

웹툰 댓글 분석을 통한 감정 예측 및 토픽 모델링:드라마화 여부에 따른 독자 반응 비교

Abstract

이 연구는 네이버 웹툰 댓글 데이터를 통해 드라마화된 웹툰과 그렇지 않은 웹툰의 댓글 양상을 비교 분석했다. 감정 분석에 머신러닝과 LLM 모델을 사용해 성능을 비교하고, 드라마화 여부에 따른 주요 댓글 주제를 토픽 모델링으로 분석했다. 결과적으로 드라마화된 웹툰에서 독자와 스토리 관련 토픽의 비중이 높았으며, 이를 통해 웹툰 산업 발전에 대한 시사점을 제시하였다.

Introduction

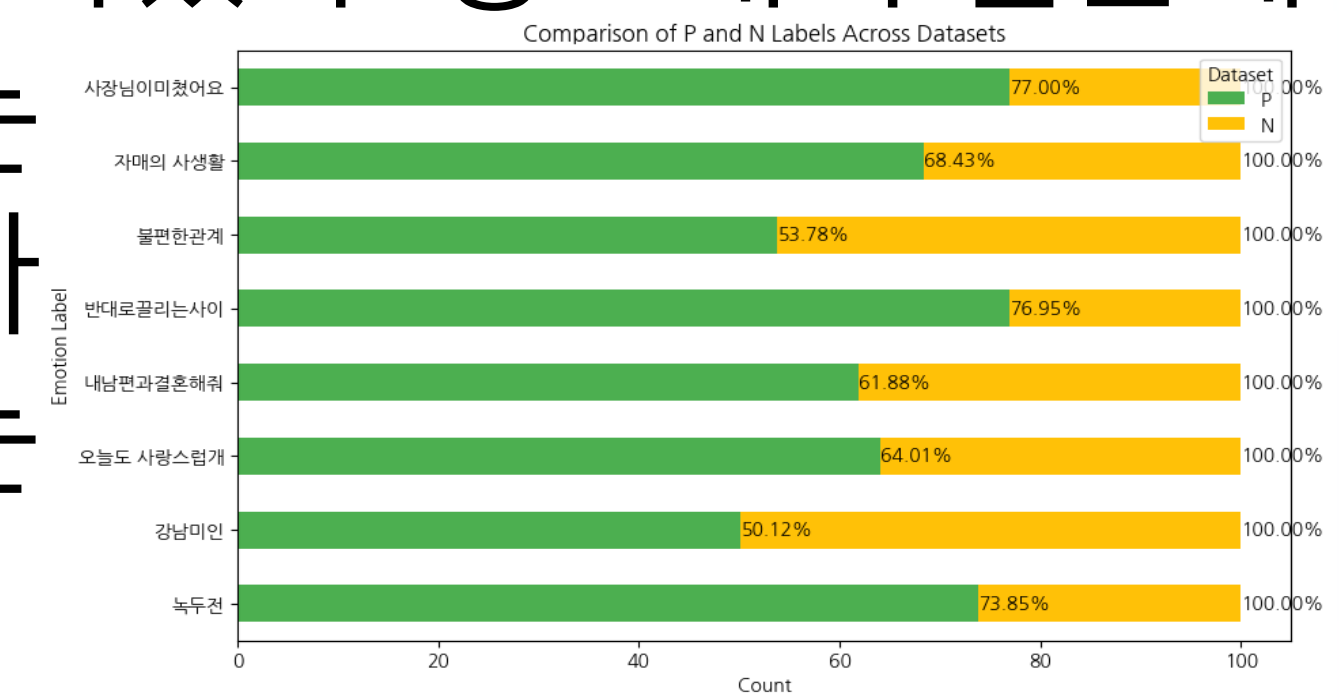
최근 웹툰 산업은 빠르게 성장하며 드라마나 영화로 재탄생하는 사례가 많아지고, 상업적으로 큰 성공을 거두고 있다. 독자들의 즉각적인 반응을 반영한 댓글은 웹툰의 인기를 강화하며, 이러한 댓글 분석을 통해 작품의 인기도와 반응을 예측할 수 있다.

Aim

웹툰 댓글 데이터를 라벨링하여 감정 분석 모델을 학습하고자 한다. 특히, 드라마화가 많이 이루어지는 '멜로' 장르를 중심으로 연구를 진행하며, 드라마화된 웹툰과 그렇지 않은 웹툰의 댓글 양상을 비교 분석할 것이다.

Methods

본 연구에서 사용된 데이터는 네이버 웹툰에서 댓글을 추출하여 구성하였다. 각 웹툰에서 추출된 댓글 수가 다양했기 때문에 특정 회차에 치우치지 않도록 2~3,000 개를 선정하였다. 총 8개의 웹툰에서 추출한 데이터 개수는 20,227개였으며 감정사전을 기준으로 긍정 또는 부정으로 라벨링하였다.



감정 예측 모델

데이터를 바탕으로 머신러닝 기반 앙상블 모델, LSTM 기반 분류 모델, LLM 기반 분류 모델, Clova 모델을 이용한 감정분류를 각각 수행하였다.



토픽 모델링

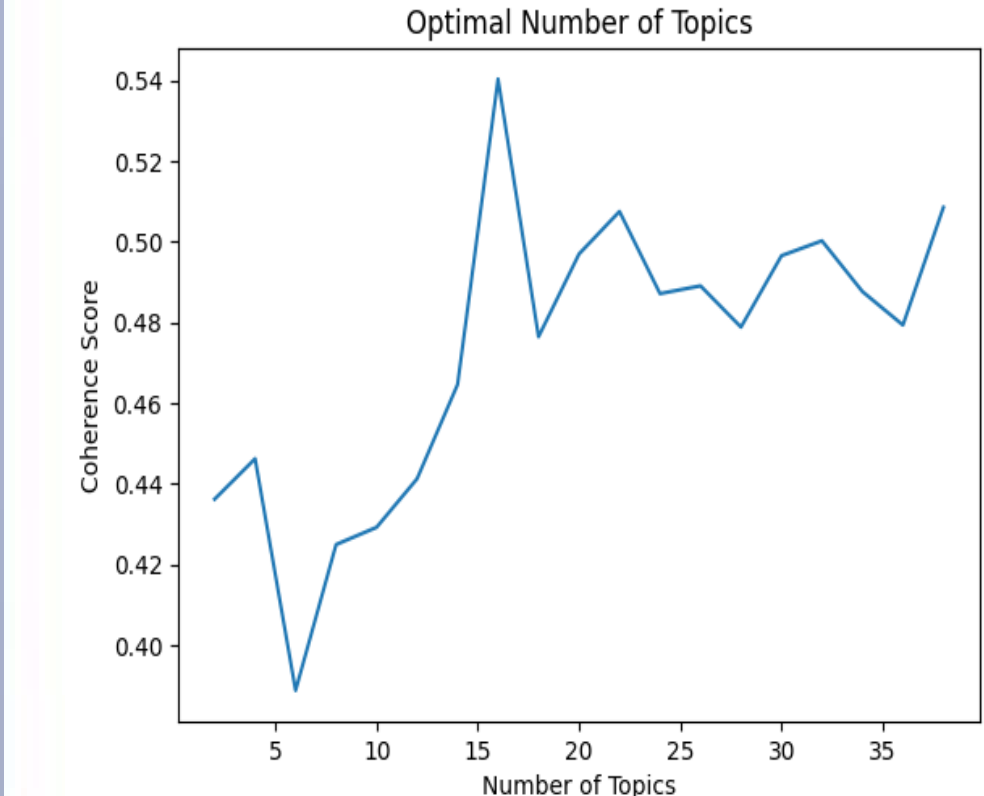
웹툰 댓글에 내재된 주제를 식별하고 분류하기 위해 잠재 디리클레 할당(LDA) 기법을 적용하였다.

- 데이터 정제

반복되는 단어를 정규 표현식으로 감지하고 하나로 치환했으며, Okt 형태소 분석기로 주요 단어를 추출했다. 또한, 불용어 사전을 사용해 불용어를 제거하여 최종 단어 리스트를 생성하였다.

- 분석 과정

1.최적의 토픽 개수



2.토픽별 대표 단어 추출

word1	word2	word3	word4	word5
0.087**정주행	0.073**쿠키	0.050* "작가"	0.031**처음	0.014**어머"

3.토픽별 대표 댓글 추출

Representative Content
안녕하세요 작가님 제가..내일 입대입니다 정말로 뭐라 설명할수가없어서 정말 답답했는데 항상 작가님 웹툰 볼 때 만큼은 아무생각없이 즐기는거같아요.. 이 웹툰 (...)

4.토픽 라벨링

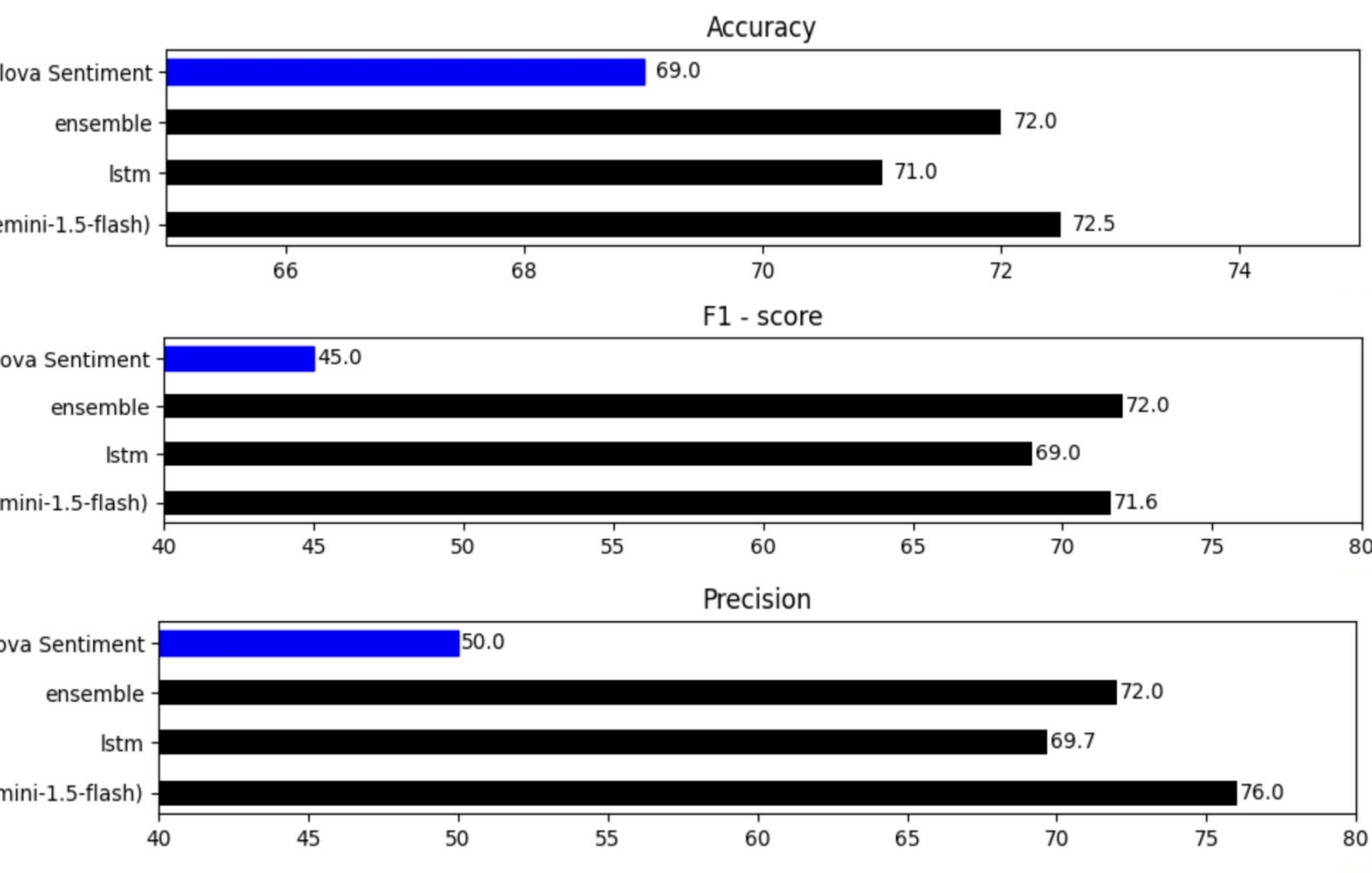


elbow method를 활용해 웹툰별로 토픽 개수를 추출해주고, 각 단어와 문서가 특정 토픽에 속할 확률을 추정하여 토픽 라벨링을 진행하였다.

Results

감정 예측 모델

네 가지 모델 모두 웹툰 댓글 감정 분석에서 높은 성능을 보이지 않았으며, 가장 높은 정확도는 72.5% 였다.

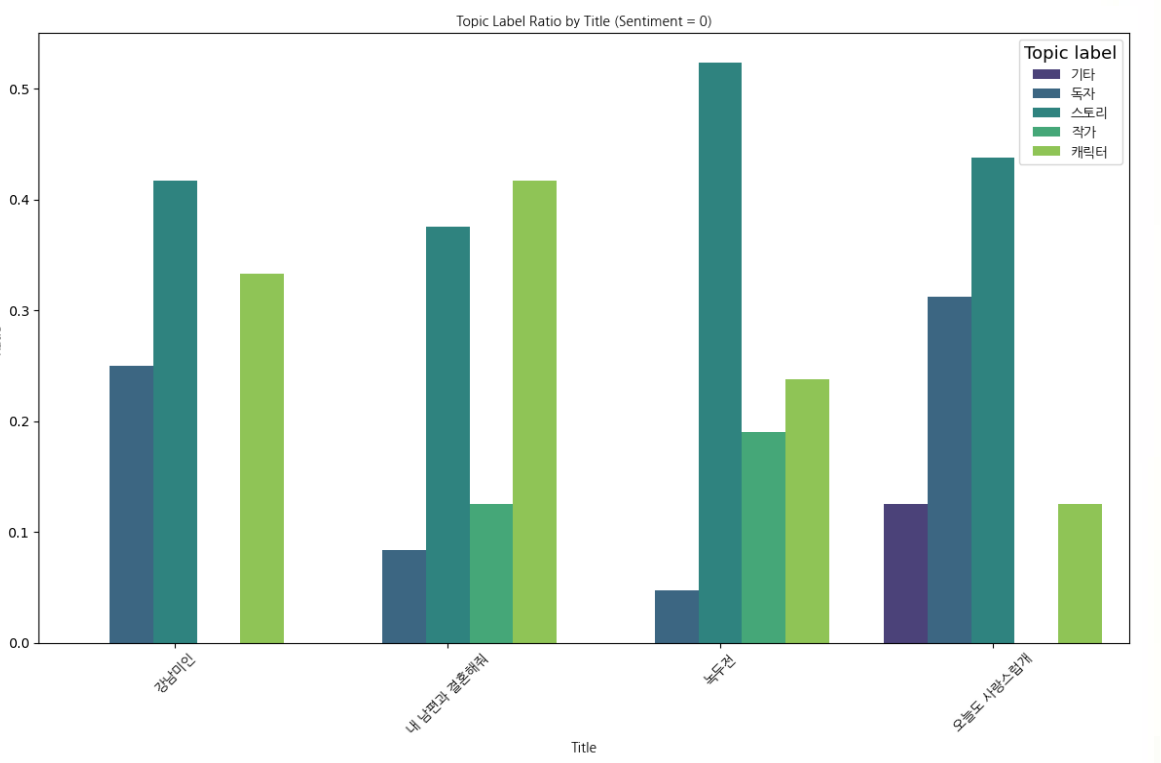
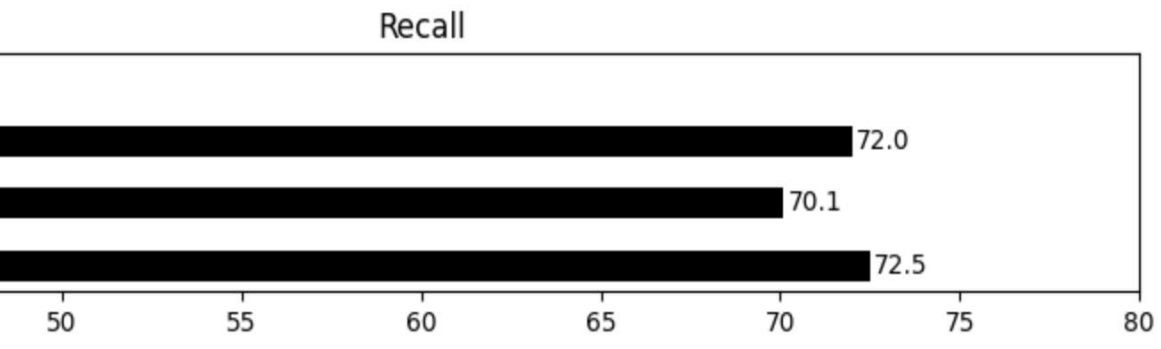


그 중에서도 LLM

모델이 가장 나은 성능을 기록하며, 복잡한 문맥과 감정적 뉘앙스를 파악하는 데 강점을 보였다.

토픽 모델링

두 경우 모두 '스토리'와 '캐릭터' 토픽이 주요 비중을 차지했지만, 드라마화된 웹툰에서는 '독자' 토픽의 비중이 비드라마화된 웹툰에 비해 훨씬 더 두드러졌다. 또한,



토픽분류	드라마화	비드라마화
스토리	시대적 배경 설명, 사회 이슈 언급, 전개 속도	전개 속도, 분위기, 개연성, 연출
작가	업로드 속도, 오타	현실 고증, 작화

- 스토리**
 - 분위기, 연출, 개연성 등의 플롯 완성도 향상
 - 시대적 배경을 활용해 독자는 세계관과 주제 의식 전달
 - 사회이슈를 활용한 사회적 담론의 장 마련
- 독자**
 - 독자들의 감정 이입을 통한 작품 참여도 향상
 - 자아 성찰, 추억 회상, 일상 공유를 위한 소통의 장 마련
- 작가**
 - 현실 고증 리서치를 통한 작품의 진정성 확보
 - 일관된 작화와 오타 수정 등 디테일 제공

'스토리', '작가'별 댓글 내용이 드라마화 여부에 따라 상이했다.이를 바탕으로 산업 방향성을 토픽별로 제시한다.

Conclusion

머신러닝을 이용하여 감정 분류에서 유의미한 성과를 보였고 긍정적 및 부정적 댓글에서 주요 토픽을 효과적으로 추출하였다. 다만, 특정 플랫폼에 한정된 데이터와 댓글의 문맥이나 뉘앙스를 완벽히 반영하지 못하는 한계가 있으며, 향후 다양한 플랫폼과 고도화된 자연어 처리 기법을 통해 분석 정밀성을 높여야 한다.

Reference

1) P. Lee, and S. Park. "An Exploratory Study on the Critics's Reviews Reported..." Journal of the Korea Entertainment Industry Association, Vol. 13, No. 7, pp. 1-13, 2019.
2) X. Sun, X. Li, J. Li, F. Wu, S. Guo, T. Zhang, and G. Wang. Text classification via large language models.
3) 웹툰댓글 공부정 감성사전 구축방안, (링크 생략)