

심층생성모델 기반 가수 스타일 전이형 작사 모델 구현

홍혜진^{*†}, 김소현^{**†}, 이지항^{*,**,+}

^{*}상명대학교 휴먼지능정보공학과

^{**}상명대학교 지능정보공학과

201910844@sangmyung.kr, 202131054@smu.ac.kr, jeehang@smu.ac.kr

Engineering a deep-generative model for lyric writing based upon a style transfer of song writers

Hye-Jin Hong^{*†}, So-Hyeon Kim^{**†}, Jee Hang Lee^{*,**,+}

^{*}Department of Human-Centered AI, Sangmyung University, Seoul, KR

^{**}Department of AI & Informatics, Sangmyung University, Seoul, KR

[†]Corresponding author: Jee Hang Lee, [†]These authors contributed equally

요 약

본 논문은 사전 학습된 심층생성모델을 기반으로 가수 별 가사의 특성을 반영하여 새로운 가사를 생성하는 모델을 소개한다. 베이스 모델로 한국어 사전 학습 모델 KoGPT-2 를 사용하였으며, 총 가수 10 명의 노래 823 곡을 수집하여 미세조정 기법을 바탕으로 학습하였다. 특히, 가수 별로 구분한 가사를 학습 데이터로 구축하여, 가수 별로 독특하게 나타나는 가수 스타일이 전이되도록 하였다. 가수의 이름과 시작 단어를 입력으로 주고 작사를 수행한 실험 결과, (i) 가수 별로 생성되는 가사의 어휘와 스타일이 그 가수의 기존 곡들의 가사와 유사함을 확인하였고, (ii) 작사 결과 가수 별 차이를 확인하였다. 추후 설문을 통해, 개별 가수들의 가사와 생성된 가사의 어휘와 스타일 유사성을 확인하고, 가수 별 차이 또한 확인하고자 한다.

1. 서론

AI 기술의 발전에 따라 대화형 인터페이스를 통해 사용자와 상호작용하는 시스템에 대한 연구가 다양하게 진행되어 왔다. 최근 대화형 인터페이스 기반 연구는 사람과 유사한 성격을 지닐 뿐만 아니라 성별, 나이, 성격 등과 같은 고유한 페르소나를 적극적으로 반영하는 추세를 보이고 있다 [1]. 이러한 접근은 대부분 페르소나에 따라 다른 어투로 사람과 자연스러운 대화를 하는 챗봇 시스템을 구현하는 것을 목표로 한다. 그러나, 자연어로 성격을 구현하는 어려움과, 공적 상황에서 관찰되는 대화에서 페르소나 희소성 문제가 부각되고 있는 바, 이에 대한 연구가 많이 진행되지 않은 상황이다 [2].

위와 같은 한계를 고려하여, 본 연구는 페르소나를 내포하고 있을 만한 노래 가사를 이용, 가수의 작사 스타일이 반영된 가사를 생성하고, 생성 결과가 그 특성을 포함한 가수 스타일인지 확인하고자 한다. 특히, 가수 별로 같은 입력 단어가 주어지더라도 생성된 문장에 어휘나 표현에 차이가 있는지를 우선적으로 비교해보고자 한다. 이는, 페르소나를 직접 디자인하는

대신 이미 우리에게 널리 알려져 있는 가수의 가수 스타일을 페르소나로 설정하여 수행하는 방법으로, 사용자에게 더욱 친밀감이 있는 시스템을 구현하는데 도움이 될 것으로 사료된다.

2. 실험

2.1 베이스 모델

본 논문에서는 KoGPT-2 를 베이스 모델로 선정하였다. 이는 Open-AI 사의 GPT-2 를 약 20GB 의 한국어 데이터셋으로 사전학습한 모델이다 [3].

GPT-2[4]란 Transformer decoder 를 사용하는 self-attention 기반의 딥러닝 언어 모델이다. 기존 GPT 모델[5]과 유사한 구조이나, context size 가 더 크며 약 800 만개의 텍스트 데이터셋을 통해 학습했다는 특징이 있다. 임의의 문장이나 단어를 입력하면 맥락을 고려한 다음 문장을 스스로 만들어내는데 글의 창의성이나 표현 범위를 변수를 통해 조절할 수 있다[4].

GPT-2 는 residual 층의 개수에 따라 4 가지 종류로 나뉘는데 KoGPT-2 는 그 중 가장 작은 사이즈의 모델인 GPT-2 small 을 사용하였다 [3].

<표 1> 입력 단어에 따라 가수 별 생성되는 가사 사례

가수	시작단어 - 나를	시작단어 - 너를
ITZY	나를 외치며 외쳐라 I'mma make it myself 난 멈추는 법을 몰라 I'mma make it myself 난 멈추는 법을 몰라 차갑게 보여도 어떡해 Cool 한 나니까 눈치 볼 마음 없어 Oh Come on 당당하게 Let it go Here we go 길거리를 누비고 On a rollBackground	너를 보고 싶어 나야 나지만 널 원해 원해 원 해 원해 We gon' run this town like we ain' t got no tomorrow20 carat diamond 너 는 확 눈이 멀어 Can' t nobody hold
IU	나를 부를 때 나는 왜 숨을 쉬지 못하고 비 틀거리는 거야 눈물이 멈추질 않아서 잠시만 가만히 있으면 눈을 뜨면 눈을 뜨지 않아 꺼 져 있는 꿈속을 바라보면서 나도 모르게 눈 물이 흘러 나오고 마음을 다 비우면 나 혼자 만의 방으로 들어가서 잠시만 가만히 있으면 눈을 뜨지 않아	너를 부르고 싶었어 그래 불러줘줘 정말 예뻐 지자고 불러줬던 그 목소리 널 불러줘 정말 예 뻐지자고 불러줘 정말 예뻐지자고 불러줬던 그 목소리 잊고 살다 보니까 이렇게 아름다운 날만 있다면 참 좋겠습니다 좋은 꿈 꾸시고 안녕히 계세요
루시드 폴	나를 부르러 찾아왔단다 공공장소에서 흘러 나오는 뜨거운 온기에 몸을 맡긴듯 작은 몸 을 실으리 지금 설 곳이 없어 성큼 걸어가자 쉽 없이 몰아치는 더 큰 불길 세상은 두렵지 않아 우리 함께 뛰어들자 우리 함께 뛰어들 자 우리 함께 가자 손 흔드는 순간 두근두근 대며 나를 보내준 사람 하얗게 마른 땅에 파 묻힌 얼굴들 아무 소리도 나지 않으니 멀리 사라지고 있어 성큼 걸어가자 함께 가자 손 흔드는 순간 두근두근두근대는 우리 둘만의 홀로 아니 그냥 셋이서 함께 가자 우리 함께 걸어가자	너를 보고 있으면 강물이 생각나 강물이 생각 나 상류도 하류도 아니라 아마 중류 어딘가쯤 굽이굽이 허위허위흐르는 강물 강물 당신을 보고 있으면 강물이 생각나 강물이 생각나 느 리지도 빠르지도 않은 때론 맑은 낮으로 때론 슬픈 빛으로 흐르는 강물 갈대도 억새도 모래 도 철새도 조개도 돌게도 물고기도 친구가 되 고 흐르는 강물을 보면 당신이 생각나 당신이 생각나 요란하게 고고하게 그림자 드리운 산 이 아니라 그냥 보름달의 친구 강물 가을도 겨울도 봄도 여름도 구름도 어른도 아이들도 친구가 되고 흐르는 강물을 보면 당신이 생각 나 거칠게 광활하게 넘치고 파도치는 바다가 아니라 그냥 그..
몬스타 엑스	나를 노래해 이 시대 가장 아름다운 목소리 남긴 노래 가사 중 하나가 돈과 사랑에 미쳐 허전해 이런 나를 두고 넌 왜 자꾸 Hope MeSwitch 을 불러 My baby you're so beautiful 왜냐면 널 갖고 싶어 You're so beautiful 왜냐면 It's your beautiful beautiful 왜냐 면 It's your beautiful beautiful 왜냐면 It's your beautiful beautiful 왜냐면 It's your beautiful beautiful	너를 본 그 순간부터 머릿속은 black out 그 어둠 속에서 너 혼자서만 빛나 날 보는 눈빛 은 블랙홀 위험할 것 같아 근데 한 없이 빠져 들어 분홍빛 입술과 pony tail 목소리 다 내 스타일 이미 다른 여자들과는 비교가 안돼 너 가 걸어오면서 뿜어내는 빛이 두 눈을 감아도 선명하게 보여 거짓말처럼 니가 눈에 보여 다 른 일들은 할 수 없어 매일 매일 이제 난 너 만 생각해 하루종일동안 my baby 오직 너만 생각해 다시 만나게 될 날을 기다려 나는 Oh This is
노라조	나를 부르지 말고 나를 비하더라가 아니라 나를 환하게 웃어주더라가 더러워도 숨을 쉴 수조차 없이 아픈데넌 치료약도 발라먹고 치 료효과도 역시 역시 탁월한미양행 구호재단 인 한미글로벌의 사회공헌 활동인 나를 따듯 하게 만들어줘요 미칠듯이 말을 걸어주더라 가 오히려 독하다 나는 오히려 약해 보여도 그게 사랑해줄수 있는 사람이라 나에게말 걸 어줘요 내가 참 멋진 사람이라는 걸 환하게 들려줘요	너를 보내며 세 번의 눈물이 내 가슴에 새겨 진다 내 마음 속 어두운 장막으로 부딪히고 너를 보내며 세 번의 눈물이내 가슴에 새겨진 다 내 마음 속 어두운 장막으로 부딪히고 너 를 보내며세 번의 눈물이 내 가슴에 새겨진다 내 마음 속 어두운장막으로부딪히고 너를 보 내며 세 번의 눈물이내 가슴에 새겨진다 내 마음 속 어두운장막으로부딪히고 너를 보내며 세 번의 눈물이내 가슴에 새겨진다내 마음 속

2.2 가사 스타일 전이를 위한 학습 데이터

본 논문에서는 현재 활동하는 가수 중에서 가사의 어휘나 분위기가 독특해 스타일이 분명하게 드러나는 가수 10 팀을 임의로 선정하였다. 해당 가수들의 전곡 가사를 크롤링하여 데이터셋을 구축하였다. 이 때 가사 구분의 편의와 전처리 과정 용이성 확보를 위해 가수와 제목 컬럼 또한 데이터셋에 추가하였다. 한국어 가사 생성 모델로 한정할 바, 데이터 수집 시 제목 컬럼에 English version 이나 Japanese version 등의 단어가 포함된 곡들은 제거하였다. 또한, 제목에 remix 와 instrumental 단어가 포함된 곡들을 제거함으로써 중복된 곡들이 없도록 하였다. 결과적으로 총 823 곡 데이터를 구축하였으며 각 곡은 평균 87.8 개의 단어로 구성되어 있다. 한 가수 당 평균은 곡 수는 82.3 곡, 표준편차는 40.60 으로 가수들 간 데이터셋의 크기 차이가 있음을 알 수 있다. GPT-2 모델이 end-to-end 모델 기반이며, encoder 를 통해 텍스트 데이터가 처리되는 바, 각 데이터의 별도 레이블링은 수행하지 않았다.

2.3 미세 조정을 통한 모델 학습

앞서 구축한 한국어 노래 가사 데이터셋을 이용, KoGPT-2 모델에 미세 조정을 시행하였다 [6]. 이 때 베이스 모델에서 제공하는 최대 시퀀스의 길이는 기본값 1024 로 설정하였기 때문에, 그 이상의 시퀀스 길이를 가진 데이터는 학습에서 제외되었다. Optimizer 는 Adam 을 사용하였고, learning rate 는 $3e-5$ 로 고정하였으며, 손실 함수로는 CrossEntropy 를 사용하였다.

3. 실험 결과

상기 미세 조정된 심층생성모델 기반 작사 모델에 가사 시작 단어를 입력으로 주면, 해당 단어로 시작하는 가사가 출력된다.

가사 생성 시 하이퍼파라미터로 temperature 를 조절할 수 있다. 이것은 생성된 단어의 뒤에 올 다음 단어의 임의성을 조절하는 역할을 수행한다. temperature 조절을 통해 생성되는 가사의 창의성을 높여줄 수 있지만 temperature 값이 너무 높아지는 경우 가수의 스타일에서 멀어질 수 있기 때문에 경험적으로 설정한 값(여기서는 1.0)을 사용하였다.

또한 생성되는 가사의 스타일이나 문맥에 맞지 않는 단어를 필터링하기 위해 top_k, top_p 변수를 조절하였다. 두 변수의 역할은 다음과 같다 - 높은 확률부터 정렬한 k 개의 단어 중 누적확률이 p 범위에 포함되지 않는 단어는 제외하고 가사를 생성할 수 있도록 진행하였다.

표 1 은 입력으로 해당 단어가 주어졌을 때 가수별

로 생성되는 가사를 나타낸 결과이다. 입력 단어로는 수집한 데이터에서 공통으로 등장한 빈도수가 많았던 단어를 선정하였다.

4. 고찰

실험 결과, 가수별로 같은 입력이 주어져도 다른 가사가 생성된 것을 볼 수 있었다. 원래 가수가 가지고 있는 가사의 장르는 물론 어휘가 비슷하게 생성이 되었는데 이러한 부분에서는 가사 생성이 잘 이루어졌다고 평가할 수 있지만 몇몇 결과의 경우 원래 학습데이터와 동일한 가사가 생성된 부분도 있었다.

학습데이터와 생성된 가사의 스타일 간 유사성을 비교해보기 위해 학습데이터의 단어 중 영어 단어가 차지하는 비율을 측정해보았다. 그 결과 ITZY 52.31%, 몬스타엑스 38.56%, IU 15.43%, 노라조 2.73%, 루시드 폴 1.24%의 순서로 나타났다. 국내 k-pop 아이돌(ITZY, 몬스타엑스)의 가사에는 영어 단어의 비율이 매우 높다. 그에 따라 생성된 가사에서도 전체 단어 중 영어 단어가 차지하는 비율을 측정해본 결과, ITZY 49.23%, 몬스타엑스 27.20%로 원곡에서의 영어의 비중이 결과에서도 비슷한 양상으로 나타난 것을 볼 수 있다. 이에 반해 IU, 루시드폴, 노라조의 가사를 보면 학습데이터에서 단어 빈도를 측정했을 때 영어 단어의 비중이 크지 않았고 생성된 결과도 모두 한글로 이루어진 것으로 보아 유사한 스타일을 갖는 가사가 생성되었다고 볼 수 있다.

생성된 가사를 각각 살펴보면, 기존에 선정한 10 팀 중 가장 적은 데이터셋을 가진 ITZY 의 경우, 첫 문장을 제외하고는 원래 학습 데이터와 동일한 가사가 생성되었다. 데이터가 적어 충분한 학습이 이루어 지지 않고 과적합이 이루어진 것으로 사료된다. 이는 루시드폴과 몬스타엑스의 경우에도 나타났다. 입력 단어가 '나를' 인 경우, 가수에 맞는 새로운 가사가 잘 생성된 것으로 보이지만 입력 단어가 '너를' 인 경우, 기존 곡의 가사와 동일한 가사가 생성되었다.

반면, 가장 많은 데이터셋을 가진 IU 의 경우, 기존 가수의 스타일을 유지하면서도 입력 단어에 따른 새로운 가사가 잘 생성된 것을 볼 수 있었다. 이렇게 충분한 데이터셋을 가지고 있다면 위와 같은 문제는 충분히 극복 가능할 것이라고 사료된다.

노라조의 경우, 학습 데이터와는 다른 새로운 가사가 잘 생성되었지만 몇몇 구간에서 문법적으로 어색하거나 구간의 반복이 많은 가사가 생성되었다. Transformer 에서는 Byte Pair Encoding(BPE)을 통해 문장들을 토큰으로 나누고 이 토큰들은 임베딩 행렬로 변환되어 Decoder Stack 에 들어가게 되는데, Self-attention 과정 이후 들어온 임베딩 벡터를 통해 확률

값을 구한 뒤 이 확률 값이 가장 높은 문장을 출력하는 형태로 진행된다 [7]. 그렇기 때문에 노라조의 경우처럼, 학습 데이터에 단순히 반복되는 구조가 많이 나온다면 결과 생성에 어려움이 생긴다고 볼 수 있다. 이를 해결하기 위해 추후 연구에서는 반복되는 학습 데이터의 전처리 과정을 고려할 계획이다.

현재 결과는 휴리스틱 방법으로 가사의 유사성에 대해 논한 바, 추후 생성한 가사들의 품질을 정량적, 정성적으로 평가할 필요가 있다. 일반적으로, 텍스트 생성 모델의 경우 생성된 문장의 동일 의미 대비 구조적 다양성, 생성문장 품질 수용도(perplexity)[8-10] 등의 지표를 사용한다. 하지만 본 연구의 경우 통상적인 의미/구조적 다양성 등 일반적인 텍스트 평가 지표로 성능을 측정하는 것에는 무리가 있다. 생성된 가사의 라벨에 해당하는 비교 대상이 없어, 문장 구조, 의미 유사도를 측정하는 것이 어렵다. 가사가 시적 예술성을 포함하고 있는 바 일반적 문장 구조 및 내용으로 평가하기에는 한계가 존재한다.

이에 따라, 가사 품질에 대한 사용자 설문 기반 평가를 바탕으로 향후 품질 평가를 진행할 계획이다[11]. 예를 들어, (i) 품질: 가사가 자연스럽게 잘 생성되었는지, (ii) 개인화: 가수와 가사가 어울리는지 등을 제시할 예정이다. 모델의 성능 평가 관점에서, 가수마다 생성된 가사가 일관성이 있는지, 입력 단어에 맞는 가사가 생성되었는지, 특정 가수의 가사 특징을 반영하는지에 대해 설문을 진행하여 성능을 평가할 예정이다 [12]. 추가로, 생성된 가사를 보고 가수를 떠올릴 수 있는 문항을 추가하여 사용자들이 체감할 수 있을 정도의 가수 개성 반영도 또한 확인할 예정이다.

5. 결론

본 논문에서는 페르소나의 한 예로 간주할 수 있는 가수의 가사 스타일을 고려하여, 가사 생성 시 가수별 특성이 발현되는 지 보았다. 실험 결과, 같은 입력 단어에 대해 각각 다른 특징을 갖는 가사를 잘 생성해내는 것을 확인하였다. 다만, 아직 반복되는 문장이 여러 번 등장하거나 문법적으로 어색한 부분을 갖는 한계를 보이기도 했다. 또한, 학습 데이터와 동일한 가사가 출력되는 경우도 존재했는데 이는 데이터셋이 보다 큰 가수에 대해서는 특징을 갖는 가사를 잘 생성해낸 것으로 보아 이 연구가 추후에 가수가 아닌 페르소나로 확장을 하였을 때 페르소나 별로 다른 특징을 갖는 텍스트를 생성해 낼 수 있는 가능성을 보여준 것에 상당히 의미가 있다고 사료된다. 추후 페르소나에 따라 다른 어투로 대화가 가능한 챗봇 시스템으로 확장하여, *verbal communication* 시 사용자의 만족도를 높이고, 감성적 측면에서 더 깊은 사용자 경험을

제공하는 시스템을 제공할 수 있도록 할 예정이다.

Acknowledgement

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2020R1G1A1102683). 본 연구는 삼성미래기술육성센터의 지원을 받아 수행하였음 (No. SRFC-TC1603-52). 본 결과물은 교육부와 한국연구재단의 재원으로 지원을 받아 수행된 사회맞춤형 산학협력 선도대학(LINC+) 육성사업의 연구결과임.

참고문헌

- [1] PRADHAN, Alisha; LAZAR, Amanda. Hey Google, Do You Have a Personality? Designing Personality and Personas for Conversational Agents. In: CUI 2021-3rd Conference on Conversational User Interfaces. 2021. p. 1-4.
- [2] ZHENG, Yinhe, et al. A pre-training based personalized dialogue generation model with persona-sparse data. In: Proceedings of the AAAI Conference on Artificial Intelligence. 2020. p. 9693-9700.
- [3] KoGPT2 모델, <https://github.com/SKT-AI/KoGPT2/tree/589182bc85a0741b8ac20b49cdd56d9e44b9479c>, retrieved 20211007
- [4] Radford, Alec, et al. "Language models are unsupervised multitask learners." OpenAI blog 1.8 (2019): 9.
- [5] RADFORD, Alec, et al. Improving language understanding by generative pre-training. 2018.
- [6] KoGPT-2 미세조정 모델, <https://github.com/forum-ai/KoGPT2-Lyrics-Generation-FineTuning-Version1>, retrieved 20211007
- [7] 안세훈, 정옥란, “감정분석 기반 심리상담 AI 챗봇 시스템에 대한 연구”, 한국 IT 서비스학회지, 제 20 권, 제 3 호, 75-86, 2021.
- [8] Ke Wang, Xiaojun Wan, SentiGAN: Generating Sentimental Texts via Mixture Adversarial Networks IJCAI, 4446-4452, 2018.
- [9] PAPINENI, Kishore, et al. Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 2002. p. 311-318.
- [10] LIN, Chin-Yew. Rouge: A package for automatic evaluation of summaries. In: Text summarization branches out. 2004. p. 74-81.
- [11] 박요한, 정혜지, 강일민, 박천용, 최용석, 이공주, “SeqGAN 모델을 이용한 한국어 시 자동 생성”, 한국정보과학회 언어공학연구회:학술대회논문집(한글 및 한국어 정보처리), 580-583, 2018.
- [12] 이한솔, 이주연, “와파린 복용 환자를 위한 메신저 기반 챗봇 개발 연구”, 서울대학교 대학원, 2020.