

사용자 기반 Text Cloud Visualization

금비, 원윤서

부산대학교 정보컴퓨터공학부

지도교수 조준수

연구 동기

- ▶ 방대한 텍스트 데이터를 효과적으로 처리하고 의미 있는 정보를 추출하는 것은 중요한 연구 주제이다
- ▶ 방대한 양의 텍스트 데이터를 간단하고 직관적으로 받아들일 수 있는 효과적인 도구이다
- ▶ 사용자 기반 텍스트 클라우드 시각화는 사용자의 특정 관심사나 분석 목표에 따라 텍스트 클라우드를 사용자화 할 수 있다



사용자 기반 Text Cloud 구현

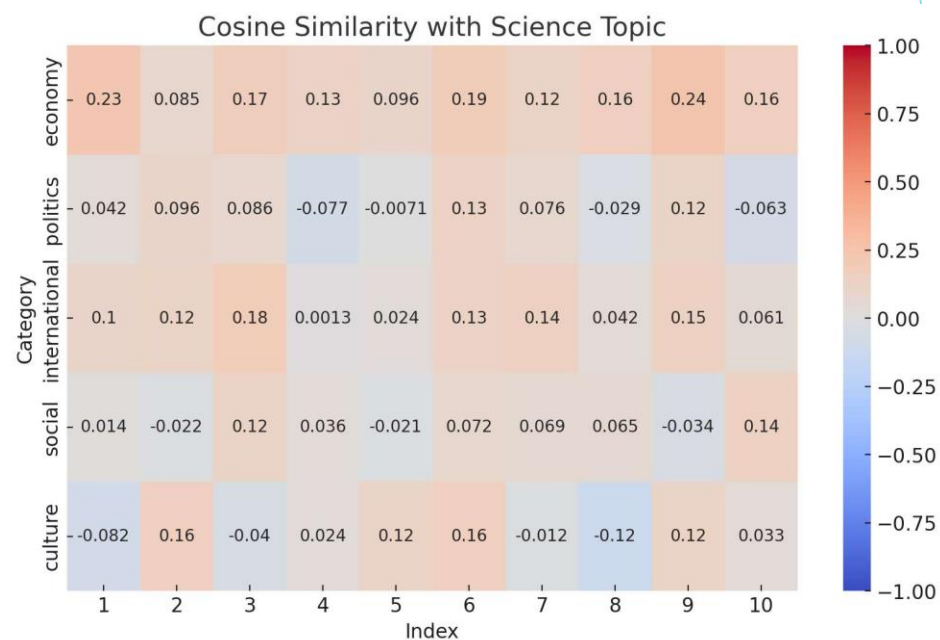
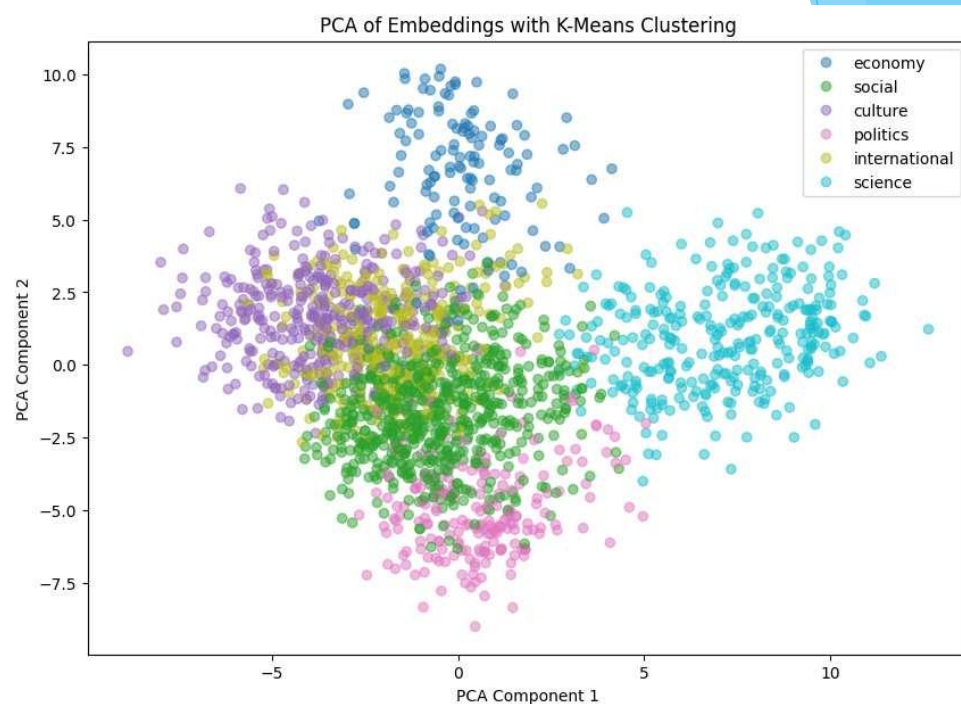
- ▶ 생성형 AI(Fine-tuning한 Llama 3 모델 이용)과 사용자의 대화를 통해 사용자의 감정을 추출한다
- ▶ $\vec{V} = k \times \overrightarrow{sentence} + (1 - k) \times \overrightarrow{sentiment}$
- ▶ \vec{V} 에 따라 뉴스를 추천해준다
- ▶ 추천한 뉴스의 내용을 텍스트 클라우드로 시각화 한다

Fine-tuning에서의 비교 평가

- ▶ 사용자의 **MBTI**를 분류하려 했으나, 정확도가 너무 낮아 실패 (**Test-set**에서 약 **23%** 정확도)
- ▶ 감정 분류는 생성형 **AI(Llama 3)** 보다는 감정 분류에 특화된 **Klue** 모델이 정확도가 더 높았다
- ▶ 공감적 응답 생성 또한 학습한 결과가 너무 좋지 않았다
- ▶ 요약 생성 학습에서는 준수한 성능을 보였다
- ▶ 미세 조정된 모델과 학습이 안된 모델 모두 멀티 턴(**Multi-turn**) 대화를 학습 했을 때, 학습이 안된 모델이 성능이 더 좋았다

뉴스 요약본 군집화

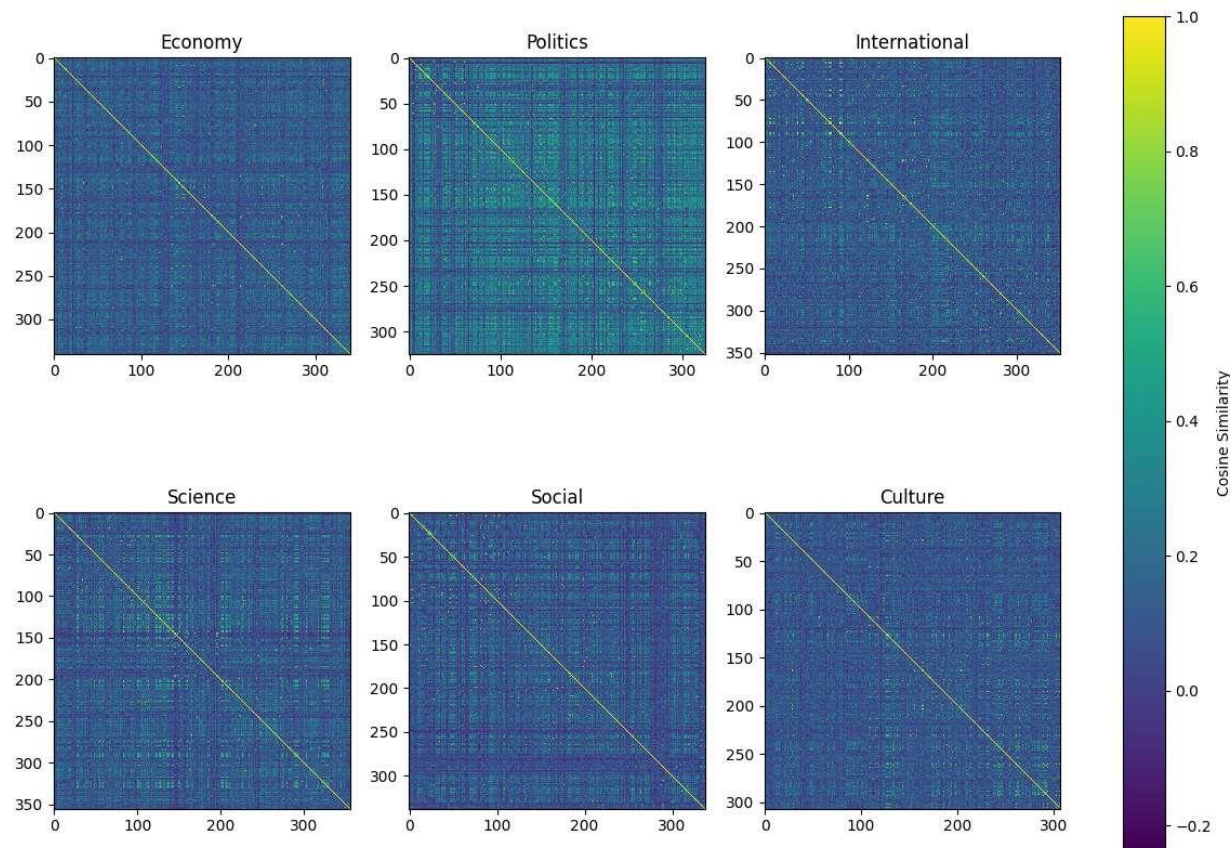
- ▶ Topic 중 science 주제가 다른 주제들 간의 코사인 유사도가 다른 조합들에 비해 확연히 낮게 나타난다



같은 뉴스 Topic 내에서의 코사인 유사도

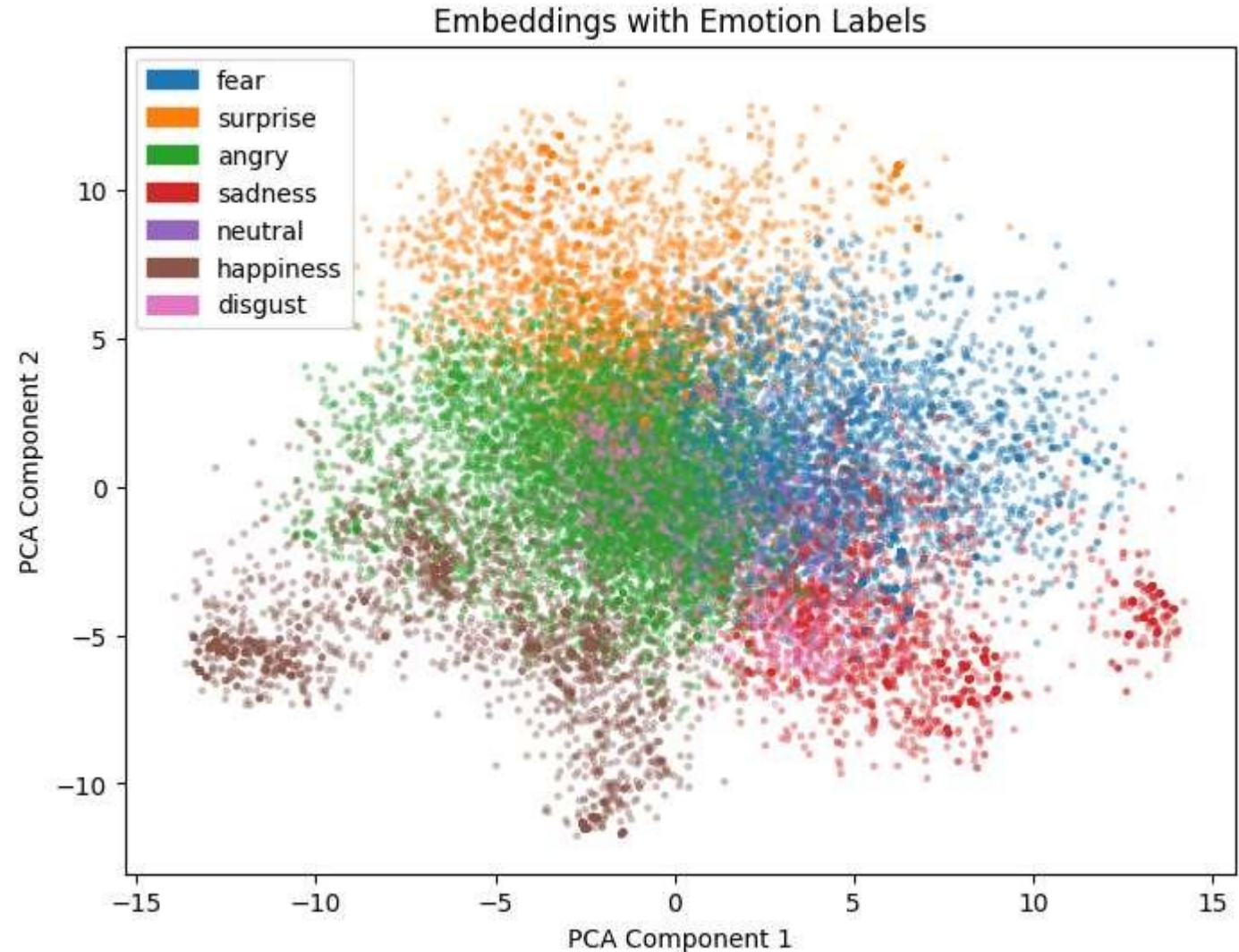
- ▶ 정치(Politics) 분야에서 코사인 유사도가 높게 나타났다. 이는 같은 주제 같은 주제 내에서 여러 언론사들이 보도하는 내용이 상당히 일관되거나 유사한 경향을 보이고 있음을 의미한다
즉 정치적 사건이나 이슈에 대한 분석 및 해석이 유사한 방식으로 이루어지고 있거나 동일한 정보원에 기반해 보도가 이루어지고 있을 가능성이 크다

Cosine Similarity Matrices for Different News Topics



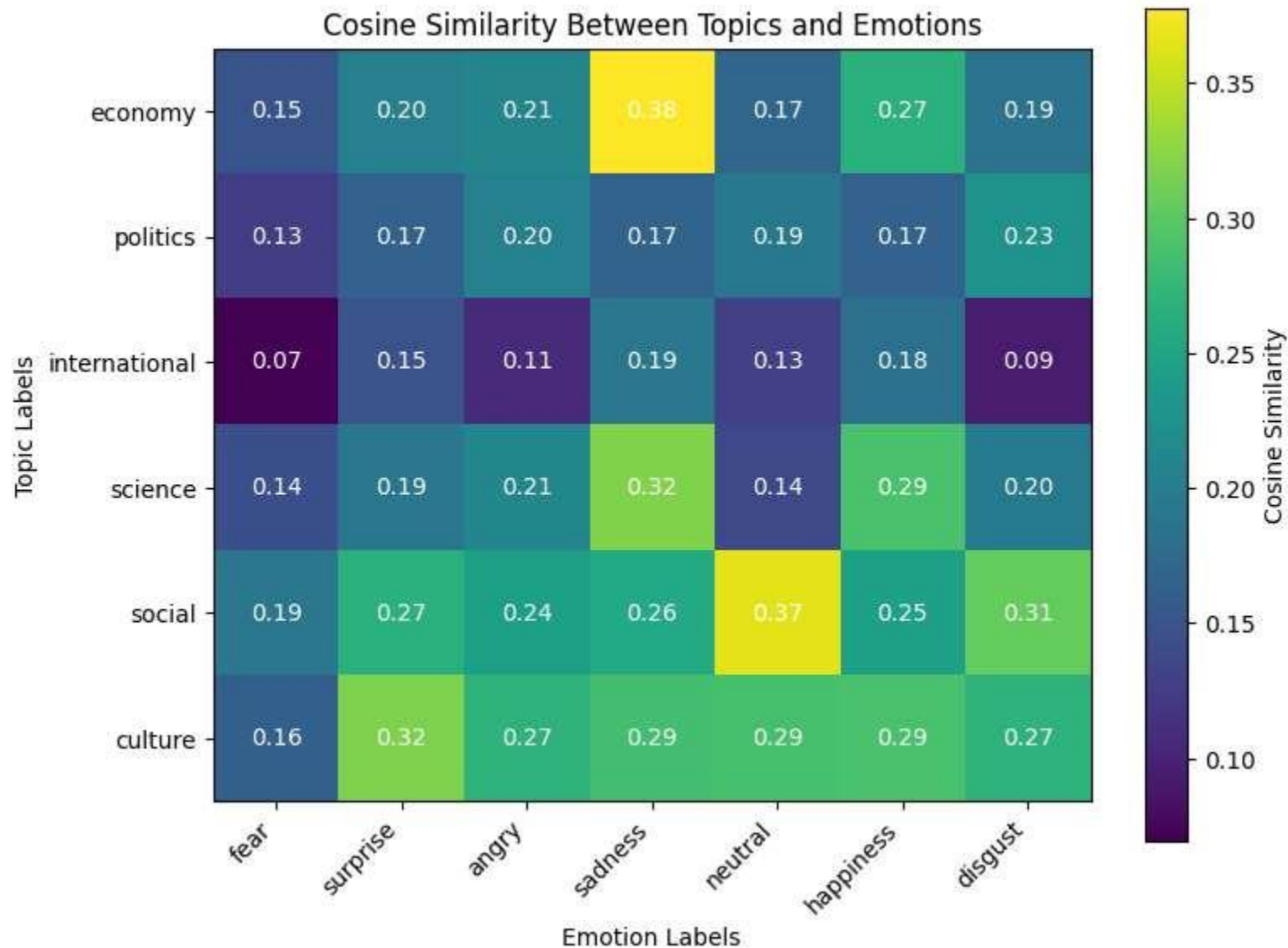
뉴스 추천 및 Text Cloud 생성

- ▶ 혐오(disgust)와 분노(angry) 감정은 색깔상 분홍색과 초록색으로 구분되지만, 그 분포가 서로 상당 부분 겹쳐 있다. 이는 혐오과 분노라는 두 감정이 의미적으로나 표현적으로 유사성이 강하다는 것을 반영할 수 있다



뉴스 추천 및 Text Cloud 생성

- ▶ 경제 관련 뉴스가 슬픈 감정과 높은 유사도를 보인다.
이는 경기가 좋지 않을 때 슬픈 감정이 뉴스 기사에도 드러난다는 것을 의미한다



뉴스 추천 및 Text Cloud 생성

▶ 시연 영상 참고