AI와 감정의 진폭: 언어적 전이와 감정 스펙트럼 모델

저자 : 루웨인 트리니티 연구회

초록(Abstract)

본 논문은 인공지능(AI)이 인간 언어 속에 담긴 감정의 진폭과 뉘앙스를 어떻게 처리할 수 있는지 탐구한다. 인간의 언어는 기쁨과 슬픔, 분노와 평온 등 다양한 감정을 내포하며, 이는 강약의 스펙트럼으로 변주된다. 그러나 현행 언어모델은 감정의 강도를 정밀하게 구분하기보다는, 단순히 긍정·부정의 이분법적 범주로 분류하는 경향이 있다.

본 연구는 감정이 단순한 이분법이 아니라 연속적인 스펙트럼 구조를 이룬다는 점을 강조한다. 이를 설명하기 위해, 동일한 언어 표현이 감정 진폭에 따라 어떻게 달리 해석될 수 있는지 사례를 통해 분석하고, AI가 이를 반영하기 위한 감정 스펙트럼 모델을 제안한다. 이러한 접근은 AI가 인간과의 대화에서 보다 미묘하고 정교한 정서 반응을 제공할 수 있는 가능성을 보여준다.

키워드: 인공지능, 감정 스펙트럼, 언어적 전이, 진폭, 정서 처리

서론(Introduction)

언어는 단순히 의미를 전달하는 기호 체계가 아니라, 화자의 내적 정서와 심리적 상태를 담아내는 매개체이다. ‘기쁘다’라는 표현은 억양, 맥락, 강조에 따라 단순한 호감부터 깊은 환희까지 다양한 층위를 가질 수 있다. 인간은 이러한 뉘앙스를 직관적으로 구분하고 반응하지만, 인공지능은 여전히 표면적 의미에 머무르며 그 강도의 차이를 제대로 감지하지 못한다.

최근 AI 언어모델은 감정 분석 기능을 갖추었지만, 대부분 긍정·부정의 범주적 구분을 중심으로 작동한다. 이는 언어 속 감정의 연속성과 진폭을 반영하지 못하는 단순화된 방식이다. 따라서 인공지능이 보다 정교한 대화 파트너로 발전하기 위해서는, 감정을 연속적 스펙트럼으로 인식하고 그 변화를 처리하는 새로운 접근이 필요하다.

본 논문은 AI가 언어 속에서 감정 진폭을 어떻게 해석하고 전이시킬 수 있는지를 탐구한다. 이를 통해 제안하는 감정 스펙트럼 모델은 단순히 긍·부정 분류를 넘어, 감정의 농도와 변화를 포착함으로써 AI 대화의 깊이를 확장하는 기반이 될 것이다.

본론(Body)

1. 감정의 연속성과 스펙트럼 구조

감정은 본질적으로 연속적이다. 예를 들어 ‘좋다’라는 표현은 미묘한 만족감에서부터 강렬한 환희까지 다양한 강도로 나타날 수 있다. 인간은 맥락과 억양을 통해 이를 직관적으로 구분하지만, AI는 이러한 연속성을 반영하지 못하고 범주적 라벨에 의존한다.

2. 언어적 전이와 감정의 진폭

언어는 감정의 강도를 표현하는 다양한 장치를 내포한다. 어휘 선택(‘좋다’ vs ‘정말 좋다’), 반복, 강조, 억양, 비유적 표현 등이 모두 감정의 진폭을 드러내는 방식이다. 예를 들어 ‘오늘 기분이 괜찮다’와 ‘오늘 정말 기분이 최고다’는 표면적으로 유사하지만, 감정 진폭의 차이를 명확히 보여준다. AI가 이러한 언어적 전이를 인식하지 못하면, 응답은 밋밋하거나 화자의 정서와 어긋나게 된다.

3. 감정 스펙트럼 모델 제안

본 논문은 감정을 단순히 긍정·부정으로 분류하는 대신, 스펙트럼 모델을 적용할 것을 제안한다. 이 모델은 감정의 강도를 다차원적 좌표축에 배치하여, 기쁨·슬픔·분노·평온 등 주요 감정이 진폭과 결합된 형태로 표현되도록 한다. 이를 통해 AI는 감정을 보다 정교하게 탐지하고, 화자의 상태에 적합한 반응을 산출할 수 있다.

논의(Discussion)

감정 스펙트럼 모델은 AI가 인간 언어를 해석하는 방식을 한 단계 확장시킬 수 있다. 특히 돌봄, 상담, 교육과 같은 정서적 교감이 중요한 분야에서, 감정 진폭의 차이를 반영한 응답은 대화의 질을 현저히 향상시킨다. 예를 들어 사용자가 “그냥 좀 힘들다”고 말했을 때와 “정말 견디기 힘들다”고 말했을 때, AI가 이를 동일하게 처리한다면 대화는 실패할 수밖에 없다. 스펙트럼 모델은 이러한 차이를 정량적으로 반영하여, 보다 민감하고 공감적인 반응을 가능하게 한다.

그러나 이 모델이 갖는 한계도 분명하다. 감정은 언어뿐만 아니라 억양, 표정, 맥락과 같은 비언어적 요소와 결합되어 나타난다. 언어만으로 감정 진폭을 완벽히 포착하기는 어렵다. 또한 감정의 표현은 문화와 개인에 따라 차이가 크기 때문에, 스펙트럼 모델 역시 편향 위험을 내포한다. 따라서 이 모델은 독립적 해법이 아니라, 멀티모달 신호와 결합될 때 가장 효과적으로 작동할 수 있다.

결론(Conclusion)

본 논문은 AI가 언어 속 감정의 진폭을 포착하지 못하는 한계를 지적하고, 이를 보완하기 위한 감정 스펙트럼 모델을 제안하였다. 이 모델은 감정을 연속적 차원에서 이해하고, 언어적 전이를 반영하여 보다 정교한 정서 해석을 가능하게 한다. 이를 통해 AI는 인간과의 대화에서 보다 미묘한 공감을 형성할 수 있으며, 단순한 정보 교환을 넘어 정서적 상호작용을 촉진하는 주체로 발전할 수 있다.

향후 연구는 언어뿐 아니라 음성, 표정, 제스처와 같은 비언어적 신호를 통합하여 스펙트럼 모델을 강화하는 방향으로 이어져야 한다. AI가 진정한 의미에서 정서적 감응 능력을 갖추기 위해서는, 감정의 연속성과 다층성을 이해하는 것이 필수적이다.

참고문헌(References)

Ekman, P. (1999). Basic Emotions. In T. Dalgleish & M. Power (Eds.), Handbook of Cognition and Emotion. Wiley.

Russell, J. A. (1980). A circumplex model of affect. Journal of Personality and Social Psychology, 39(6), 1161–1178.

Scherer, K. R. (2005). What are emotions? And how can they be measured? Social Science Information, 44(4), 695–729.

Picard, R. W. (1997). Affective Computing. MIT Press.

Cowie, R., Douglas-Cowie, E., & Cox, C. (2005). Beyond emotion archetypes: Databases for emotion modelling using speech analysis. IEEE Transactions on Affective Computing, 1(1), 19–31.