감정 분석에서의 심리 모델 적용 비교 연구

김해준○, 도준호, 선주오, 정서희, 이현아
금오공과대학교, 컴퓨터소프트웨어공학과
dubo4050@naver.com, djh20@naver.com, qssz1326@naver.com, jsh001505@naver.com, halee@kumoh.ac.kr

A Comparative Study on Sentiment Analysis Based on Psychological Model

Haejun Kim○, Junho Do, Juoh Sun, Seohee Jeong, Hyunah Lee
Kumoh National Institute of Technology, Department of Computer Software Engineering

요 약

기술의 발전과 함께 사용자에게 가까이 자리 잡은 소셜 네트워크 서비스는 이미지, 동영상, 텍스트 등 활용 가능한 데이터의 수를 폭발적으로 증가시켰다. 작성자의 감정을 포함하고 있는 텍스트 데이터는 시장 조사, 주가 예측 등 다양한 분야에서 이용할 수 있으며, 이로 인해 긍부정의 이진 분류가 아닌 다중 감정 분석의 필요성 또한 높아지고 있다. 본 논문에서는 딥러닝 기반 감정 분류에 심리학 이론의 기반 감정 모델을 활용한 결합 모델과 단일 모델을 비교한다. 학습을 위해 AI Hub에서 제공하는 데이터와 노래 가사 데이터를 복합적으로 사용하였으며, 결과에서는 대부분의 경우에 결합 모델이 높은 결과를 보였다.

주제어: 기계학습, 감정분석, multi-classification

1. 서론

감정 분석 및 오피니언 마이닝은 텍스트에 나타나는 의견, 감정, 평가, 태도 및 감정을 분석하는 연구 분야이다. 자연어처리에서 최근 가장 활발한 연구 분야 중 하나이며 데이터 마이닝, 웹 마이닝 및 텍스트 마이닝 분야에서도 널리 연구되고 있다[1].

소셜 네트워크 서비스나 무선 인터넷 환경, 모바일 등의 소형 디바이스들이 대중에게 가까이 자리 잡으면서 사람들은 이전보다 자신의 의견을 자유롭게 표현할 수 있게 되었으며, 이로 인하여 감정 분석에 대한 관심과 수요가 높아졌다.

감정 분석에 대한 초기 연구는 긍/부정 극성분석 중심으로 감정 사전에 기반한 방법으로 진행되었다. 감정 사전을 이용한 방식은 어휘별 감정을 정의한 감정 사전을 이용하여 문장 어휘의 극성과 출현 빈도에 따라 감정을 분류한다. 그러나 감정 어휘는 특정 도메인에 따라 감정의 종류나 정도가 달라질 수 있으므로, 정확한 감정분석을 수행하기 위해서는 해당 도메인과 적합한 감정 사전을 구축해야 하는 문제가 있다[2][3]. 이러한 문제를 해결하기 위해 최근의 딥러닝에 기반한 방식에서는 RNN, CNN 등 다양한 방식을 이용하고 있으며[4], 최근에는 BERT[5]를 활용하여 도메인에 종속적이지 않은 다중 감정분석에서도 성과를 얻고 있다[6~8].

딥러닝에 기반한 감정 분석에서는 감정이 태그된 학습 데이터를 필요로 하며, 감정 태그 분류는 다중 감정 분석에 큰 영향을 미친다. 대표적인 감정 모델인 러셀의 핵심 정서(Russell core affect)[9]는 그림 1과 같이 심리 모델에서는 유쾌(Pleasure)/불쾌(DisPleasure)와 격함(Activation)/차분함(DeActivation)을 기반 감정(base sentiment)로 보고 세로와 가로축에 교차하여 배치하고, 이의 조합으로 감정을 8개의 분류로 나눈다.

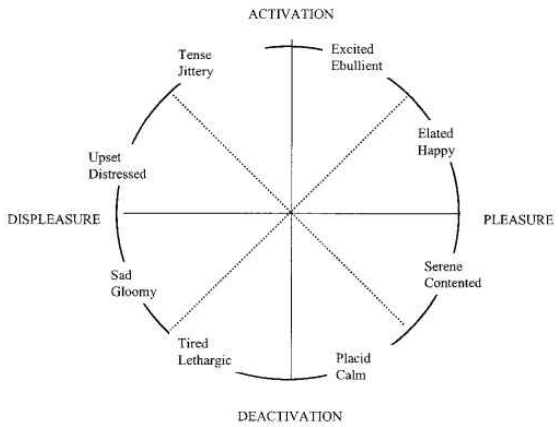

그림 1. 러셀의 핵심 정서
(Russell core affect)

본 논문에서는 이러한 심리 모델을 적용하여 기반 감정의 조합으로 다중 감정을 결정하는 결합 모델과 각 감정을 별도의 감정으로 분별하는 단일 모델을 구축하여 그 결과를 비교하고자 한다. 논문에서는 러셀의 핵심 정서의 여덟 개의 감정 분류를 단순화시켜 표 1의 네 분류로 축소하여 그 결과를 분석한다. 2-class 결합 모델에서는 기반 감정이 격함/차분함과 유쾌/불쾌 중 무엇인지 판별하고, 그 조합으로 최종 감정을 결정한다. 예를 들어 기반 감정이 격함 + 유쾌로 판별되는 경우 최종 감정

표 1. 4가지 감정 분류 기준

	Activation(격함)	DeActivation(차분함)
Pleasure(유쾌)	기쁨	평온
DisPleasure(불쾌)	분노	슬픔

은 '기쁨'이 된다. 4-class 단일 모델에서는 기쁨/평온/분노/슬픔을 별개의 감정으로 보고 감정을 분류한다.

2. 다중 감정 분류 모델

2.1 학습데이터 구축

BERT를 기반으로 감정 분석을 수행하기 위해서는 감정이 태그된 데이터가 필요하다. AI Hub의 한국어 감정 정보가 포함된 단발성 대화 데이터[10]는 공포, 분노, 슬픔, 혐오, 행복, 놀람, 중립의 7개의 감정 분류를 레이블링한 38,594문장을 제공한다. 본 논문에서는 AI Hub 데이터에 기반하여 다중 감정 분류를 위한 학습 데이터를 구축한다.

2.1.1 4-class 단일 모델 데이터 구축

본 논문의 4-class 단일 모델은 표 1의 기쁨, 평온, 분노, 슬픔의 네 감정을 하나의 모델로 분류한다. 데이터 구축에서는 AI Hub 데이터의 일곱 감정 분류 중 행복으로 태그된 6000문장을 본 모델의 기쁨에, 5600건의 분노와 5200건의 슬픔으로 태그된 문장을 각각 본 모델의 분노와 슬픔의 데이터로 추출하였다. AI Hub 데이터에서는 평온에 대응되는 태그가 존재하지 않아 중립으로 태그된 4800문장을 평온으로 간주하여 추출하였다. 추출된 각 감정 데이터 중 1,000건은 실험 집합으로 사용하고, 나머지 데이터로 학습을 진행하였다.

2.1.2 2-class 결합 모델 데이터 구축

2-class 결합 모델은 유쾌/불쾌 분류모델과 격함/차분함 분류모델을 결합하여 구성한다. 기존 극성 분류의 긍정과 부정에 해당되는 유쾌/불쾌 분류의 예측은 격함/차분함 분류에 비해 감정 분류 결과에 큰 영향을 미친다. 본 논문에서는 격함/차분함 분류에서는 작은 크기의 AI Hub 데이터를 그대로 사용하고, 유쾌/불쾌의 분류에서는 AI Hub 데이터와 함께 대량의 데이터를 추가적으로 활용하는 경우의 성능을 비교하고자 한다. 대량의 데이터로는 감정 어휘가 풍부하게 발생하는 노래 가사 데이터를 사용한다.

아래 문장은 격함/차분함 분류에 사용되는 단발성 문장 데이터의 예시이다. AI Hub 데이터는 격함/차분함의 형태로 태그되지 않아, 태그된 감정 중 행복/평온/분노/슬픔에 대하여 아래 예시와 같이 표 1의 태그를 부착하여 학습에 사용한다.

- 쉽게 구할 수 있는 재료로 특별한 맛을 낼 수 있게 도와주시는 OOO님 늘 감사해요 [**행복 → 유쾌 + 차분함**]
- 정말 한결같고 애교도많고 너무 큰사랑 줬던 사람이라 잊기가 힘들어요 [**슬픔 → 불쾌 + 차분함**]
- 역시 OOO!!! 멋짐!!! [**놀람 → 유쾌 + 격함**]
- 소비자를 우롱하는 OO은 나가라!! [**혐오 → 불쾌 + 격함**]

AI Hub 데이터의 7가지 감정 분류를 러셀의 핵심 정서의 기반 감정에 적용하면 불쾌=[공포, 분노, 슬픔, 혐오], 유쾌=[행복], 중립=[중립]으로 나눌 수 있다. AI Hub의 데이터가 부정 감정을 세분류하고 있어, 불쾌 약 21,000문장, 유쾌 약 4,600문장, 중립 약 3,800문장이 학습 데이터로 얻어진다. 유쾌나 중립보다 불쾌에 대응되는 데이터가 5배의 수준을 보이며, 이러한 데이터 불균형은 성능 하락으로 이어지기 쉽다. 이를 해결하기 위해 본 논문에서는 유쾌/불쾌 모델을 두 단계로 나누어 구성한다. 유쾌/불쾌/중립을 분류하는 3-class 모델에서 불쾌일 확률이 50% 미만인 경우에 유쾌/중립 데이터에 대해 정확한 분류를 수행하기 위한 유쾌/중립의 2-class 모델을 추가로 구축하여 사용한다.

AI Hub 데이터는 SNS와 온라인 댓글 중심으로 구성되어 있어 감정 표현에 한정적일 수 있다. 본 논문에서는 유쾌/불쾌에 대한 정확한 판별을 위해 지니 뮤직[11]의 가사 데이터 약 104만 건을 수집하여 사용한다. 노래에는 아래 예제처럼 사용자가 태그한 감정 단어가 부착되어 있으나, 가사의 감정보다는 노래의 리듬이나 음정, 박자의 영향으로 실제 문장 감정과 일치하지 않은 경우가 많다. 이에 앞서 소개한 AI Hub 데이터로 구성한 유쾌/불쾌 분별 모델로 노래 가사를 감정 분류한 결과를 학습 데이터로 사용하고, AI Hub 데이터를 사용한 결과와 자동 태그된 노래 가사를 추가로 사용한 결과를 비교 평가하여 노이즈가 포함된 대량의 데이터가 감정 분류에 기여할 수 있는지 살펴본다.

- 찬란하게 빛나고 있으면 [외로움]
- 다 쓴 누가 버리고 간 침대를 보며 [설렘]

2.2. BERT기반 문장 분류

본 논문에서는 BERT[5]를 활용해 만들어진 KoBERT[12] 모델을 사용한다. 각 분류 모델은 Linear Layer를 중첩한 MLP(Multi-Layer Perceptron)를 기본 KoBERT 모델에 적재하여 fine-tuning을 진행했다. 이후 심리 모델을 기반으로 한 2-class 결합 모델을 구축하기 위해 유쾌/불쾌 분류 모델과 격함/차분함 분류 모델을 병렬로 결합하여 감정을 분석하는 시스템을 구축해 4-class 단일 모델과 비교한다. 2-class 결합 모델은 입력으로 문장이 들어오면, 해당 문장을 유쾌/불쾌 분류 모델과 격함/차분함 분류 모델의 입력으로 하여 각각 분류를 수행한다. 이후 각 모델의 분류 결과를 통해 기쁨(유쾌+격렬), 평온(유쾌+차분), 분노(불쾌+격렬), 슬픔(불쾌+차분)의 감정으로 예측한다. 4-class 단일 모델의 경우 입력문장을 바로 4가지의 감정 중 하나의 감정으로 예측한다.

3. 실험 및 평가

2.1.1에서 언급한 바와 같이, 테스트 데이터는 AI Hub 데이터를 기쁨, 평온, 분노, 슬픔으로 변환한 감정별 천 문장, 총 4,000문장으로 구성하였다.

심리모델에 기반한 2-class 결합 모델의 성능을 평가하기 위해, 기반 감정 분류 모델의 정확도를 평가하였다. 결과는 표 2와 같다.

표 2 기반 감정 분류 정확도

유쾌/불쾌 이진 분류	격함/차분 이진 분류
76.23%	73.60%

심리 모델에 기반한 2-class 결합 모델의 결과를 조합하면 최종 감정을 얻을 수 있다. 표 3은 4-class 단일 모델과 2-class 결합 모델 각각의 정확도를 보인다. 결

합 모델에서는 유쾌/불쾌 분류에 AI Hub 데이터만 사용한 경우와 자동 태그된 노래 가사 데이터를 추가로 사용한 결과의 비교하여 보인다.

표 3 각 모델의 감정별 재현율 및 정확률

recall/precision	기쁨	평온	분노	슬픔
4-class 단일 모델	82.50% /84.96%	62.30% /63.44%	72.90% /76.02%	87.60% /80.51%
2-class 결합 모델 (AI Hub)	84.10% /63.93%	**5.10%** **/21.89%**	69.70% /54.24%	90.10% /77.20%
2-class 결합 모델 (AI Hub + 가사)	81.60% /72.86%	**3.70%** **/20.90%**	81.00% /54.73%	92.70% /75.80%

결과에서는 심리모델에 기반한 2-class 결합 모델이 전반적으로 4-class 단일 모델보다 낮은 재현률과 정확률을 보여, 딥러닝에서 일반적으로 사용하는 태깅된 감정에 기반한 분류가 심리 모델을 사용한 모델보다 좋은 결과를 내는 것으로 나타났다. 결합 모델에서는 노이즈가 포함된 자동 태깅된 노래 가사를 추가로 사용한 경우가 AI Hub만 사용한 모델보다 평균적으로 높은 정확률과 재현율을 보여, 대량의 데이터가 다중 감정 분류에 기여할 수 있음을 알 수 있었다. 감정별 분류에서는 기쁨과 분노, 슬픔에서는 결합 모델이 우수한 성능을 보였으나, 평온에서는 크게 성능이 하락하는 결과를 얻었다.

해당 실험 결과를 분석하기 위해 러셀 핵심 정서의 기반 감정인 유쾌/불쾌, 격함/차분함에 대한 재현율과 정확률을 측정하였다. 표 4의 결과에서는, 4-class 단일 방식은 전반적으로 고른 재현율을 보이는 반면에, 심리 모델에 기반한 2-class 결합모델은 차분함과 유쾌의 경우 낮은 재현율을 보였다. 표 3에서 평온의 감정의 경우 표 4의 차분함과 유쾌의 조합으로 만들어지는 감정인데 유쾌와 차분함의 재현율이 낮아 해당 감정분류의 정확도가 낮은 것으로 분석되었다. 이는 AI Hub의 중립의 비감정 태그인 중립을 평온에 강제 연관시켜 발생한 것으로 분석되어, 감정 분류의 올바른 정의가 다중 감정 분류에 미치는 영향을 짐작할 수 있었다.

표 4. 각 모델의 기반 감정별 재현율 및 정확률

recall/precision	유쾌	불쾌	격함	차분함
4-class 단일 모델	83.95% /85.97%	86.30% /84.32%	79.15% /82.02%	82.65% /79.85%
2-class 결합 모델 (AI Hub)	**63.75%** /82.35%	86.35% /70.43%	88.60% /68.15%	**58.60%** /83.71%
2-class 결합 모델 (AI Hub + 가사)	**58.65%** /90.44%	93.80% /69.40%	88.60% /68.15%	**58.60%** /83.71%

4. 결론

본 논문은 심리학에서 제시한 기반 감정에 기초해 구현한 감정 분류 모델과 단일 딥러닝 모델을 비교하였다. 결과에서는 각 감정에 따라 단일 모델과 결합 모델이 다른 성능을 보였으며, 정확도를 높이기 위해 추가로 사용한 노래 가사 데이터로 성능을 향상시킬 수 있었다.

참고문헌

[1] Bing Liu, "Sentiment analysis and opinion mining", Synthesis Lectures on Human Language Technologies, vol 5, No. 1, pp. 1-167, 2012

[2] 김승우, 김남규, “오피니언 분류의 감성사전 활용 효과에 대한 연구”, 석사, 국민대학교 비즈니스IT 전문대학원, 2014

[3] 박상민, 나철원, 최민성, 이다희, 온병원, “Bi-LSTM 기반의 한국어 감성사전 구축 방안.” 지능정보연구, vol 24, no 4, pp. 219-240, 2018

[4] 홍성준, "Recurrent - Convolutional Neural Network를 이용한 한국어 구어체 문장의 감정 분류." 석사, 경북대학교 대학원, 2017

[5] Jacov Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, Proceedings of NAACL-HLT 2019, pp. 4171-4186, 2019.

[6] Manish Munikar, Sushil Shakya, Aakash Shrestha, "Fine-grained Sentiment Classification using BERT", 2019 Artificial Intelligence for Transforming Business and Society (AITB), pp.1 - 5, 2019

[7] 박서희, "BERT Transfer Learning 을 활용한 스토리 텍스트 감정 인식", 석사, 성균관대학교 일반대학원, 2020.

[8] 이가은, "BERT 기반 한국어 감정 사전을 이용한 감정 예측기 개발.", 석사, 서강대학교 정보통신대학원, 2020.

[9] Russell, J. A, “Core affect and the psychological construction of emotion”, Psychological Review, vol 110, No 1, pp.145-172, 2003

[10] "AI Hub 한국어 감정 정보가 포함된 단발성 대화 데이터셋." (2020년 09월 18일). AI Hub, https://www.aihub.or.kr/keti_data_board/language_intelligence

[11] "지니 뮤직." (2020년 09월 18일). 지니 뮤직, 수정, https://www.genie.co.kr/.

[12] "SKTBrain/KoBERT." (2020년 09월 18일), github-SKTBrain/KoBERT, https://github.com/SKTBrain/KoBERT.