

딥러닝 모델(BERT)과 감정 어휘 사전을 결합한 음원 가사 감정 분석

윤경섭*, 오종민^o

*인하공업전문대학 컴퓨터정보공학과,

^o인하공업전문대학 컴퓨터정보공학과

e-mail: ksyoon@inhac.ac.kr*, alswhddh@naver.com^o

Analysis of Emotions in Lyrics by Combining Deep Learning BERT and Emotional Lexicon

Kyung Seob Yoon*, Jong Min Oh^o

*Dept. of Computer Science, Inha Technial College,

^oDept. of Computer Science, Inha Technial College

● 요약 ●

음원 스트리밍 서비스 시장은 지속해서 성장해왔다. 그중 최근에 가장 성장세가 돋보이는 서비스는 Spotify와 Youtube music이다. 두 서비스의 추천시스템은 사용자가 좋아할 만한 음악을 계속해서 추천해 줌으로써 많은 사랑을 받고 있다. 추천시스템 성능은 추천에 활용할 수 있는 변수(Feature) 수에 비례한다고 볼 수 있다. 최대한 많은 정보를 알아야 사용자가 원하는 추천이 가능하기 때문이다. 본 논문에서는 기존에 존재하는 감정분류 방법론인 사전기반과 딥러닝 BERT를 사용한 머신기반 방법론을 적절하게 결합하여 장점을 유지하면서 단점을 보완한 하이브리드 감정 분석 모델을 제안함으로써 가사에서 느껴지는 감정 비율을 분석한다. 감정 비율을 음원 가중치 변수로 사용하면 감정 정보를 포함한 고도화된 추천을 기대할 수 있다.

키워드: 자연어처리(NLP), 버트(BERT), 감정(emotion), 가사(lyrics), 딥러닝(deep learning), 사전(lexicon)

I. Introduction

노래를 듣다 보면 다양한 감정을 느껴본 경험이 있을 것이다. 사랑 노래를 들으면 사랑하고 싶어지고 슬픈 노래를 들으면 과거 슬펐던 기억이 떠오르기도 한다. 이처럼 노래에서 감정은 중요한 요소이다. 노래에는 멜로디, 가수의 목소리, 템포 등 많은 요소가 있지만 그 중 가사는 멜로디에 몰입할 수 있도록 도움을 줄뿐더러 리스너의 공감을 끌어낼 수 있는 중요한 매개체 역할을 한다. 또한 가사는 감정을 가장 잘 표현한 도메인이라고 볼 수 있다. 가사에서 느껴지는 감정 정보를 얻을 수 있다면 음원 스트리밍 서비스에서 고도화된 음원 추천시스템을 기대할 수 있다.

본 논문에서는 국내 인기 음원 가사를 수집하여 효과적인 가사감정 분석 방법을 제안한다.

II. Previous literature

2.1. 기존 가사 감정 분류 방법론

기존 감정 분류 방법[4]은 Fig 1과 같이 크게 2가지로 나누어 볼 수 있다. 사전 기반 접근법(Lexicon-based Approach)[1]은 어휘사전을 이용하는 방법으로 메뉴얼 기반, 사전 기반, 말뭉치 기반 접근법이 있다. 그중 말뭉치 기반 접근법은 대용량 말뭉치를 통해 감정 어휘를 재구축하는 방법이다. 도메인 의존성을 극복할 수 있어서 새로운 도메인에 대한 분석을 시도할 때 많이 사용된다. 가사말뭉치를 사용하면 도메인에 맞는 감정 어휘사전을 구축할 수 있다.

머신 기반 접근법(Machine Learning Approach)[2]은 인공신경망을 활용한 딥러닝 기법이 대표적이며 2018년 구글에서 공개된 BERT[3]는 대용량 말뭉치를 학습한 언어모델이다. 사전 학습(Pre-Trained)된 언어모델을 해결하고 싶은 작업에 맞게 파인 튜닝(Fine-Tuning)하면 높은 성능을 보여줄 수 있고 실제로도 자연어 처리 모든 분야에서 상위권 성능을 유지하고 있다. Transformer구조의 Self-Attention 기법을 사용해 어휘의 순서와 전체적인 문맥을

과약하여 분류 문제에서도 높은 성능을 보인다.

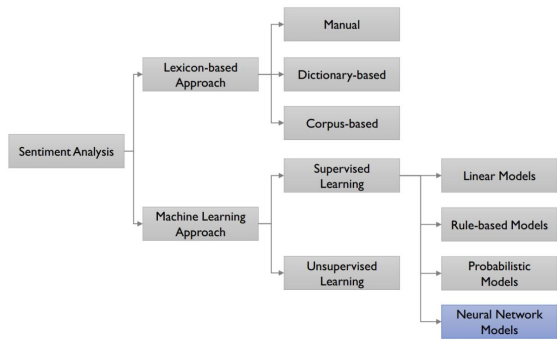


Fig. 1. Emotion Analysis Methodology

2.2 기존 기사 감정 분류의 문제

Fig 1에 두 가지 방법론에는 각각의 치명적인 단점이 있다. 사전 기반 접근법은 등장 어휘들을 구축한 사전에 대입시켜 분석하기 때문에 순서와 문맥을 전혀 고려하지 않고 오로지 어휘만 보고 판단하는 단점이 있다. 예를 들면 ‘좋아하는 사람과 만날 수 없네요’라는 문장은 ‘슬픔’에 대한 기사지만 어휘만 가지고 분석하면 [‘좋아한다’, ‘사랑’, ‘만나다’, ‘없다’]로 분석되어 ‘사랑’으로 분류할 가능성이 크고 ‘만날 수 없다’ 같은 부정어를 고려하지 못하는 단점이 있다.

머신 기반 접근법은 어휘의 순서와 전체적인 문맥을 파악하여 부정어를 잘 처리해 사전 기반 접근법의 단점을 해결하고 좋은 성능을 기대할 수 있지만 좋은 성능도 대용량의 학습데이터를 가지고 있어야 하고 특징을 찾아서 분석하기 때문에 세부 감정 분석 시 정확도가 많이 떨어질 수밖에 없다. ‘사랑’과 ‘슬픔’은 잘 구별하지만 ‘사랑’과 ‘행복’에 대한 분석 정확도는 많이 떨어지고 분류 클래스가 늘어날 경우 정확도는 더 많이 떨어지게 된다. 실제 테스트로 15,327개 문장 데이터로 BERT 9개의 다중 분류 학습을 진행했을 때 40%의 낮은 정확도를 보였다.

III. Design and Implementation

3.1 음원 기사 감정 데이터 구축 방법

3.1.1 기사 데이터 수집 및 전처리

본 연구에서는 국내 인기 음원들을 여러 장르로 구분하여 가사용 데이터를 수집하였고 음원 400곡을 18,613개의 문장(소절)단위로 분리하였다. 가사를 문장으로 나눈 이유는 다음과 같다.

(1) 가사의 텍스트 길이가 너무 길다. BERT모델의 입력 토큰 수는 제한적이다. (2) 가사의 감정은 한 가지가 될 수 없다. (3) 기사 감정은 서로 다른 클래스들과의 차이가 크지 않다. 슬픔과 외로움의 감정차이는 크지 않다. (4) 한 가사에 여러 라벨을 붙이면 어떤 감정이 더 큰지 알 수 없다. (5) 한 가사에 여러 라벨이 붙으면 지도학습하기 어렵다. (6) 문장 단위로 분리하여 합산하는 방법은 오분류율을 낮출 수 있다, 잘못 분류해도 가사에는 여러 문장들이

있어서 분석결과에 영향이 적다.

3.1.2 감정 리벨 선정

감정 연구의 대가 ‘Ekman’이 정의[5]한 감정은 행복, 분노, 슬픔, 혐오, 놀람, 두려움이다. 그러나 텍스트를 통해 느껴질 수 있는 감정들은 정해져 있다.

‘놀람’은 텍스트에서 느껴지기보단 표정으로 판단할 때가 많아 제거했고 ‘열정’, ‘그리움’, ‘외로움’, ‘즐거움’은 힙합, 발라드, R&B등 음원 장르 특성에 맞게 새로 추가하였다. 최종적으로 사랑(love), 즐거움(fun), 열정(enthusiasm), 행복(happyness), 슬픔(sadness), 분노(anger), 외로움(lonely), 그리움(logging), 두려움(fear)으로 총 9가지 감정으로 재정의 하였다.

3.2. 하이브리드 감정 분석 모델

머신기반과 사전기반 방법론은 각각의 치명적인 단점이 존재하지만, 장점들을 결합하면 단점을 보완할 수 있고 훨씬 좋은 성능을 낼 수 있다. 새롭게 고안한 하이브리드 감정 분석 모델은 먼저 사전 학습된 딥러닝 BERT 모델로

분리된 기사 문장의 긍/부정을 파악하고 그 결과를 감정 어휘 사전 점수 계산에 반영한다. 그다음 문장별 감정 점수를 합산하여 가사의 최종 감정 비율 도출한다. 문맥과 어휘 두 가지를 전부 고려하기 때문에 기존 방법론들보다 더 확실한 결과를 기대할 수 있다.

3.3. 딥러닝 긍/부정 분류 모델 학습

머신기반 방법론에서 딥러닝 모델이 잘하는 것은 이진 분류이다. 한국어 벤치마크 데이터 세트 중 네이버 영화리뷰 데이터(NSMC)[6]에서 BERT모델 대부분 Table 1과 같이 90%에 근접한 높은 정확도를 기록[7]한걸 볼 수 있다.

Table 1. Performance of NSMC by BERT Model

BERT-NAME	Accuracy
kcBERT-base	89.62
kcBERT-Large	90.68
koBERT	89.63
XLM-Roberta-base	89.49
HanBERT	90.16

구축한 데이터 셋에서 ‘사랑’, ‘즐거움’, ‘열정’, ‘행복’ 감정들을 긍정으로 치환하고 ‘슬픔’, ‘분노’, ‘외로움’, ‘그리움’, ‘두려움’ 감정들을 부정으로 치환시켜 학습데이터를 이진분류로 재구축했다. BERT 모델은 한국어 위키피디아를 기반으로 사전 학습된 KoBERT(SKTBrain)[8]를 선택하여 긍/부정 학습시켰다. 분류기의 평가 지표는 Table 2와 같고 예시 문장을 테스트 해보았을 때 Table 3와 같이 ‘사랑’이라는 단어가 모두 포함 되었지만 문맥을 잘 고려하여 분류하는 것을 확인할 수 있다.

Table 2. Positive/negative Evaluation Metrics

	Precision	recall	F1-score
Negative	0.79	0.78	0.79
Positive	0.88	0.88	0.88
accuracy			0.85

Table 3. Positive/Negative Classifier Test Results

lyrics	Positive	Negative
사랑 때문에 눈물이 흘렀어	7%	93%
사랑하기 싫어요	13%	87%
사랑해서 웃음이 나와	98%	2%
사랑하기 좋아요	99%	1%

3.4. 가사 감정 어휘 사전 구축 방법

사전 기반 접근법 중 말뭉치 기반 접근법을 통해서 가사 도메인에 알맞은 가사 감정 어휘 사전을 구축했다. 감정 어휘 사전은 감정별로 0~1사이의 점수가 매겨져 있으며 모든 감정의 점수를 합하면 1이 되도록 하였다. 전체 빈도수에서 각각의 감정 등장 빈도수를 나눠준 값을 감정 점수로 사용했다.

3.4.1 한국어 형태소 분석기

형태소 분석기는 Komoran 형태소 분석기[9]를 사용했으며 형태소는 일반명사(NNG), 고유명사(NNP), 의존명사(NNB) 형용사(VA), 동사(VV), 일반부사(MAG), 감탄사(IC), 외국어(SL)만 포함시켰다. 동사와 부사에는 다른 형태소와 구별하기 쉽게 “+다”를 붙여서 Table 5와 같이 감정 점수를 포함한 감정 어휘사전을 구축하였다.

Table 4. Emotional Lexicon Words Filtering

테스트 문장	형태소 분석/ 불용어 필터링
사랑을 잃은 이 마음을	사랑, 잃다, 마음
후회는 없잖아 그렇게 이유 없이 사랑했잖아	후회, 그렇게, 이유, 사랑
재미있게 듣던 말 이 책 중에 있으니	재미있다, 듣다, 말, 책, 중
이렇게 지껄어도 아무 의미 없다	이렇다, 지껄이다, 의미
가만히 있어도 내 생각은 하는지	가만히, 생각

Table 5. Emotional Lexicon constructed with a corpus of lyrics

	사랑	즐거움	열정	행복	슬픔	분노	외로움	그리움	두려움
사랑해	0.78	0	0	0	0.05	0.05	0	0.05	0.08
안다	0.78	0	0	0	0.13	0	0	0.04	0.04
아름답다	0.74	0	0	0.12	0.04	0	0	0.08	0.01
고맙다	0.66	0	0	0.07	0.07	0	0.03	0.17	0
영원	0.74	0.11	0	0.07	0	0	0	0	0.07
이별	0	0	0	0	0.8	0	0.02	0.16	0.02
미워하다	0	0	0	0	0.8	0	0	0.2	0
눈물	0	0	0	0	0.79	0.02	0.02	0.09	0.03
헤어지다	0.06	0	0	0	0.69	0.09	0.07	0.04	0.06
힘들다	0.05	0	0.11	0.07	0.59	0.03	0.1	0.01	0.04

3.5. 하이브리드 모델 구현

구축된 감정 어휘 사전을 살펴보면 “사랑”이라는 단어는 Table 6와 같이 사랑 노래에 일반적으로 61% 등장하지만 슬픈 노래에도 18% 등장한다.

Table 6. Emotional Lexicon Example words

	사랑	행복	즐거움	슬픔	분노	외로움
사랑	0.61	0.02	0.01	0.18	0.07	0.04
마음	0.27	0.04	0.05	0.33	0.00	0.09

사전에 대입하여 점수화하면 사랑 노래에 등장한 “사랑”은 슬픔 감정 점수 0.18이 매번 점수에 합산될 것이다. Fig 2와 같이 딥러닝으로 학습시킨 긍/부정 분류기로 문장에 긍/부정 문맥을 파악하고 만약 긍정이면 뒤에 부정 감정인 [‘슬픔’, ‘분노’, ‘외로움’, ‘그리움’, ‘두려움’]을 Table 7과 같이 점수 계산에서 제외시킨다. 반대로 딥러닝의 결과가 부정이면 Fig 3와 같이 긍정 감정인 [‘사랑’, ‘행복’, ‘열정’, ‘즐거움’]을 점수 계산에서 제외해 최종 점수를 합산한다.

Table 7. Emotional Lexicon Score If Positive

	사랑	행복	즐거움	슬픔	분노	외로움
사랑	0.61	0.02	0.01	X	X	X
마음	0.27	0.04	0.05	X	X	X

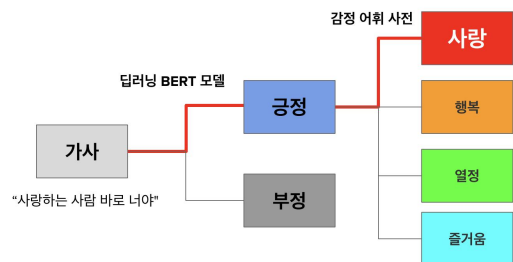


Fig. 2. Love Emotion Classification Case

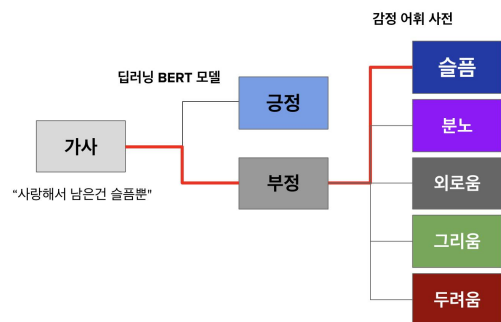


Fig. 3. Sadness Emotion Classification Case

IV. Evaluation

감정의 평가지표는 정량적으로 판단할 수 없다. 사람마다 가사에서 느껴지는 감정이 다르기 때문이다. 하지만 하이브리드 감정 분류

모델의 결과는 대부분이 납득할 수 있는 감정의 비율을 도출한다. 핵심은 한 가지 감정을 결과로 뽑는 것이 아닌 9가지 감정 비율을 도출한다는 것이다.

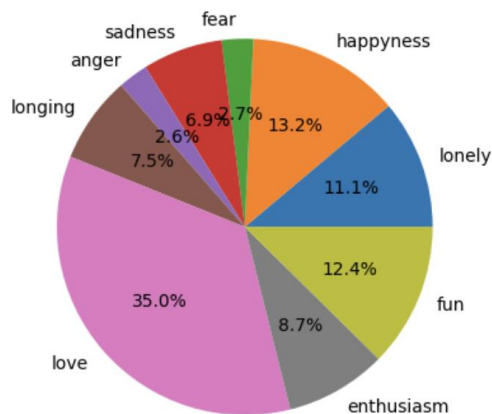


Fig. 4. Hybrid Emotional Analysis Report

Fig 4는 가수 “아이유”의 “내 손을 잡아” 감정 분석 결과이다. 사랑/행복한/신나는 감정이 TOP3로 랭크되었다. 물론 이해할 수 없는 감정들도 적지않은 비중을 차지하고 있지만 결과적으로 사랑 노래에 가장 가깝다고 볼 수 있다. 또한 하이브리드 모델은 가사를 문장(소절)으로 분리하고 문장을 다시 단어로 분리하여 정밀하게 분석하기 때문에 설상 딥러닝 예측이 틀리거나 감정 어휘 분석이 잘못되어도 다른 모델에 비해 결과에 영향이 적다.

V. Conclusion

본 연구를 진행하기 전 다양한 감정 분류 방법을 가사 도메인으로 시도해 보았지만 기존의 두 방법론 모두 치명적인 단점이 존재했다. 본 논문에서는 최신 딥러닝 BERT 모델과 감정 어휘사전 구축을 통해 기존 방법론의 장점을 유지하면서 단점을 보완한 하이브리드 감정 분석 모델을 제안하였고 그 결과로 9가지 감정의 다중 분류 모델을 구현할 수 있었다. 이를 통해 음원 가사별로 9가지 감정에 대한 비율을 가중치로 사용하면 감정 정보를 포함한 고도화된 추천을 기대할 수 있다.

REFERENCES

[1] Jurek, A., Mulvenna, M.D. & Bi, Y. “Improved lexicon-based sentiment analysis for social media analytics“. Secur Inform 4, 9 (2015). <https://doi.org/10.1186/s13388-015-0024-x>

[2] Kawade, Dipak R., and Kavita S. Oza. "Sentiment analysis: machine learning approach." International Journal of

Engineering and Technology 9.3 (2017): 2183-2186.

[3] Evlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv:1810.04805 (2018).

[4] Pilsung Kang, "Text-Analytics", <https://github.com/pilsung-kang/text-analytics>

[5] Wikipedia, "Paul Ekman", https://ko.wikipedia.org/wiki/Paul_Ekman

[6] Naver, "Naver sentiment movie corpus v1.0", <https://github.com/e9t/nsmc>

[7] Junbum Lee, "KcBERT", github.com/Beomi/KcBERT

[8] SKTBrain, "KoBERT", <https://github.com/SKTBrain/KoBERT>

[9] Shineware, "Komoran Morpheme Analyzer", <https://docs.komoran.kr/>