챗봇의 효과적 정서적 지지를 위한 한국어 대화 감정 강도 예측 모델 개발

정세림 [1,+], 노유진 [1,+], 오은석 [1], 김아연 [1], 홍혜진 [2], 이지항 [3,*]
[1] 상명대학교 휴먼지능정보공학과, 학사과정
[2] 상명대학교 지능정보공학과, 석사과정
[3] 상명대학교 휴먼지능정보공학과, 교수
(+: 공동 1 저자, *: 교신저자)

Saelim8992@gmail.com, shdbwls2001@naver.com, ihaom@naver.com, zn122@naver.com, hhz2000@naver.com, jeehang@smu.ac.kr

On the Predictive Model for Emotion Intensity Improving the Efficacy of Emotionally Supportive Chat

Sae-Lim Jeong[1,+], You-Jin Roh[1,+], Eun-Seok Oh[1], A-Yeon Kim[1], Hye-Jin Hong[2], Jee Hang Lee[3,*]
[1]Department of Human-Centered AI, Sangmyung University
[2] Department of AI & Informatics, Sangmyung University
[3]Department of Human-Centered AI, Sangmyung University

요 약

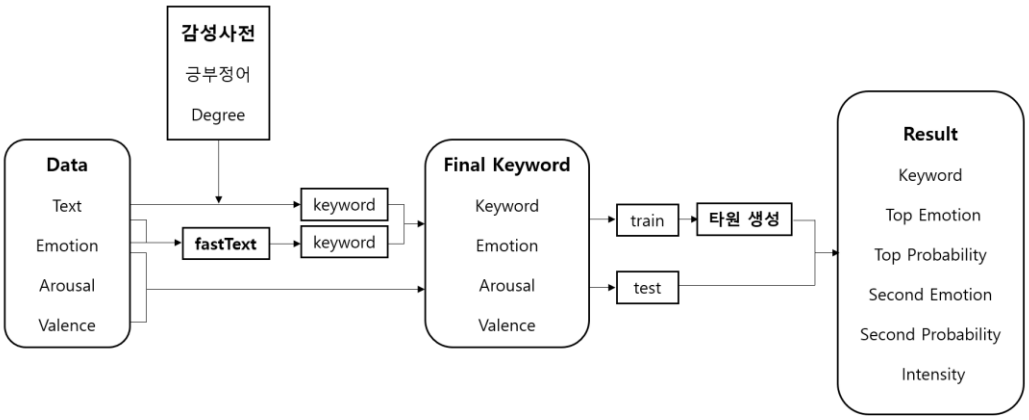
정서적 지원 대화를 위한 챗봇 개발 시, 사용자의 챗봇에 대한 사용성 및 대화 적절성을 높이기 위해서는 사용자 감정에 적합한 지원 콘텐츠를 제공하는 것이 중요하다. 이를 위해, 본 논문은 사용자 입력 텍스트의 감정 강도 예측 모델을 제안하고, 사용자 발화 맞춤형 정서적 지원 대화에 적용하고자 한다. 먼저 입력된 한국어 문장에서 키워드를 추출한 뒤, 이를 각성도 (arousal)과 긍정부정도(valence) 공간에 투영하여 키워드가 내포하는 각성도-긍정부정도에 가장 근접한 감정을 예측하였다. 뿐만 아니라, 입력된 전체 문장에 대한 감정 강도를 추가로 예측하여, 핵심 감정 강도 – 문맥상 감정강도를 모두 추출하였다. 이러한 통섭적 감정 강도 지수들은 사용자 감정에 따른 최적 지원 전략 선택 및 최적 대화 콘텐츠 생성에 공헌할 것으로 기대한다.

1. 서론

정신건강 문제가 사회적 이슈가 되고 있는 가운데, 최근 정신건강의 예방, 치료, 관리에 큰 강점을 보이는 비대면 디지털치료제에 대한 관심이 고조되고 있다 [1, 2]. 생성 모델 기반 챗봇은 사용자의 입력에 따라 적절한 정서적 지원을 동반한 대화를 제공하여 사용자의 감정과 맥락에 훈련 프로그램을 제안하고, 언제 어디서든지 훈련을 진행할 수 있어 등 정신 건강 분야에서 중요한 역할을 차지하고 있다 [3].

다만 정서적 지원 대화의 효과를 높이기 위해서는 사용자의 챗봇에 대한 사용성 및 대화 적절성 고취가 매우 중요하다 [4]. 이에 본 연구는 사용자 입력에 초점을 맞추어 정서적 지원을 할 수 있도록 입력 텍스트의 감정 강도를 추출하여, 사용자의 입력 내에서 감정과 맥락과 같은 다양한 정보를 파악하고 이를 정서적 지원형 대화의 기초로 사용하고자 한다 [5, 6].

이를 위해, 본 논문에서는 챗봇 사용 시 사용자가 입력한 한국어 문장에서 감정과 관련된 키워드를 추출하고, 각성도 (arousal)과 긍정부정도(valence) 공간에 투영하여 키워드가 내포하는 각성도-긍정부정도에 가장 근접한 감정을 예측하고 이를 통해 입력 문장의 감정을 예측하는 시스템을 제안하고자 한다. 이러한 모델은 챗봇과의 대화를 통해 정신 건강 문제를 가진 사람들의 감정 상태를 보다 정확하게 파악하여, 개인화된 정서적 지지형 대화를 제공하고, 사용자의 감정에 알맞은 정서적 지지 전략을 통해 상담의 효과를 제고하는 데 공헌할 것으로 사료된다.



(그림 1) 한국어 문장 감정 분류 및 감정 강도 예측 모델 구조도

2. 한국어 문장 감정 분류 및 감정 강도 예측 모델

본 연구에서 사용한 한국어 문장 감정 분류 및 감정 강도 예측 모델은 그림 1 과 같다. 크게 두 과정으로 구분되는데, 먼저 입력된 한국어 문장에서 감정 관련 키워드를 추출하고, 감성 사전에 근거하여 해당 키워드의 감정 각성도 (이후 arousal)과 감정 긍부정도 (이후 valence)를 추출한다. 해당 arousal-valence 조합은 이후 감정 및 감정 강도 예측에 사용되는데, Russell 의 Circumplex model[12]를 확장한 감정 차원 모델에 사영된 12 개 감정 영역과 arousal-valence 벡터 사이의 거리를 구한다. 이 중 최근접 감정을 감정 예측 값으로 삼고, 12 개 감정 영역들 사이의 거리 벡터를 Softmax 연산하여 감정 강도를 추출하였다.

2-1. 학습데이터

2-1-1. 멀티 모달 영상 데이터셋

본 연구에서는 AI hub 에서 배포 중인 멀티 모달 영상 데이터[7]를 사용하였다. 해당 데이터는 감정인식, 개체인식, 성별/나이 인식, 관계분석, 멀티모달 영상 질의응답, 단일 발화 의도 분석, 복수 발화 의도 분석, 음성인식으로 규정되는 AI 8 종 임무유형을 고려하여 구축 후 공개된 데이터이다. 멀티 모달 영상 데이터는 총 6,000 개의 영상과 그에 상응하는 85,077 개의 발화로 구성되어 있다. 본 논문에서는 해당 데이터에서 발화 스크립트(Text)와 문장에 대한 감정 정보인 감정 종류(Emotion) 과 긍부정 영향도(Valence) 각성 정도 (Arousal)를 사용하였다.

2-1-2. KNU 한국어 감성 사전

한국어 문장 키워드 추출을 위해 KNU 에서 배포하는 KNU 한국어 감성 사전[8]을 사용하였다. 특정 도메인에서 사용되는 긍부정어가 아닌 인간의 보편적인 기본 감정 표현을 나타내는 긍부정어로 구성된다. 국립국어원 표준 국어대사전의 뜻풀이 분석을 통한 긍부정 추출, 국어 감정동사 연구의 긍부정어 목록, SentiWordNet[9] 및 SentiNet-5.0[10]에서 주로 사용되는 긍부정어 번역, 최근 온라인에서 많이 사용되는 축약어 및 긍부정 이모티콘 목록의 과정을 거쳐 구축된 데이터셋이다. 총 14,843 개의 1-gram, 2-gram, 관용구, 문형, 축약어, 이모티콘 등에 대한 긍정, 중립, 부정 판별 및 정도(degree)값으로 구성되어 있다.

2-2. 감정 키워드 추출

입력된 한국어 문장으로부터 감정 관련 키워드 추출을 위해 감성 사전[8]과 fastText[11]를 활용하였다. 먼저, 앞서 설명한 KNU 한국어 감성 사전[8]에서 각 긍부정어의 긍부정 정도에 절대값을 취해 강도 점수를 구하였다. 동시에, 멀티 모달 영상-발화 데이터[7] 각 문장에 대해서는 형태소에서 n-grams 까지 각 요소를 구하였다. 발화 데이터 내 단어가 KNU 한국어 감성사전에 존재하는 경우 해당되는 단어들을 추출하고, 그 중 강도 점수가 가장 높은 단어 하나를 그 문장의 키워드로 선택했다.

그러나, 감정을 유추하기 어려운 단음절 단어들이 키워드로 많이 추출되어, 감성 사전 하나만을 이용하여 키워드를 추출하는 데에는 한계가 있다고 판단하였다. 이를 해결하기 위해 Facebook 에서 제공하는 오픈 소스 라이브러리 fastText[11]를 추가로 이용하였다. 멀티 모달 영상-발화 데이터 문장에 대해 앞서 감성 사전을 활용한 방식과 동일하게 각 요소를 구한 뒤, 해당 문장에 라벨링 되어있는 감정 타입과 가장 유사도가 높은 요소를 그 문장의 키워드로 추출하였다. 최종적으로, 추후 감정 분류에 어려움을 줄 수 있는 한 글자 키워드를 최소화하고자, 감성 사전으로 추출한 키워드가 한 글자인 경우, fastText 를 이용하여 추출한 1-gram 키워드와 2-grams 키워드로 대체해주었다.

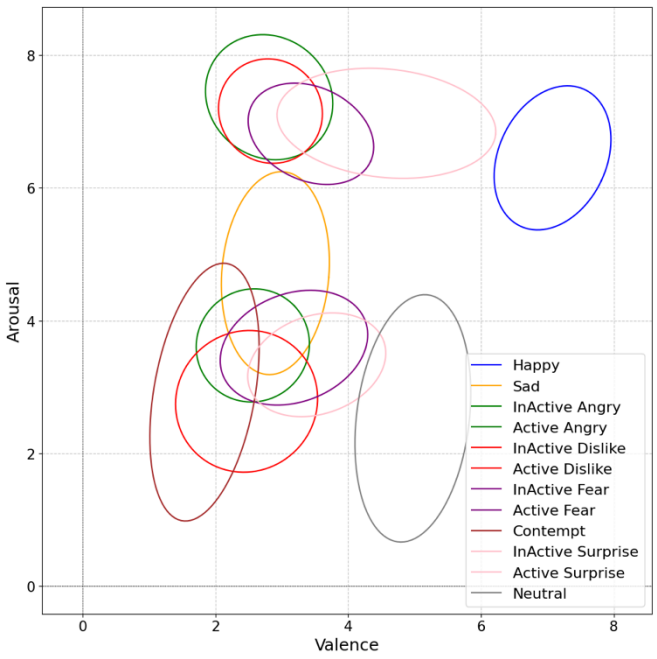
2-3. 감정 차원 모델 생성

공학 분야에서 사용되는 대표적인 감정 차원 모델은 Russell 의 Circumplex model 로, 이는 Valence 와 Arousal 2 개의 축을 고려하여 감정을 표현한다[12].

<표 1> 감정 분류 및 감정 강도 예측 결과

Text	Emotion	Valence	Arousal	Top Emotion	Top probability	Second Emotion	Second probability	Intensity
어머, 나는 안돼. 법에 걸려. 그런거 안하셔도 된다고 말씀 드려.	Surprise	3	3	Dislike	0.494715	Sad	0.041459	3
아빠 나 좋은 소식이 있어	Happy	7	8	Happy	0.089319	Sad	0.083128	1
다음주에 엄마 생일이잖아. 잊은 건 아니지?	Surprise	2	5	Fear	0.098744	Sad	0.081169	1

이 모델은 각 감정을 하나의 점으로만 표현하므로 개인의 차이를 나타내기에 부족하다는 한계가 있어, 최근 연구에서는 각 감정의 영역이 표현 가능한 타원의 형태를 채택했다[13]. 데이터를 정규분포를 따른다고 가정할 경우, 가우시안 분포의 3 차원 곡선 단면이 타원으로 나타나는 바 [13], 멀티 모달 영상-발화 데이터의 감정 정보가 정규분포를 따른다고 가정하였다. 이 때, 한 감정에 대해 타원의 중점을 각 단어의 평균값으로, 회전각 theta 는 상관계수를 통해 구한 값으로 대치하였다. Arousal 의 경우는 개인의 특성에 따라 동일한 감정의 각성 정도를 다르게 판단할 수 있기에[13], 범위를 절반으로 나누어 Arousal 값이 active 인지 inactive 를 구분할 수 있도록 하였다. 이때 감정에 대해 동적 점수와 정적 점수 비율이 0.3 ~ 0.7 에 포함된다면 하나의 감정을 동적감정과 정적감정으로 나누어 각각의 감정 영역을 만들도록 하였다. 그림 2 는 이렇게 생성된 감정 차원 모델을 보여준다.



(그림 2) Russell 의 Circumplex model [13] 확장형 감정 차원 모델

2-4. 감정 분류 프로세스

타원이 생성된 감정 차원에 테스트 데이터로부터 추출된 키워드의 감정 점수 (Valence, Arousal)을 투영하였을 때, 감정 별 타원의 중심 값과 키워드의 감정 점수는 모두 Valence 와 Arousal 로 만든 좌표 값이기 때문에 두 점사이의 거리를 계산할 수 있다. 두 점 사이의 거리의 역수를 취한 후, 이에 Softmax 값을 취한다면 데이터셋에 존재하는 8 가지의 감정 중 어떤 감정일 확률이 높은지에 대한 결과값을 추출할 수 있다.

3. 실험 결과

본 연구에서는 한국어 감정 분류를 위해 감정 각성도에 따라 키워드를 추출하고, 기존 모델의 문제점을 보완한 타원형 감정 차원 모델을 이용하여 감정 분류를 진행하였다. 제안한 시스템을 이용하여 얻은 최종 데이터 셋은 <표 1>과 같다. 최종 감정 분류 결과 8 가지 감정 중 가장 높은 확률을 가진 두 가지 감정인 Top emotion, Second emotion 과 그 확률 값인 Top probability, Second probability 를 기존 테스트 데이터에 추가할 수 있었다. 예측한 Top Emotion 과 기존 라벨링 감정이었던 Emotion 과의 AUC 점수 확인 결과 0.60 으로 확인되었다.

추가적으로, 감정의 강도(Intensity)를 구하기 위한 접근 방법을 고민하였다. 이를 위해, 앞에서 구한 Top probability 와 Second probability 를 사용하여 그 값의 차를 구하고, 이를 10 점 척도로 변환하여 예측된 각 감정에 대한 confidence 를 0~10 의 값으로 표현하였다. 해당 방법은 기존 연구[14]를 바탕으로, 감정의 강도를 수치적으로 정확하게 측정하기보다는 해당 감정을 다른 감정과 상대적으로 비교하여 상대적인 강도를 파악하는 접근을 따라 진행하였다. 이에 따라, 계산된 confidence 값을 각 감정의 강도로 볼 수 있다고 판단하여, 실수로 표현된 confidence 값을 1~5 의 정수 값으로 구간화 하고 감정 intensity 로 사용하였다.

마지막으로 우리가 활용한 방식과 기존 감정 분류 모델의 성능을 비교해보았다. 대표적 감정 분류 모델로 알려진 KoBERT[15]를 활용하여 감정 다중 분류를 진행하였다. 예측 값에 대한 비교는 앞서 설명한 방식과 같이 예측한 Top Emotion 과 기존 라벨링 감정

이었던 Emotion 을 비교하였고, 그 결과 AUC 점수가 0.667 로 확인되었다. 성능 비교 결과, 기존 KoBERT[15]를 활용한 방식이 타원을 활용한 방식보다 미세하게 우수한 성능을 보였지만, 현재 데이터셋에는 감정 종류가 대체로 부정적 감정이기에 그림 2 와 같이 타원들의 영역이 겹치는 경우가 많았다. 따라서 향후 긍정 감정의 태그에 대한 데이터를 추가하여 균형적인 데이터를 구축한다면, 충분히 개선의 여지가 있다고 판단된다. 또한, KoBERT[15]를 활용하여 앞서 시도한 방식과 동일하게 각 감정에 대한 Intensity 를 측정하였다. 다만, 본 연구에서는 Intensity 에 대한 새로운 접근을 사용하였기 때문에 정량적인 성능 비교는 어려웠다. 그러나 전반적으로 KoBERT[15]의 경우에는 더 강한 확신을 갖는 점수가 더 자주 나타나는 반면, 본 연구에서 제안된 모델의 결과에서는 두 감정 사이의 확률 값이 모호한 부분이 다수 발견된다는 한계점이 존재하였다.

4. 결론 및 향후 연구

본 연구에서는 한국어 감정 분류를 위한 키워드 추출과 감정 차원 모델을 생성하였고 만들어진 감정 차원 내 개별 감정 영역과의 거리를 이용한 감정 분류 시스템을 제시하였다. 분류 결과, 한국어 문장 키워드에 대한 감정 분류가 테스트 데이터에 추가되었음을 확인할 수 있었다. 단순히 문장내에 중요도가 높은 키워드를 추출하는 것이 아닌 각성 정도가 높은 단어를 추출함으로써 문장내 감정 영향도가 높은 단어를 찾아낼 수 있었다. 또한, 본 논문에서는 예측된 감정의 Confidence 를 활용하여 감정에 대한 상대적인 강도, 즉, Intensity 를 도출해낼 수 있다고 판단하여 예측된 감정에 대한 강도(Intensity)를 새롭게 제안하였다. 완성된 테스트 데이터셋을 통해 확인한 결과, Intensity 값이 Arousal 값에 비해 크기가 작은 경우가 존재함을 확인할 수 있었다. 이는 상위 두 감정의 확률값 차이가 미미하다는 것을 의미하며 결론적으로 두 감정의 영향도를 세밀하게 구분하기 어렵다는 점을 확인할 수 있다. 또한, 기존 감정분류 모델인 KoBERT[15]에 비해 모호한 결과값을 다수 보인다는 것을 확인할 수 있었다. 본 연구에서 제안하는 시스템은 Intensity 에 대한 예측 시스템이 아니었기 때문에 향후 이를 중점으로 시스템을 보완해 나간다면 보다 더 안정적인 결과를 가지는 시스템 구축이 가능할 것을 기대한다.

추후에는 보다 정확한 키워드 추출을 위해 더 많은 단어가 담긴 감정 사전을 구축하여 키워드를 추출하고 보다 안정적인 타원 모델을 구현하기 위해 추출한 감정 단어들을 그룹화 하는 연구를 진행할 예정이다. 그리고 최종적으로는 본 연구가 사용자의 감정, 감정 각성 정도, 감정 강도를 보다 정확하게 파악하여 정서적 지원 대화를 제공하는 챗봇 개발에 공헌할 것으로 사료된다.

Acknowledgement

이 논문은 2023 년도 정부재원(과학기술정보통신부 여대학원생공학연구팀제 지원사업; 협약번호 WISET-2023-126 호)으로 과학기술정보통신부와 한국여성과학기술인육성재단의 지원 및 과학치안진흥센터 (인공지능과 클라우드를 활용한 아동 목격자 맞춤형 비대면 진술조서 지원 시스템 개발; RS-2023-00281194)의 지원을 받아 수행되었음.

참고문헌

[1] 강한승 외, 디지털 헬스 산업의 글로벌 동향과 전략적 발전방안에 대한 연구. *e-비즈니스연구*, *23*(5), 227-241, 2022.

[2] Bucci, S. *et al.*, The digital revolution and its impact on mental health care. *Psychology and Psychotherapy: Theory, Research and Practice*, 92(2), 277-297, 2019.

[3] Cameron, G. *et al.*, Towards a chatbot for digital counselling. *In Proceedings of the 31st International BCS Human Computer Interaction Conference*, 2017.

[4] 전용호. 노인 돌봄의 연속성 측면에서 바라본 의료· 보건· 복지 서비스의 이용과 연계. *보건사회연구*, 38(4), 10-39. 2018.

[5] Saima, A. *et al.*, Identifying expressions of emotion in text. *International Conference on Text, Speech and Dialogue*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[6] Nandwani, P. *et al.*, A review on sentiment analysis and emotion detection from text. *Social Network Analysis and Mining*, 11(1), 81, 2021.

[7] AI hub 멀티모달 영상 데이터, https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=115&topMenu=100&aihubDataSe=realm&dataSetSn=58 (Retrieved 20230927)

[8] KNU 한국어 감성사전, http://dilab.kunsan.ac.kr/knusl.html (Retrieved 20230927)

[9] SentiWordNet_3.0.0_20130122, http://sentiwordnet.isti.cnr.it/ (Retrieved 20230927)

[10] SenticNet-5.0, http://sentic.net/ (Retrieved 20230927)

[11] fastText, https://fasttext.cc/ (Retrieved 20230927)

[12] Russell, J. *et al.*, Circumplex Model of Affect, *Journal of Personality and Social Psychology*, 39(6), 1161-1178, 1980.

[13] 한의환 외, 모델의 확장을 통한 감정차원 모델링 방법 연구, *감성과학회*, 제 20 권, 1 호, 2017.

[14] Troiano, E. *et al.*, Emotion Ratings: How Intensity, Annotation Confidence and Agreements are Entangled, *In Proceeding of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, Online, 40-49, 2021.

[15] KoBERT, https://github.com/SKTBrain/KoBERT (Retrieved 20230927)