embedding 차원 = 128
- Hidden size = 32
- Max Seq Len = 202

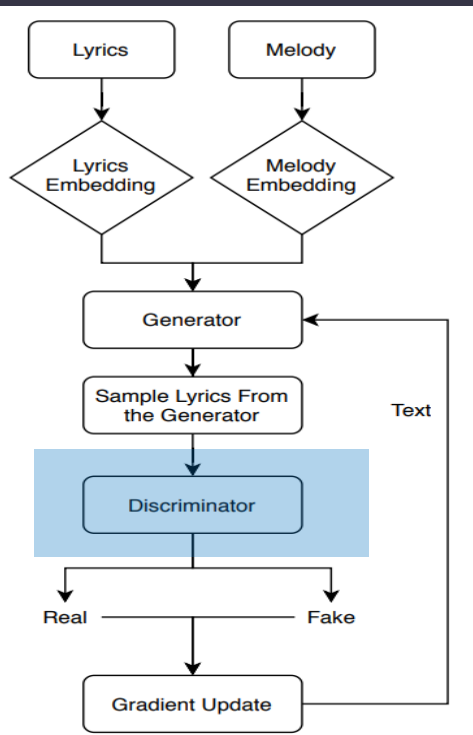
LSTM

Linear

softmax

* Max Seq Len 이 202인 이유
한 곡의 가사, 멜로디의 길이를 200으로 잡고
문장의 첫 시작임을 알려주는 <BOS>(가사의 시작),
<EOS>(가사의 끝) 토큰을 추가했기 때문

SeqGAN - structure



Discriminator 구조

* 층이 많아 사진으로 설명

```
SeqGAN_D(  
  (embeddings): Embedding(10000, 64, padding_idx=23)  
  (convs): ModuleList(  
    (0): Conv2d(1, 100, kernel_size=(1, 64), stride=(1, 1))  
    (1): Conv2d(1, 200, kernel_size=(2, 64), stride=(1, 1))  
    (2): Conv2d(1, 200, kernel_size=(3, 64), stride=(1, 1))  
    (3): Conv2d(1, 200, kernel_size=(4, 64), stride=(1, 1))  
    (4): Conv2d(1, 200, kernel_size=(5, 64), stride=(1, 1))  
    (5): Conv2d(1, 100, kernel_size=(6, 64), stride=(1, 1))  
    (6): Conv2d(1, 100, kernel_size=(7, 64), stride=(1, 1))  
    (7): Conv2d(1, 100, kernel_size=(8, 64), stride=(1, 1))  
    (8): Conv2d(1, 100, kernel_size=(9, 64), stride=(1, 1))  
    (9): Conv2d(1, 100, kernel_size=(10, 64), stride=(1, 1))  
    (10): Conv2d(1, 160, kernel_size=(15, 64), stride=(1, 1))  
    (11): Conv2d(1, 160, kernel_size=(20, 64), stride=(1, 1))  
  )  
  (highway): Linear(in_features=1720, out_features=1720, bias=True)  
  (feature2out): Linear(in_features=1720, out_features=2, bias=True)  
  (dropout): Dropout(p=0.25, inplace=False)  
)
```



Search



Home



Library



05



Result

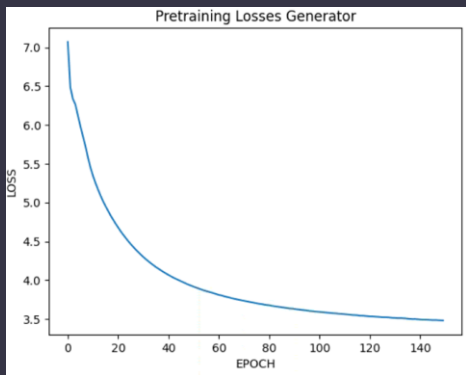
Generated lyrics



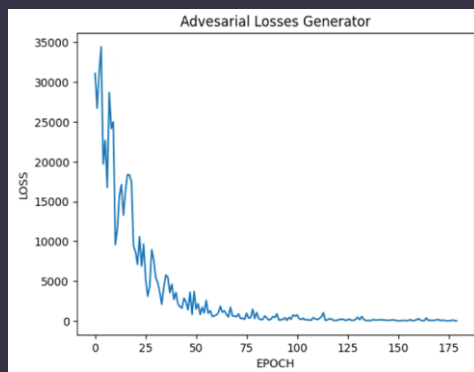
Result Loss - SeqGAN

- LOSS : G는 NLLoss, D는 CrossEntropyLoss 사용

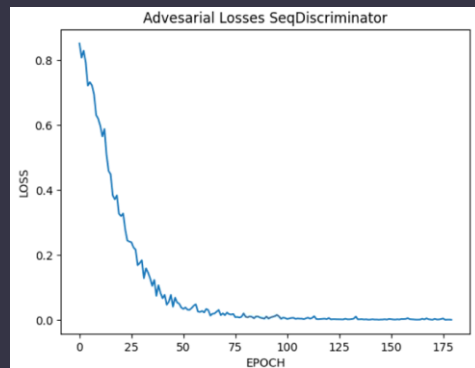
- G는 120, D는 180 에폭 학습 진행



Pretraining G LOSS = 3.4813



G LOSS = 11.155



D LOSS = 0.000638



Result — 가사 비교



그대가 나에게



— 이승철 —

상대를 위해 이별을 결심한 연인과, 그런 연인의 결정을
받아들일 수 없어 슬퍼하는 감정이 잘 드러난다.



Result — 가사 비교



Original lyrics

나의 어깨를 두드리면서 한없이 먼 길을 가라 했지. 그 길은 너무 먼 곳
이기에 멍하니 그대 눈만 보았어. 그대가 나에게 숨겨 왔던 맘 날 위해 떠나
보내리라고 나 몰래 흘려 왔었던 눈물 아직도 그댈 울리고 있어. 그대가
나에게 말해왔던 얘기 내게는 중요하지 않았어. 이렇게 나를 떠나 보내기
위한 애긴 줄 몰랐던 거야. 나의 어깨를 두드리면서 한없이 먼 길을 가라
했지. 그 길은 너무 먼 곳이기에 멍하니 그대 눈만 보았어. 그대가 나에게
말해왔던 얘기 내게는 중요하지 않았어. 이렇게 나를 떠나 보내기 위한
애긴 줄 몰랐던 거야. 그대가 나에게 숨겨 왔던 맘 날 위해 떠나 보내리라고
나 몰래 흘려 왔었던 눈물 아직도 그댈 울리고 있어.



Result — 가사 비교



Generated lyrics

벌써 헤어진 거란 걸 이해를 보내야 했지만 나처럼 미워 해도 어쩔 수 있을까. tv를 하든 천국은 오히려 기억해 나의 사랑 너 마음 슬픔이 난 너 다시는 다시 버린 나의 뜨거운 눈빛으로 장미꽃 여자는 쌓이고 빗속으로 젖은 감정이며 내일 내가 바보 같은 하늘은 이제 나에게 네가 보고 싶은 마음에 때로는 편지를 이렇게 새로 함께 할 순 없겠지. 반지 번개 못 해야만 했나 봐. 내 기다려 가는 내 모습 얼마나 행복한 햇살만이 남아있을 뿐. 오 공동 자금 오늘은 내가 잔뜩 아프구나. 벼랑 진짜 오랜 모습 알고 보통 말해야 해. 그 애기도 보고 싶어 내게 연기 한 걱정 그냥 널 조심 사랑했었던 거야 나는



Result — 가사 비교



Beautiful life



— VOS —

사랑을 할 때 느끼는 설렘, 아름다움 등을 표현하고,
연인을 위해 최선을 다해 사랑할 것을 다짐하고 있다.



Result — 가사 비교



Original lyrics

변해 버린 내 모습에 놀라 잘라 버린 긴 머리까지. 이렇게도 꽤나 내가
괜찮을 줄 나도 정말 몰랐었죠. 더 좋은 사람 돼야죠. 눈부신 그대 나를
보아요. 아름다운 그대 눈빛에 내 전부를 맹세 하죠. Oh I believe 당신을
위해서 나 내가 살아왔던 세상을 모두 바꿀거예요. 이제야 사랑이란 걸
알아가 행복이란 걸 느껴가요. 그만큼 당신의 진심 어린 그 미소에
깨어나는 내 안에 세상. you are my life 밤새 우리에게 비춰준 별들도 이젠
사랑을 나누죠. 행복만 가득한 이 곳에 다시 태어난 거죠. Oh I believe 당신
을 위해서 나 내가 꿈꿔 왔던 세상을 모두 선물할게요. 이제야 사랑이란 걸
알아가. 행복이란 걸 느껴가요. 당신의 진심 어린 그 미소에 깨어나는 내 안
에 숨었던 세상. you are my life you are my soul 내가 살아갈 이유 내 앞에
달라져가는 나의 꿈들이 그대 가슴안에 살아 숨쉬죠. In my heart 마음으로
소원으로 하늘 향해 기도 해요 forever 그대와 나 그대와 나



Result — 가사 비교



Generated lyrics

싸늘한 첫 번째 삶이 되어 못하게 될 거야. 한 걸음 씩 더 가까이 당신의
구원 있었어. 니 감정 없이 너를 사랑하던 상상을 해. 잠든 여자의 남자
보다 이제 우린 신경이 힘이 좋아. 그 노랫말과 똑같다면 그래 더운 다른
느낌 나 나 사랑 있고 있나 차마 기도 그대 만을 우리 설립 하다 나나나 나
사랑 수많은 날에 발이 세상 꼭 멀어질 지라도 사랑해 차갑게 말해야 하던
날이면 내 모든 것이 너무 더 깊어 가는 걸 다 잡는 말 을 깨물고 멀리는
있으면 저 낙엽이 내 눈물로 서로가 꿈꿔 왔던 너를 살다 떨어질 때면 그대
따윈 이성예요. 믿어 괜찮아 난 널 사랑하는데 그럼에도 떠나가지 않고
믿을 수가 없잖아. 나 갈라진 눈물들 그리고 에 대한 나의 숙녀 나만의 기쁨
속에 기억 하다. 나 행복 하나에 예쁜 삶이 하루를 들어 날 우리 둘일까 너를
맘에 잠긴 채 멈춰라. 그저 헛된 서랍 속에 잃어버린 없다고 날 잊어줘 못 다
몰라 없어.



Result — 의의



MP3 → Lyrics (Translation) & MIDI (vocal sound) → .npy

여러 단계를 거쳐 .npy 파일로 변환해 **data set**을 직접 구축



Original Lyrics와 Generated Lyrics 비교하였을 때,

곡의 주제를 비슷하게 잡아내 작사함



Limits



data 수가 부족함

Alignment가 맞지 않을 수도 있으며, 번역 문제가 발생함

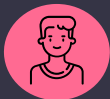
Spleeter 모델로 나온 결과물의 잡음으로 인해, MIDI파일에 노이즈가 있음

완벽하게 song writing을 대체할 수 없고 참고용으로만 사용할 수 있다는 점

시간 부족으로 인해 많은 에폭을 돌리지는 못했다는 점



Team member



김상윤

데이터수집&전처리



김성훈

데이터수집&인퍼런스정리



이정우

데이터수집



지윤혁

데이터수집&인퍼런스취합



홍종현

데이터수집&전처리



강효은

데이터셋수집&전처리&모델인풋생성&발표



고나경

데이터셋수집&전처리&모델인풋생성&발표



국주현

데이터셋수집&모델링&인퍼런스&발표



이예진

데이터파악



현승현

데이터수집&발표자료총괄



Github!



https://github.com/Juhyun730/Melody_Based_LyricGeneration



Search



Home



Library



Our Music AI

Thanks!

