

감정 분석 알고리즘에 관한 연구

동덕여자대학교 컴퓨터학과

김지민, 유승연, 이어진, 정채원

요약

해당 프로젝트 주제는 AR 여행 기록 다이어리로 사용자 다이어리 내용에 감정 분석을 적용해 이모지를 달력에 표시한다.

네이버 클라우드 Sentiment와 ChatGPT를 활용하며, 다양한 감정 범주를 고려한 정확도 향상을 추구하고 입력 데이터에 따라 긍/부정 분석을 수행한다. 중립과 놀람은 모호함으로 분리하여 감정 분석을 실시한다.

1. 서론

현대 사회에서는 디지털 기술의 급격한 발전으로 인해 여행 기록과 공유의 형태가 크게 변화하고 있다. 특히 MZ 세대를 중심으로 SNS를 통해 자신의 일상과 여행 경험을 기록하고 공유하는 새로운 트렌드를 만들어내고 있다.

그러나 페이스북, 인스타그램 등의 기존의 플랫폼에서는 기록한 내용을 바탕으로 감정을 정량화하거나 분석하는 것은 불가능하다는 한계를 갖고 있다.

본 연구는 이러한 제약을 극복하고자 AR 기반 여행 기록 다이어리를 바탕으로 네이버 클라우드 Sentiment와 ChatGPT를 활용한 감정분석 결과를 제공한다. 핵심이 되는 네이버 클라우드 Sentiment API의 흐름은 다음과 같다.

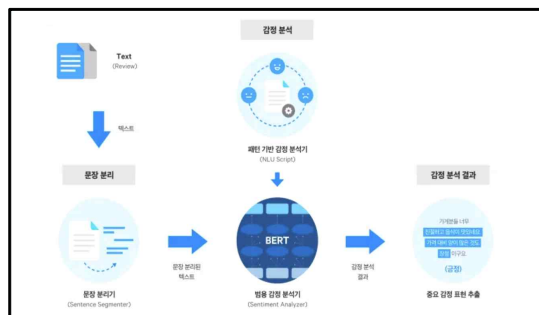


사진 1. 네이버 클라우드 Sentiment API 흐름도

Sentiment를 활용한 감정 분석의 특징은 다양한 감정 범주를 고려한 접근 방식이다. 단순히 긍정과 부정만을 분류하는 것이 아니라, 각 분류층에서 분류하는 라벨의 수를 감소시키는 방법을 활용하여 정확도를 향상시켰다.

예를 들어, 입력된 데이터가 부정적인 경우, 슬픔, 공포, 분노, 놀람, 중립 등 다양한 감정을 분류하는 분류층을 활용하여 감정을 분석하며, 긍정적인 경우에는 행복, 놀람, 중립 등의 감정에 대한 분석을 수행한다.

2. 본론

하나의 분류층에서 수행하는 것보다 각 분류층에서 분류하는 라벨의 수를 감소시킴으로써 정확도를 증가시키기 위해 긍/부정을 분류하는 감성 분석을 먼저 수행한다. 만약 입력한 데이터가 부정일 경우 슬픔, 공포, 분노, 놀람, 중립에 대해 학습시킨 분류층을 통해 감정을 분석하고, 긍정일 경우에는 행복, 놀람, 중립의 감정에 대해 분석하도록 구성한다.

이때, 놀람과 중립의 경우에는 '모호함'으로 분류되어 긍/부정 모두에 포함될 수 있으므로 긍/부정 범주의 감정에서 모두 분석을 할 수 있도록 분리한다. 이를 바탕으로 입력된 문장에 대해 감성 분석을 진행한 후, 해당 감성에 대한 감정 분석을 하도록 모델 생성 후 학습을 수행한다.



사진 2. AI 기반 감정 분석 모델

실제로 API 요청을 보냈을 때 다음과 같은 응답값을 받을 수 있으며 우리는 이 응답값을 GPT에게 넘겨 정의해둔 기준에 따라 happy, smile, not bad, sad, upset 등 어떤 감정에 부합하는지 최종적인 결과를 반환받는다.

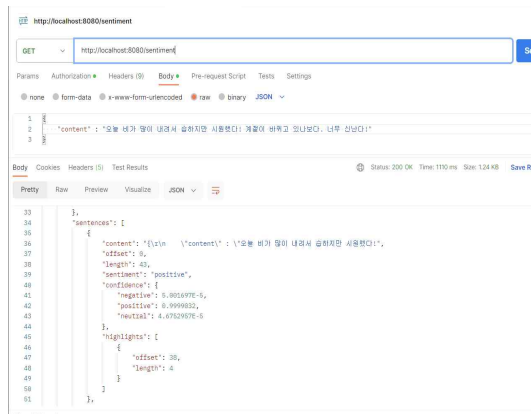


사진 3. 네이버 클라우드 Sentiment 테스트 응답값

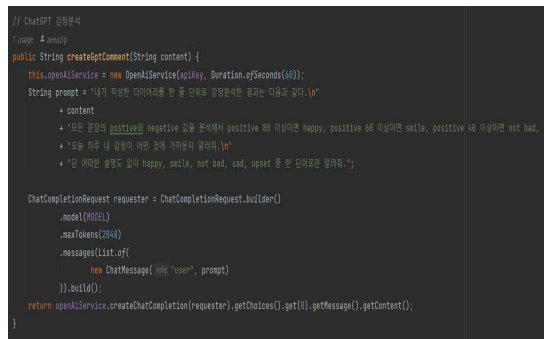


사진 4. ChatGPT를 활용한 최종 분석 핵심 코드

3. 결론

데이터 확장(data expansion) 및 전이 학습(transfer learning) 기술을 활용하여 효율적으로 신규 도메인을 확장할 수 있도록 설계해 범용성을 확보했다.

사용자들이 사용하는 단어와 표현들은 각 도메인마다 매우 차이가 크다. 예를 들면, “쇼핑 리뷰”에서는 “배송, 가격, 품질” 등의 쇼핑 관련 표현이, “뉴스 댓글”에서는 “집값, 정치, 임대” 등의 정치 관련된 표현이 자주 나온다.

따라서 “뉴스 댓글” 도메인만을 지원하는 감정 분석 모델을 그대로 “쇼핑 리뷰” 서비스에 적용하게 되면 만족스러운 품질이 나오지 않을 가능성이 매우 커진다.

그러나 신규 도메인을 확장하면 대량의 데이터를 구축하는 것은 많은 시간과 비용이 소모된다. 이를 해결하기 위해 소규모 데이터를 이용하여 효율적으로 신규 도메인을 확장하는 방법이 있다.

학습 데이터가 소규모인 신규 도메인 데이터를 확장하기 위해, 전이 학습의 특성을 활용하여 전이학습의 공통점은 유지하면서 신규 도메인만의 특색 있는 데이터를 자동으로 추출 및 확장하도록 설계했다. 즉 데이터 확장의 효과로 기존의 도메인들의 품질이 유지되거나 오히려 향상되는 효과를 얻을 수 있다.

참고문헌

- [1] 이수안, “인공지능-07.자연어 처리”, <http://suanlab.com/assets/lectures/ai/07.pdf>
- [2] NAVER CLOUD PLATFORM, “텍스트 감정 분석 서비스 구현하기”, <https://medium.com/naver-cloud-platform/이렇게-사용하세요-텍스트-감정-분석-서비스-구현하기-clova-sentiment-활용기-5d9db7b0209b>, 2021.09.10.
- [3] 네이버 검색, “감정을 분석하다 - 네이버 감정 분석 AI”, https://blog.naver.com/naver_search/222080132277, 2020.09.04

※ 본 프로젝트는 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.