

프랑스 영화 감정분석 (주제 변경)

권성우 (KWON SUNGWO) 2023-82030

2023/11/29

1 연구개요

1.1 연구배경

오늘날 여러가지 머신러닝과 딥러닝 기법을 사용한 감정 분석은 그 정확도면에서 각광 받고 있는 분야이고, 특히 그 예시로 많이 사용되는 주제 중 하나가 영화 감상평을 감정의 polarity에 따라 분류하는 작업이다: 그에따라 한국어는 네이버, 영어는 IMDb 같이 대용량 데이터베이스를 가지고 있는 전문적인 영화 리뷰 사이트의 데이터를 가지고 연구가 진행된 바가 있는 상황이다. 이번 연구에서는 프랑스 텔레비전 산업 전문 사이트 SensCritique의 영화 감상평 데이터를 가지고 현존하는 머신러닝(및 딥러닝) 기법을 사용한 감정분석을 실행 해 보려 한다 (프랑스 사이트 그리고 다른 감정 분류 task 중에서 영화 감상평을 선택한 이유는 연구자가 한국계 프랑스인이고 영화를 보는 취미가 있는 까닭이다).

1.2 연구목적

이 프로젝트를 통하여 실행 해보려 하는 주요한 작업은 크게 두 가지이다. 두 가지 모두 수업 시간에 배운 내용들의 연장선에 위치하는데, 첫번째로는 Beautifulsoup, pandas, pickle과 같은 파이썬 패키지와 라이브러리들을 이용해 웹에서 정보들을 스크래핑 해온 다음 가공하는 것이고, 두번째는 수업 시간에 배운 TF-IDF 기법을 적용한 로지스틱 회귀로 간단한 감정 분류기를 구현해보는 것이다.

1.3 데이터 선정과 수집

이 목적을 달성하기 위해 상기된 사이트의 영화들의 제목, 감상평, 별점에 해당하는 데이터를 스크래핑 해 활용하려한다.

2 데이터 구축 및 활용 계획

2.1 데이터 가공

데이터 전처리와 가공에는 수업 시간에 배운 기술(정규표현식, Pandas, pickle etc.)들을 활용할 계획이다.

2.2 데이터 활용

필요한 데이터가 모두 수집되면 분류기 모델 구현과 더불어 감상평 글자 수 분포, 별점과 polarity 별 분포도 살펴 볼 수 있을 것이다.

3 참고문헌

SensCritique : <https://www.senscritique.com/>