

Review Decomposer Application







17기 김재윤

문제 상황

영화 리뷰의 극단화

- 차별적 리뷰의 극단화
 - 여성이 주연이면 불편(캡틴 마블)
 - 유색인종이면 불편(인어공주)
 - 원하는 방식의 전개가 나오지 않은 경우 기존팬의 테러(스타워즈: 라스트 제다이)



	DreamtOy 4시간전	
얼굴의 반을 가려도 이경실 ㅠㅠ		
답글2	 55	 3
<hr/>		
	ItsShawTime 3시간전	
엠버 허드, 샤를레스 테론, 케이트 업튼.....		
좋은 여주는 넘치는데..어디 가족 드라마나 찍을 여자를 히어로 주연으로 뽑음..		
답글7	 24	 1
<hr/>		
	Tarian 3시간전	
캡틴 페미겠지.		
답글 작성	 10	 2

서비스 소개

- 리뷰 시스템의 문제:
 - 높은 평점은 재미를 보장할 수 없음 (내 취향이 대중과 맞지 않다면?)
 - 별점 테러를 걸러낼 수 있는 장치가 부재함 (일부 극렬한 트롤이 있다면?)
 - 별 다섯개만 가지고는 복합적인 평가가 부족함 (내가 좋아하는 요소가 따로 있다면?)
- 해결책
 - 리뷰를 텍스트를 감정분석하여 개인화된/보정된 점수체계를 만들어보자!

서비스 소개


- 여성 주연 영화는 별점테러를 받음
 - 하지만 나는 여성주연 영화가 좋은걸?
 - 이런 리뷰들은 보통 여성혐오적 표현이나 공격적 표현을 담고있음
 - 공격성과 혐오표현지수를 추출한 후, 보정된 점수체계를 만들자
- 예시: 별점 2.0, 40명
 - 공격적 리뷰 30개가 1.0점을 줌 / 트롤 제외평점 4.0, 10명
 - 공격적 리뷰에 +3점 보정 -> **최종 평점 4.0으로 보정됨.**

**DreamtOy**

4시간전

얼굴의 반을 가려도 이경실 πππ

답글2 😊 55 😞 3


**ItsShawTime**

3시간전

엠버 허드, 샤를레스 테론, 케이트 업튼.....

좋은 여주는 넘치는데..어디 가족 드라마나 찍을 여자를 히어로 주연으로 뽑음..

답글7 😊 24 😞 1

**Tarian**

3시간전

캡틴 페미겠지.

답글작성 😊 10 😞 2

파이프라인

- 트위터 데이터를 기반으로 한 혐오표현/공격성 분석기(Linear Regression Model, 출처: kaggle)
- 로튼 토마토에서 리뷰 수집(Beautiful Soup)
- 리뷰를 공격성 분석기로 분석, 분석 결과를 DB에 저장(MongoDB)
- 해당 어플리케이션을 웹에 구현(Flask)
- 모인 리뷰를 대시보드로 구현(Looker studio)
 - 주의: 크롤링 과정에서 문제가 있어서 데이터가 빈약한 상태

서비스 시연 및 대시보드 분석

- 로튼 토마토 영화 주소에서 영화 제목 따오기
- 따온 제목을 입력할 경우 평점 분석
- 대시보드
 - 보정된 평점과 평론가 평점, 관객평점 비교 -> 어느정도 신뢰도 있는 지표임
 - 공격적 리뷰가 준 평점 통계 -> 공격적 리뷰어는 골고루 평점을 준다고 해석됨
 - 평론가평점/관객 평점/평점차에 따른 공격적인 리뷰의 비율 -> 특히 관객평점에서 낮은 점수에서 공격성이 높아지는 경향 있기는 함.

응용 가능성

- Sentiment analysis 분석기를 여러가지 설치, 리뷰에서 더 많은 감정이나 평가 방식을 추출: 배우에 대한 선호 등
- 리뷰에서 추출해낸 여러 지표들을 입력, 개인이 매겨놓은 평점을 타겟으로 지도학습하여 철저히 개인화된 모델을 만들어낸 후, 이를 기반으로 개인화된 평점 제공
- 예: 이 사람은 혐오표현 리뷰에 전혀 동의하지 않으므로, 혐오표현 리뷰들은 전부 +3점으로 보정. 이 사람은 배우에 대한 선호표현에 신경쓰지 않으므로, 해당 리뷰들은 무시하고 집계

마무리

- 이슈

- 로튼토마토 사이트 구조 특성상 크롤링에서 여러가지 이슈가 발생. 크롤링이 차단당하거나 리뷰를 첫페이지밖에 수집할 수 없게 되어있어 우회해야한다거나.
- 때문에 불완전한 데이터, 모델로 분석을 진행할 수 밖에 없었음

- 아쉬움

- 더 다양한 감정분석을 할 수 있었으면 좋았을 것 같다.
- 로튼토마토에서 더 많은 영화를 스크랩할 수 있었다면 더 다양한 추이분석이 가능했을 것