

# Russell 모델 기반 문장 감정 분류와 강도 정량화: BERT 임베딩과 수식어 분석을 활용한 연구

최다녕, 정예찬, 정치윤\*, 김무섭\*, 김채규\*\*

국립부경대학교, \*한국전자통신연구원 감각확장연구실, \*\*국립부경대학교 인공지능 연구소

choidaday@gmail.com

## Emotion Classification and Intensity Quantification of Sentences based on the Russell Model: A Study Using BERT Embeddings and Modifier Analysis

Choi Da Nyeong, Jung Ye Chan, Jeong Chi Yoon\*, Kim Moo Seop\*, Kim Chae Kyu\*\*

Pukyong National University., \*Digital Convergence Research Lab., Etri., \*\*Artificial Intelligence Laboratory, Pukyong National University

### 요약

본 연구는 Russell의 감정 원형 모델을 기반으로 문장의 감정 강도를 정량적으로 측정하고 분류하는 방법론을 제안한다. BERT 임베딩을 활용하여 문장의 valence와 arousal 값을 예측하고, 이를 바탕으로 감정 강도를 계산하는 혼합식을 설계하였다. 혼합식은 valence 중심 편차, 감정 공간 중심점으로부터의 유clidean 거리, 강조 수식어의 존재 여부에 따른 수식어 점수를 조합하여 구성되었으며, 유전 알고리즘을 통해 최적화되었다. 최종적으로 가우시안 혼합 모델을 활용하여 감정 강도를 5단계로 분류하였다. 실험 결과, 제안된 방법론은 텍스트에서 표현된 감정의 강도를 효과적으로 측정하고 의미 있는 단계로 분류할 수 있음을 확인하였다.

### I. 서론

텍스트에서 감정을 인식하고 분석하는 것은 자연어 처리(NLP) 분야의 중요한 연구 주제로 자리 잡았다. 특히 감정 분석은 소셜 미디어 모니터링, 고객 피드백 분석, 마케팅 전략 수립 등 다양한 응용 분야에서 활용되고 있다. 그러나 기존의 감정 분석 연구는 주로 긍정, 부정과 같은 감정의 극성(polarity)이나 기쁨, 슬픔, 분노와 같은 개별 감정 카테고리 분류에 초점을 맞추고 있어, 감정의 강도(intensity)를 정량적으로 측정하는 방법론은 여전히 초기 단계에 머물러 있다.

감정 강도는 감정의 존재 여부를 넘어 그 감정이 얼마나 강하게 표현되었는지를 나타내는 중요한 요소이다. 예를 들어, “행복하다”와 “매우 행복하다”는 모두 긍정적인 감정을 표현하지만, 후자가 전자보다 더 강한 감정 강도를 가진다. 이러한 감정 강도의 차이는 사용자의 감정 상태를 더 정확히 이해하는 데 중요한 단서를 제공한다[1].

본 연구에서는 텍스트의 감정 강도를 예측하고 분류하기 위한 새로운 방법론을 제안한다. 구체적으로, BERT[2] 모델을 미세 조정하여 GPT로 생성된 데이터셋의 valence와 arousal 값을 예측하고, 이를 바탕으로 감정 강도를 계산하는 혼합식을 설계하였다. 혼합식의 가중치는 유전 알고리즘을 통해 최적화되었으며, 최종적으로 가우시안 혼합 모델(GMM)[3]을 활용하여 감정 강도를 5단계로 분류했다.

### II. 본론

본 연구에서는 그림 1과 같이 문장의 감정을 Valence와 Arousal 두 축으로 표현하는 Russell의 이차원 감정 모델을 기반으로 감정 분석을 수행하였다. 해당 모델은 감정을 원형 공간에 분포시키는 방식으로, 심리학적으로 감정분류 기준을 제공한다. 이를 위해, 사전 학습된 언어 모델인 BERT 모델을 활용하여 문장 임베딩을 생성하고, 이를 바탕으로 Valence와

Arousal 값을 회귀적으로 예측하는 감정 회귀 모델을 구성하였다. 문장의 시작을 나타내는 특수 토큰인 [CLS] 토큰의 출력 벡터를 활용하였으며, 해당 벡터가 문장의 전체 의미를 함축하고 있다는 특성을 이용하여 감정 상태 예측에 적용하였다.

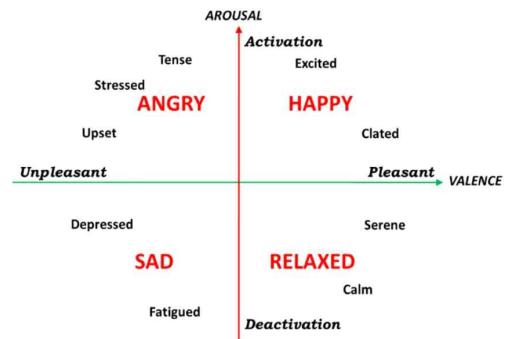


그림 1. 러셀의 원형 모형[4]

학습은 감정 레이블(Valence, Arousal)이 존재하는 Emobank 데이터셋[5]을 활용하여 진행되었으며, 본 연구에서는 두 값을 0~1 범위로 정규화한 뒤 Russell의 원형 감정 모델에 따라 그림 2와 같이 2x2 사분면으로 나누어 감정 범주를 정의하였다.

본 연구에서는 BERT 모델을 통해 예측된 Valence와 Arousal 값을 활용하여 감정 강도를 계산하는 혼합 가중합 식을 정의하였다.

$$I = w_1 \times |V - 0.5| + w_2 \times A + w_3 \times D_{center} + S_{modifier} \quad (1)$$

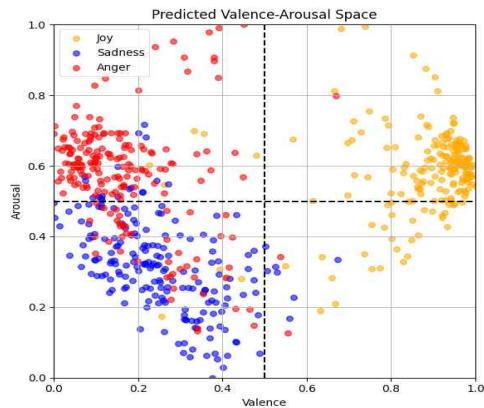


그림 2. 텍스트 데이터의 예측된 러셀 공간

위의 식에서  $I$ 는 문장의 최종 감정 강도를 나타내며, 각 구성 항은 다음과 같이 정의된다.

$V$ 는 감정의 극성,  $A$ 는 감정의 각성도,  $D$ 는 문장 임베딩이 감정 공간 상에서 중립점인  $(0.5, 0.5)$ 로부터 떨어진 정도를 측정하는 유클리드 거리이다.  $S$ 는 문장 내 강조 수식어의 유무 및 강도를 기반으로 산출한 수식어 점수로 해당 표현이 포함될 경우 해당 문장의 감정 정도를 보정하여 증가시키는 역할을 한다.

가중치 항은 각 요소가 감정 강도 산출에 미치는 상대적 기여도를 결정하는 계수로, 본 연구에서는 유전 알고리즘을 활용하여 최적화하였다. 제안된 수식은 심리학적 기반의 감정 구성 요소(Valence, Arousal), 언어적 강조 표현(Modifier) 및 임베딩 기반의 의미적 거리 정보를 통합함으로써, 감정 강도의 다차원적 특성을 정밀하게 반영할 수 있다. 이러한 접근은 기존의 단순한 규칙 기반 또는 극성 중심 강도 측정 기법에 비해 더욱 높은 표현력과 적용 가능성을 가진다.

실험에는 GPT-4.5 기반으로 생성된 Joy, Sadness, Anger 각 200문장 총 600개의 텍스트 데이터를 사용하였으며, 사전 학습된 BERT 모델을 통해 각 문장의 Valence 및 Arousal 값을 회귀적으로 예측하였다. 이후 Valence 중심편차, 감정 공간 중심점  $(0.5, 0.5)$ 로부터의 유클리드 거리, 그리고 강조 부사의 존재 여부에 따라 정의된 수식어 점수를 조합하여 강도 점수를 산출하였다. 강도 식 내 가중치는 실루엣 점수를 적합도로 설정한 유전 알고리즘을 통해 최적화되었으며, 그 결과  $w1 = 0.789$ ,  $w2 = 0.195$ ,  $w3 = 0.016$ 으로 도출되었다.

이후 산출된 강도 점수는 0~1로 정규화된 뒤, Gaussian Mixture Model(GMM)을 이용한 확률 기반 군집화 방식으로 자동 단계화되었다. BIC(Bayesian Information Criterion)를 기준으로 최적 군집 수를 탐색한 결과, 그림 3과 같이 5개의 클러스터가 최적값으로 선택되었으며, 이 방식은 데이터 분포의 밀도 기반 경계를 반영함으로써 사전 임계값 기반 분류보다 더 유연하고 해석 가능한 감정 강도 단계를 제공한다.

### III. 결론

본 연구는 텍스트에 나타난 감정 강도를 정량화하고 분류하는 새로운 방법을 제시했다. BERT 모델을 활용해 예측한 valence와 arousal 값을 기반으로 감정 강도를 산출하는 혼합식을 개발했으며, 실험 결과 감정의 극성(valence)이 강도 계산에 가장 큰 영향을 미치는 것으로 나타났다. 향후 연구 방향으로는 텍스트뿐 아니라 음성이나 얼굴 표정 같은 다양한 정보

를 함께 분석하는 멀티모달 접근법을 도입한다면, 보다 자연스럽고 정확한 감정 인식 시스템을 구현할 수 있을 것이다.

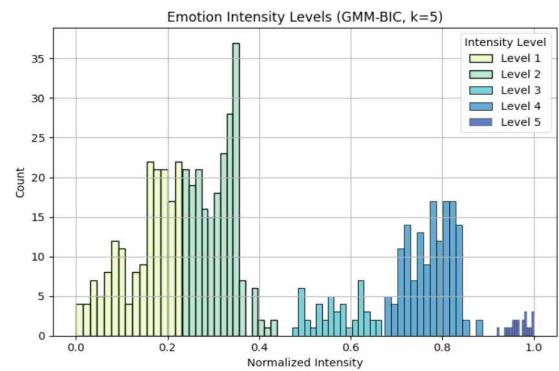


그림 3. 감정 강도 결과 분포

### ACKNOWLEDGMENT

본 연구는 한국전자통신연구원 연구 운영비지원사업(기본사업)의 일환으로 수행되었음[25ZB1200, 인간중심의 자율지능시스템 원천기술연구]

### 참 고 문 헌

- [1] S. M. Mohammad and F. Bravo-Marquez, "WASSA-2017 shared task on emotion intensity," *Sentiment and Social Media Analysis*, pp. 34-49 2017.
- [2] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *Human Language Technologies*, pp. 4171-4186, 2019.
- [3] D. A. Reynolds, "Gaussian mixture models," *Encyclopedia of Biometrics*, pp. 741-744, 2009.
- [4] S. J. Grimm and G. G. Mengel, "An emotion-based method for automated supervised learning of linguistic cues and non-linguistic game events," *Scientific Reports*, vol. 4, no. 4998, 2014.
- [5] S. Buechel and U. Hahn, "EmoBank: Studying the impact of annotation perspective and representation format on dimensional emotion analysis," *European Chapter of the Association for Computational Linguistics*, pp. 578-585, 2017.