

## 노래 가사의 감성 분석을 활용한 음악추천 시스템

CUAI 4기 문화콘텐츠 추천시스템-가사 팀  
서희재(컴퓨터공학부), 음호준(물리학과), 허인(응용통계학과)

### [요약]

개인 맞춤형 음악 추천 알고리즘이 각종 음악 플랫폼에서 활용되고 있다. 현재의 추천 알고리즘은 (1)비슷한 취향(장르, 감성)을 가진 다른 사용자의 음악 추천 - 협업 필터링 (2)음악의 오디오 데이터를 기반으로 음악 추천 - 콘텐츠 기반 필터링 (3)유사한 태그 아이디를 가진 음악 추천 - 콘텐츠 기반 필터링 - 3가지를 목표로 작동한다. 이에 본 연구는 기존의 음악 시스템 구현과는 다른 방향으로, 가사를 활용한 음악 추천 시스템을 제안하고자 한다. Genius API를 이용해 얻은 가사 데이터에 대해 자연어 처리를 실시하고 감성 분석 모델을 활용해 각 음악의 감성을 분류하여 유사한 감성을 지닌 음악을 추천하는 시스템을 제안한다.

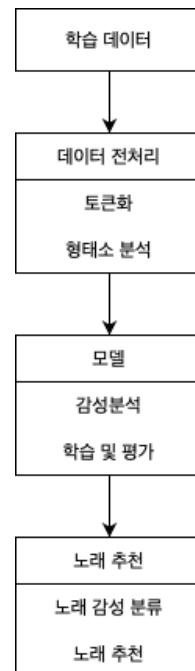
### 1. 서 론

개인화 시대가 도래했다. 기술의 발전 및 데이터의 축적과 함께 서비스 및 상품(e.g. OTT, 의류 및 음악 플랫폼)이 개인의 취향을 겨냥해 출시되고 있다. 시장조사전문기업 엠브레인 트렌드모니터가 전국 만 19세~59세 성인 남녀 1,000명을 대상으로 ‘취향’에 대한 인식 조사를 실시한 결과, 80% 이상이 개인의 취향은 존중되어야 하며, 개인의 취향을 존중하는 것 자체가 하나의 트렌드가 되었다고 응답했다. 대표적인 음악 플랫폼 스포티파이는 (1)협업 필터링(콘텐츠, 아이템 기반 필터링) (2)노래의 오디오 관련 요소(e.g. 파형(Mel-Spectrogram))를 분석하여 분위기 혹은 무드 분석 (3)노래의 언어, 가사 및 메타데이터 분석 - 3가지 방법을 활용한다[1]. 이에 본 연구는 노래의 가사에 내재된 감성을 분석하여 음악을 추천할 수 있는 시스템을 제안한다.

본 연구는 이를 위해 1980년부터 2019년까지의 한국 노래 및 가사, 아티스트 데이터를 연대 별로(10년 단위) 수집했다. 가사에 대한 형태소 분석 및 토큰화를 진행하였고, 감성 분석 모델을 생성하여 가사 데이터에 대한 학습, 평가를 실시한 후, 문서별 벡터간 유사도 알고리즘을 활용해 유사도가 높은 가사를 가진 노래를 추천하는 것을 최종 목표로 한다.

### 2. 본 론

본 연구는 다음의 방향으로 진행된다.



#### 2.1 데이터 전처리

##### 2.1.1 데이터 수집

멜론 시대별 차트 중 1980년대부터 2010년대까지의 국내 음악 차트[2]를 참조하여 연대별로 상위 20명 내외의 아티스트를 가사 수집 대상으로 선정하였다. 멜론 웹사이트는 웹크롤링을 통한 가사 수집이 불가능하기 때문에, Genius.com[3]에서 제공하는 API를 이용해 가사 텍스트 데이터를 수집하기로 결정하였다. Genius API는 아티스트 이름에 한정된 쿼리에 대해서도 곡의 모든 정보가 포함된 정보를 JSON 포맷의 데이터로 반환하기 때문에, 가사 수집에 용이한 인터페이스를 제공하는 클래스로 랩핑된 파이썬 패키지인 LyricsGenius[4]를 이용하여 텍스트를 수집하였다.

수집된 텍스트 데이터를 csv 포맷으로 저장하고 텍스트를 검수한 결과, Genius를 통해 수집된 가사 텍스트 중에는 외국어로 번역된 가사도 포함되어있는

경우가 있어, 이러한 경우는 외국어 가사를 모두 제거하였다. 또한, 시대 흐름에 따라 외국어가 가사에서 차지하는 비중이 커지는 것을 Figure 2에서 확인할 수 있다. 본 연구는 한국어를 대상으로 자연어 처리를 진행하는 것에 의의를 두어 가사에서 한국어의 의미만을 근거로 분류하는 것으로 결정하여 외국어 가사는 제거하였다. 국내 음악 및 가사는 테스트 데이터로 활용하였다.

Figure 2. 시대에 따른 가사의 외국어 비율 증가

감성 분류 모델 train 및 validation을 위하여 AIHub의 감성 대화 말뭉치[5]을 수집했다. 크라우드 소싱 수행을 통해 제작된 세대별 감성대화 텍스트 데이터로, 각 대화 별로 6개의 기본 감정(분노, 슬픔, 불안, 상처, 당황, 기쁨)과 9개의 세부 감정 중 각각 1개의 감정이 레이블링 되어있다. 사람과 시스템의 번갈아가며 대화하는 형태로, 최대 인간의 발화 3개, 시스템 발화 3개로 6개의 대화문을 각 레코드 별로 보유하고 있다. 총 데이터 개수는 46,009개이다.

시스템의 발화는 감성 분류와 무관하다고 판단해 사람의 발화문 중 첫번째 발화문을 학습 대상 텍스트로 선정하였다. 또한, 세부 감정을 모두 포함한 60가지 분류는 성능 저하를 가져올 수 있으므로 6가지 기본 감정만을 사용했다.

### 2.1.2 노래 가사 토큰화 및 형태소 분석

발화문 데이터셋을 KoNLPy의 Mecab 클래스를 활용해 토큰화 및 형태소 분석을 진행했다. 형태소 및 라벨링 데이터를 활용해 모델 성능을 평가한 뒤, 성능 개선을 위하여 형태소 분석을 재진행했다.

1차 분석에서는 NNG, NNP, NNG, NNBC, NR, NP, VV, VA, VX, VCP, VCN (각 태그는 일반 명사, 고유 명사, 의존 명사, 단위를 나타내는 명사, 수사, 대명사, 동사, 형용사, 보조 용언, 긍정 지정사, 부정 지정사, 관형사, 일반 부사, 접속 부사)[6] tag를 지닌 토큰을 선택하였다. 2차 분석에서는 감정의 의미를 담고 있는 토큰을 추출하기 위하여 1차 분석에서 활용한 NNB, NNBC, NR, NP을 제외하고 NNG, VV, VA, VX, VCP, VCN, MM(관형사), MAG(일반부사) tag을 지닌 토큰을 추출하였고, 이를 활용해 모델 학습 및 테스트, 성능 평가를 재진행하였다.

|   | sent1   | sent2   | sentiment | sent   |
|---|---|---|-----------|--|
| 0 | 아내가 드디어 출산하게 되어서 정 말 신이 나.                      | 아 지금 정말 신이 나.                                     | 0         | [아내, 출산, 되, 신, 신]                                    |
| 1 | 당뇨로 합병증 때문에 먹어야 할 약 이 일 가지가 넘어서니까 스트레스 야.       | 건강할 때 관리 좀 잘할걸 하는 생각 이 들더라고.                      | 3         | [당뇨, 합병증, 먹, 할, 넘어야, 스 트레스, 아, 건강, 때, 관리, 할, ...]    |
| 2 | 고등학교 올라오니 중학교 때보다 수업이 딱지기 어려워서 당황 스러워.          | 아직 학기 초인데 내가 수업에 잘 따 라갈 수 있을지 걱정돼.                | 1         | [고등학교, 올라오니, 중학교, 때, 수업, 어려워서, 당황, 학기, 안 대, 수업]      |
| 3 | 제취업이 되어서 받게 된 첫 월급으로는 그녀가족이 외식을 할 예정이야. 나도 행복해. | 퇴직 후 다시는 돈을 못 벌 줄 알았는데 이렇게 월급으로 가족에게 맛있는 밥을 살 ... | 0         | [제취업, 대체, 받, 된, 월급, 그녀가족, 외식, 할, 예정이야, 나도, 행복해, ...] |
| 4 | 빛을 드리다 갑자기 되어서 아내에게 안도감을 들어.                    | 빛도 다 같았으니 당분간은 저녁에 생각 이 안 하면 괜찮을 걸.               | 0         | [빛, 갑, 되, 대신, 듣, 빛, 괜찮, 생각, 하, 괜찮, ...]              |

Figure 3. 토큰화 및 적절한 형태소 추출

### 2.1.3 Word Embedding

본 연구의 감성분류 모델의 임베딩 레이어에는 사전 학습된(pre-trained) 워드 임베딩 벡터를 사용하기로 결정하였다.

단순한 형태의 형태소-벡터 변환은 단어사전 크기의 원-핫 인코딩 벡터로도 가능하기는 하지만, 원-핫 인코딩 벡터는 해당 단어의 인덱스가 1이고 이를 제외한 모든 인덱스에서의 값이 0인 희소 벡터이기 때문에, 전체 단어 사전 내에서 단어와 단어 간의 유사도 정보를 내포하기에는 어려움이 있다. 이를 개선하기 위해 제안된 것이 단어를 밀집 임베딩 벡터로 변환하는 방법이다. 밀집 임베딩 벡터는 임의의 크기와 실수값을 가지는 벡터로, 같은 크기의 단어 사전이라도 원-핫 인코딩 벡터(희소 벡터)에 비해 상대적으로 작은 크기(밀집 벡터)가 된다.

단어간 유사도를 표현하기 위한 단어 벡터는 대용량의 말뭉치 데이터가 필요한 작업이므로 이번 테스크에서 단어 벡터를 형성하기 위한 충분한 크기의 데이터를 수집하기에는 어려움이 있다. 따라서 사전 학습된 워드 임베딩 벡터로는 Kyubyong/wordvectors[8] 을 사용하기로 결정하였다. Kyubyong/wordvectors는 크기 200의 단어벡터로 구성되어있으며, 339Mb 크기의 한국어 말뭉치, 크기 30185의 단어 사전으로 사전 학습된 단어 벡터이다.

## 2.2 감성 분석 모델

본 연구에 사용되는 모델은 분석 대상인 문장을 토큰화하고, 일련의 토큰의 연속을 분류하고자 하는 값으로 귀결시키는 것을 목표로 한다. 따라서 신경망 모델 중 연속된 값을 분류하는 데 적합한 RNN 모델을 사용하기로 결정하였다.

LSTM(Long-Short-Term-Memory)는 RNN 모델의 하나로, 기존 RNN 모델에서 입력의 크기가 커짐에 따른 정보의 소실로 인한 분류 성능 저하 현상을 개선하기 위한 모델이다.

본 연구에서는 임베딩 벡터 크기 200, 은닉 상태 100,

은닉층 1, 분류 클래스 6의 양방향(Bi-directional) LSTM 모델을 적용하고, 드롭아웃 0.3, 소프트맥스 함수를 사용하여 분류 작업을 수행하였다. 그 결과 55~57%의 검증 데이터셋에 대한 정확도를 얻을 수 있었다.

### 2.3 연구 한계

#### 2.3.1 학습 데이터의 레이블 오류

학습 데이터의 발화문의 실제 감정과 레이블링된 감정의 불일치가 존재하였다. 잘못된 레이블링은 학습 모델의 성능에 직접적인 영향을 끼친다. 레이블 오류를 수정하기 위해서는 기존의 데이터에 대하여 다시 레이블링을 진행해야 하는 어려움이 있다. AIHub에서 수집된 발화문 데이터의 감정에 관한 레이블은 '감정 대분류', '감정 소분류' 두 가지가 있다. 감정 대분류 레이블은 각 발화문에 대한 대표적인 감정 레이블로 6 가지이다. 감정 소분류 레이블은 각 발화문에 대한 세분화된 감정 레이블로 60 가지이다. AIHub에서 제공된 최종 데이터셋을 직접 분석, 검수한 결과, 데이터 품질에 여러가지 문제가 있었음이 확인되었다.

첫째로, 감정 대분류와 감정 소분류의 간극이 큰 데이터가 존재했다. 예를 들어, 감정 대분류가 '분노'이고 감정 소분류가 '좌절'이며 첫번째 발화문이 "내가 일할 수 있는 자리가 점점 줄어든다는 걸 생각하니 너무 속상해." 인 레코드가 존재했다. 이를 살펴보면 사람의 생각으로는 '좌절'에 해당하는 감정이 '분노'의 감정을 내포하는 것이 가능하다고 생각되고 어느 정도 타당하다고 생각 될 수 있지만, 기계 학습을 위한 데이터로써는 상당히 부적합하다고 판단되어진다. 이와 같이 감정 대분류와 소분류 간의 차이가 큰 경우에는, 분류하고자 하는 범주의 핵심 특징과 차이가 큰 경우까지 해당 범주로 분류되는 것으로 학습에 반영되어, 정상적인 학습을 방해할 수 있다.

둘째로, 감정 레이블링 자체가 문제가 있는 데이터가 존재했다. 첫째 문제점에서 살펴본 데이터에서도 나타나는 문제점인데, "내가 일할 수 있는 자리가 점점 줄어든다는 걸 생각하니 너무 속상해."라는 문장이 "좌절"이라는 소분류로 레이블링 되어있는 것을 확인할 수 있다. 아무리 자연어라고 하지만 기계학습 모델에서 위의 문장을 토큰화한 단어들만으로는 "좌절"이라는 레이블을 유추하기에는 무리가 있다. 심지어 우리는 이 문장을 소분류인 "좌절"이 아니라 대분류인 "분노"에 맵핑시켜 학습 데이터로 사용하였는데, 특히나 주로 적용되는 분류 작업인 '부정', '긍정' 이진 분류가 아니라 다중 클래스 분류 작업이 목표인 본 연구에 있어서는 더욱 학습에 부적절한 데이터가 된다.

셋째로, 본 연구의 목적에 맞지 않는 데이터가 수집되었다. 본 연구의 최종 목적은 한국어 노래 가사의

텍스트를 입력으로 받아 각 문장별로 나누고, 나누어진 문장들을 감정 분류 모델의 입력으로 사용하여 분류 클래스 별 확률 벡터를 출력으로 구한 후, 이를 종합하여 노래 가사간의 유사도를 구하는 것이다. 따라서 중요한 것은 '노래 가사 텍스트에 대하여 얼마나 정확한 감정을 유추하는가'에 대한 것인데, 이를 간파하였다. 노래 가사는 크게는 '정제된 텍스트'에 해당하며 좁게는 '예술 문화 및 문학 텍스트'에 속한다. 이처럼 좁은 주제에 속한 데이터는 당연히 해당 도메인의 특징을 보유하게 된다.

노래 가사의 경우에는 단어와 단어간의 음성적 관계성(운율 등)이나, 단어가 암시적으로 가지는 의미, 일상적으로 잘 쓰이지 않는 선별적 어휘 등을 특징으로 가진다. 하지만 감정 분류 모델의 학습 데이터로써 수집하였던 AIHub의 말뭉치 데이터셋은 일반 사용자들을 대상으로 수집한 일상 대화의 문장들이다. 이는 위에서 언급한 노래 가사 텍스트가 가지는 특징들과는 전혀 다른 특징들을 가지고 있다. 따라서 만약 AIHub의 데이터로 정확도 높은 학습이 진행되었다고 하더라도 노래 가사 텍스트에 적용했을 때 본 연구의 본질적인 목적과 부합하는 의미있는 결과가 나오지 않을 가능성이 높을 것으로 짐작된다.

### 3. 결 론

개인 맞춤형 서비스에 대한 수요가 증가함에 따라 개인의 취향, 감성에 적합한 서비스 제공의 필요성이 증대되었다. 이에 본 연구는 가사의 감성을 분석하여, 사용자가 원하는 감성의 가사를 지닌 노래를 추천하는 것을 목표로 하였다.

연구를 진행하는 과정에서 예상보다 상당히 낮은 수치의 정확도를 얻게되어, 문제점에 대한 고찰을 통해 데이터의 결함을 발견하였다. 데이터의 레이블 오류와 학습 및 테스트 데이터 특성의 간극으로 인하여 모델 성능이 저하되는 문제가 발생하였다. 문제 해결을 위해 데이터 라벨링 작업을 추가로 진행하거나, 알맞게 라벨링이 이루어진 데이터 수집이 필요할 것으로 보인다. 분류하고자 하는 감정에 대한 특징을 잘 포함하고 있는 문장들을 잘 선별하거나 수집한 후, 적절하게 오버샘플링 기법을 적용하여 양질의 데이터를 증가시키는 것이 학습에 큰 도움을 줄 것으로 예상된다. 또한, 본 연구의 목표가 노래 '가사'에 내재된 감성을 분석하는 것이기에, 노래 가사와 유사한 특성을 지닌 데이터(e.g. 문학 작품 구절)를 테스트 데이터로 활용해야 할 것으로 보인다. 이러한 점을 보완한다면 본 연구의 모델 성능이 높아져 최종 목표인 감성 분류에 따른 노래 추천이 용이할 것으로 기대한다.

## 참고 문헌

[1] How does Spotify's algorithm Work? Streaming hacks for musicians

<https://dittomusic.com/en/blog/how-does-spotifys-algorithm-work-streaming-hacks-for-musicians/>

[2] 멜론

<https://www.melon.com/chart/age/index.htm?chartType=YE&chartGenre=KPOP&chartDate=2020>

[3] Genius API <https://docs.genius.com/#/songs-h2>

[4] LyricsGenius

<https://github.com/johnwmillr/LyricsGenius>

[5] AIHub 감성 대화 말뭉치

<https://aihub.or.kr/aidata/7978>

[6] Korean POS tags comparison chart

<https://docs.google.com/spreadsheets/d/1OGAjUvalBuX-oZvZ-9tEfYD2gOe7hTGsgUpiiBSXI8/edit#gid=0>

[7] AIHub 한국어 감정 정보가 포함된 단발성 대화 테이터셋

[https://aihub.or.kr/keti\\_data\\_board/language\\_intelligence](https://aihub.or.kr/keti_data_board/language_intelligence)

[8] Kyubyong/wordvectors

<https://github.com/Kyubyong/wordvectors>