

가사 데이터 기반의 작사 지원 시스템 연구

박영재¹, 조희련²

¹중앙대학교 국어국문학과 학부생

²중앙대학교 인문콘텐츠연구소 HK 교수
pyj789@cau.ac.kr, heeryon@cau.ac.kr

A System for Supporting Lyrics Writing Using Lyrics Data

Young-Jae Park¹, Heeryon Cho²

¹Dept. of Korean Language and Literature, Chung-Ang University

²Humanities Research Institute, Chung-Ang University

요 약

본 논문은 과거 한국 가요(K 팝)의 가사를 수집하여 (1) 특정 키워드와 관련된 기존 가사를 검색하거나, (2) 작사가가 작성한 새로운 가사와 유사한 기존 가사를 검색하거나, (3) 특정 키워드와 관련된 가사 속 어휘를 제안하는 작사 지원 시스템을 제안한다. 지금까지의 음악 관련 시스템은 음악을 소비하는 사람들을 위한 음악 추천 시스템에 집중해 왔으나, 이 연구에서는 음악을 생산하는 작사가에게 초점을 맞춰 이들을 돕는 작사 지원 시스템을 제안하고자 한다. 제안 시스템은 TF-IDF와 word2vec을 활용하여 가사와 단어 벡터 공간에 가사와 어휘를 배치하고 코사인 유사도를 계산한다.

1. 서론

지금껏 음악 시스템에 관한 연구는 주로 음악을 듣는 사람이 좋아할 만한 새로운 음악을 추천하는 방법을 연구개발하는 것이 목적이었다. 이러한 음악 추천 시스템의 대표적인 방법으로 협업 필터링(collaborative filtering)이 있는데, 이는 사용자 A와 B가 n개의 항목을 비슷하게 평가한다면 다른 항목에 대해서도 비슷하게 평가할 것이라는 가정에서 출발한다[1]. 또 음악 오디오 데이터로 합성곱 신경망(Convolutional Neural Network)를 학습하여 음악을 추천하는 방법도 제안되었다[2]. 이렇듯 기존 연구가 음악 소비자의 음악 선택을 도와주는 데 초점을 맞추고 있는데 반해, 음악을 생산하는 작사가를 도와주는 시스템은 아직 많지 않다. 한편, 가사 데이터로 곡의 장르를 예측하는 연구가 있지만[3], 가사 작성에 도움을 주는 정보를 제안하기 보다는 가사가 속한 장르를 예측하는데 초점을 맞추고 있다. 본 연구에서는 기존 연구와는 다르게 한국 가요(K 팝) 가사 데이터를 이용하여 가사 작성에 도움을 주는 세 가지 정보를 제시하는 기능을 제안한다.

2. 가사 작성 지원 시스템의 기능

본 연구가 제안하는 가사 작성 지원 시스템의 기능은 총 세 가지로, (1) 주어진 검색어와 관련이 있는 기

존 가사를 검색하는 ‘관련 가사 검색’ 기능, (2) 작사가가 새롭게 작성한 가사와 유사한 기존 가사를 검색하는 ‘유사 가사 검색’ 기능, (3) 사용자가 제시한 단어와 관련이 있는 K 팝 어휘를 제안하는 ‘관련 어휘 제안’ 기능이다. 각 기능의 입력과 출력, 기대 효과를 정리하면 아래와 같다.

첫째, ‘관련 가사 검색(get_related_lyrics_with(keyword))’ 기능은 사용자로부터 키워드 형태의 검색어를 입력으로 받고, 그 키워드와 관련이 높은 가사의 목록을 출력한다. 이 기능은 작사가가 곡의 주제를 어렵듯이 정했지만 구체적인 가사가 떠오르지 않을 때 도움을 줄 수 있다. 제안 시스템에 주제 키워드를 검색어로 입력하면 해당 주제와 관련된 과거 가사들을 제시하고, 관련 가사들을 살펴봄으로써 작사가는 구체적인 가사의 이미지를 떠올릴 수 있다.

둘째, ‘유사 가사 검색(get_similar_lyrics_to(lyric_text))’ 기능은 작사가가 창작한 가사를 입력으로 받고, 해당 가사와 유사도가 높은 과거 가사의 목록을 출력한다. 작사가가 새로 작성한 가사가 기존 가사와 어떤 점에서 유사한지를 점검해 줌으로써 작사가는 새로 창작한 가사의 독창적인 부분을 확대해 나갈 수 있다.

셋째, ‘관련 어휘 제안(get_similar_words_to(word))’ 기능은 사용자로부터 특정 어휘를 검색어로 입력 받고, K 팝 어휘 중 유사도가 높은 어휘들을 출력한다. 이 기능의 가장 큰 기대 효과는 작사가의 어휘 선택의

폭을 넓혀준다는 점이다.

3. 시스템 구현

가사 데이터는 국내 최대 음원 사이트인 멜론의 1980년 3월부터 2023년 1월까지의 월간 차트 100위 내 곡을 대상으로 수집하였고, 당초에는 가사 9,849 개를 수집하였으나, 여기에 영어로 된 가사가 섞여 있어 이를 제거하고 최종적으로 9,591 개 가사를 연구 대상 가사 데이터로 삼고, 가사 작성 지원 시스템을 구현하였다. 가사 수집에는 파이썬의 Selenium과 BeautifulSoup 라이브러리를 사용하였다.

수집한 가사 데이터는 전처리 과정을 거쳤다. 한국어 형태소 분석기(KoNLPy)를 이용하여 형태소를 추출하였고, 불용어도 제거하였다.

단어 목록으로 된 가사(문서) 데이터를 가지고 ‘관련 가사 검색’과 ‘유사 가사 검색’을 진행하기 위해 가사(문서) 속 단어들의 Term Frequency-Inverse Document Frequency (TF-IDF) 값을 계산하고, 단어들의 TF-IDF 값을 요소로 하는 가사(문서) 벡터를 마련하였다. 또, ‘관련 어휘 제안’을 진행하기 위해 가사(문서) 데이터로부터 단어 임베딩 벡터(word embedding vector)를 학습하였다. 이 때 임베딩 벡터는 주변 단어를 통해 중심 단어를 예측하는 Continuous Bag Of Word (CBOW) 방식으로 word2vec 벡터를 학습하였다. Word2vec 학습에는 파이썬의 genism 라이브러리를 사용하였다.

마지막으로 이렇게 마련한 문서 벡터 또는 단어 벡터의 유사도를 계산하기 위해 벡터 간의 코사인 유사도(cosine similarity)를 계산하고, 유사도가 높은 순으로 검색 결과를 정렬하였다.

4. 시스템 출력 결과와 결론

(그림 1)은 ‘관련 가사 검색’의 출력 예시이다. 사용자가 ‘이별’이라는 검색어를 시스템에 입력하면, ‘이별’과 관련이 높은 가사를 검색하여 제시해 준다.

artist	title	release_date	lyric	
오복	소녀	2007.05.15	사람이라면 (사람이라면) 사랑이라면 (사랑이라면) 난 두려워져요 이별	슬픈 이별
어반자카파	똑같은 사랑 똑같은 이별	2012.10.30	욕심에 점점 소라해 지니를잡으려 할수록 더 멀어지는 지켜라 할수록 지쳐 만 가느...	반복되는 이별
백지영	무시로	2011.03.20	이미 와버린 이별인 데 술피도 좋지 않아요 이때 늦은 이별인데 미련	미련없이 이별
김민수, 태연 (TAEYEON)	달라	2011.01.31	죽었던 운동화 끈이 플리츠 점점 멀어진 다 인사말의 반대 말은 이별이 아니다	다가울 이별
인피니트 (The Eye)	태동	2016.09.19	다 끝났다 다 잊었다 이제야 너를 지운 다 잠 걸었던 잠 킴 들었던 이별과 이 별한...	남겨둔 미련과의 이별

(그림 1) ‘이별’이라는 검색어로 관련 가사를 검색한 결과 가사 내용을 살펴 보면 ‘슬픈 이별’, ‘반복되는 이

별’, ‘미련없이 이별’, ‘다가울 이별’, ‘남겨둔 미련과의 이별’ 등 ‘이별’의 내용을 다루고 있음을 알 수 있다.

(그림 2)는 ‘유사 가사 검색’의 출력 예시이다. 실험에서는 새로운 가사를 따로 작성하기 어려워 기존 가사를 선택하여(예: 김현식의 ‘비처럼 음악처럼’) 해당 가사와 유사한 가사를 검색해 보았다. (그림 2)의 왼쪽은 ‘비처럼 음악처럼’의 가사이고(입력), 오른쪽은 해당 가사와 유사한 가사를 출력한 결과이다. 출력된 유사 가사들의 노래 제목을 살펴 보면, ‘비가 오네’, ‘비 오는 날 듣기 좋은 노래’, ‘비가 온다’, ‘추적이는 여름 비가 되어’, ‘비가와’로 모두 비와 관련된 노래임을 알 수 있다.

""""비가 내리고 음악이 흐르면 난 당신을 생각해 당신이 떠나시면 그 밤에 이렇게 비가 왔어요	키스 (Geeks) 비가오네 (Feat. 박수민)	2013.04.29	비가 오네 너와 둘이서 그리워하네 그 시간을 비가 오네 너와 둘이서 간...
비가 내리고 음악이 흐르면 난 당신을 생각해 당신이 떠나시면 그 밤에 이렇게 비가 왔어요	에픽하이 (EPIK HIGH) 비 오는 날 듣기 좋은 노래 (Feat. Colde)	2021.06.29	비가 오고네가 생각나 비가 오고네로 비가 오고 말은 음악이 흐르면 누...
난 오늘도 이 비를 맞으며 하루를 그냥 보내요	강승윤 비가 온다	2013.07.16	장문내려온 울음비가 내 리고니론 뒤로 부는 바람은 비장가의 눈물물 흘리고 뉘...
오아름다운 음악 같은 우리의 사랑의 이야기들은 흐르는 것처럼 너무 아프기 때문이죠	장별준 추적이는 여름 비가 되어	2021.06.11	인생이란 그 짧은 우리 세로누우려는 다시 만날 데 너그날 웃으며 걸어가요 너그...
난 오늘도 이 비를 맞으며 하루를 그냥 보내요	소유 (SOYOU), 백현 (BAEKHYUN) 비가와 (Rain)	2017.02.14	비가 오면 우산 속 너의 어깨 끝까지 떨릴 수 있었어 너의 거...
오아름다운 음악 같은 우리의 사랑의 이야기들은 흐르는 것처럼 너무 아프기 때문이죠 그렇게 온 비가 왔어요""	나달 그대 떠난 뒤	2005.01.08	나 지울 수 없는 지난 추억 물안방울과 생각해보네 이 비를 맞으며 걸던 이별 물만...

(그림 2) ‘비처럼 음악처럼’의 가사(왼쪽)와 유사한 가사를 검색한 결과(오른쪽)

마지막으로 ‘유사 어휘 제안’의 출력 예시인데, ‘추억’이라는 어휘의 유사 어휘를 시스템에서 찾은 결과, ‘기억’, ‘흔적’, ‘아쉬움’, ‘지난날’, ‘잊혀지다’, ‘아픔’, ‘지워지다’ 등이 제안되었다. 이 어휘들은 K 팝이라는 도메인의 특성을 잘 반영한 어휘들이라고 할 수 있다.

결론적으로 본 연구에서는 TF-IDF, 단어 임베딩, 코사인 유사도를 활용하여 가사 작성 지원 시스템을 개발하였는데, 1만 건에 가까운 가사 데이터 덕분에 간단한 기술로도 유의미한 검색 결과를 얻을 수 있었다.

5. Acknowledgment

이 논문은 2017년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017S1A6A3A01078538).

참고문헌

- [1] Yading Song, Simon Dixon, Marcus Pearce, “A survey of music recommendation systems and future perspectives,” The 9th International Symposium on Computer Music Modeling and Retrieval (CMMR), June 2012.
- [2] Aaron van den Oord, Sander Dieleman, Benjamin Schrauwen, “Deep content-based music recommendation,” Advances in Neural Information Processing Systems 26 (NeurIPS 2013), 2013.
- [3] Teh Chao Ying, Shyamala Doraisamy, Lili Nurliyana Abdullah, “Lyrics-based genre classification using variant tf-idf weighting schemes,” Journal of Applied Sciences, 15: 289-294, 2015.