

인터넷 기사의 댓글 속 혐오 양상에 대하여

20205178 박 현

인공지능 알고리즘을 이용한 사회과학연구 – 기말 프로젝트

최근 온·오프라인 상에서의 혐오가 사회적 문제로 떠오르고 있다. 이에 수 년간의 인터넷 기사 댓글을 수집하여 Korean UnSmile Dataset에서 제공하는 인공지능 모델을 이용하여 각 댓글에 해당하는 혐오 분류를 찾아 이들에 대한 분포 및 양상을 군집화 및 회귀를 진행하고자 한다.

[R Packages]

future: Unified Parallel and Distributed Processing in R for Everyone

future.apply: Apply Function to Elements in Parallel using Futures

reticulate: Interface to 'Python'

rvest: Easily Harvest (Scrape) Web Pages

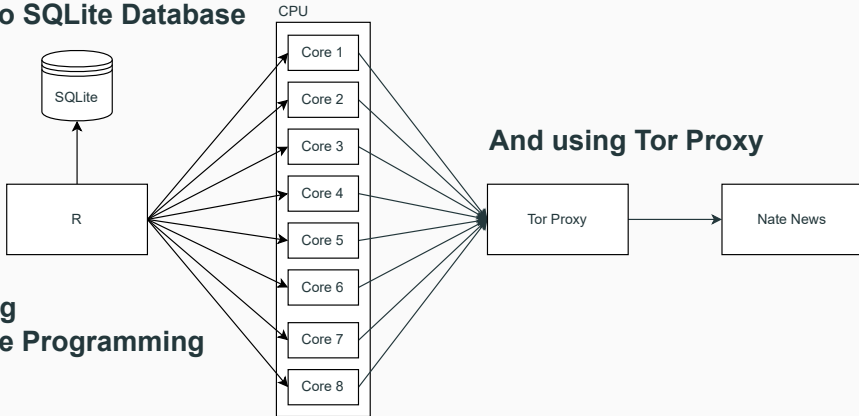
tidyverse: Easily Install and Load the 'Tidyverse'

[Python Libraries]

Hugging Face: Datasets, Transformers.

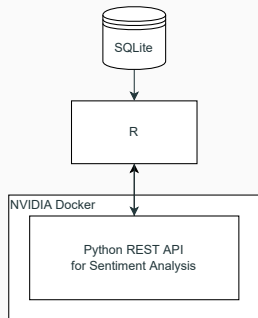
Pandas: Python Data Analysis Library

Finally, store to SQLite Database



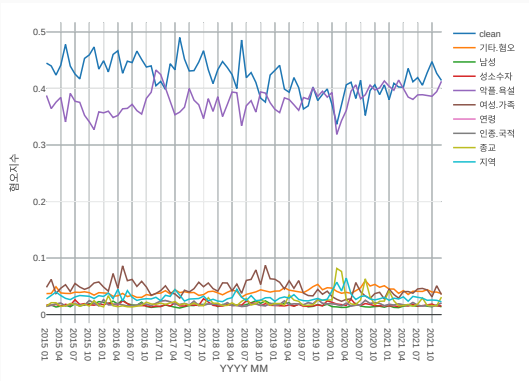
**Web Scrapping
with Multi Core Programming**

And using Tor Proxy



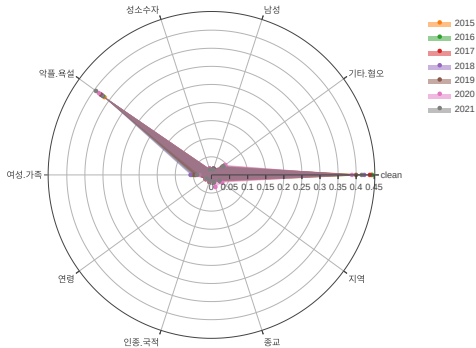
- ① SQLite로부터 수집한 댓글을 가져오기.
- ② 댓글마다 R의 `httr::GET(url)`을 이용하여 도커 내 REST API 서버에 GET 요청을 보냄.
- ③ Docker 안에서 BERT 모델을 이용해 감정 분석을 GPU로 가속하며 진행.
- ④ R에서 감정 분석 결과들을 `do.call(rbind, dfs)`를 이용해 병합.

시각화 – 꺾은선 그래프



- clean과 악플의 비율은 비슷하다.
- 남성과 여성의 양상이 유사하다.
→ 상관분석을 해보자.

시각화 – 레이더 차트



- clean과 악플의 비율은 비슷하다.
- 남성과 여성의 양상이 유사하다.
→ 상관분석을 해보자.

감정 분석

악플과 그렇지 않은 댓글의 비율은 비슷했다.

상관 분석

여성에 대한 혐오와 남성에 대한 혐오가 매우 큰 상관성을 갖는 것으로 나왔으며, 종교와 지역 간의 상관성과 남성과 성소수자 간의 상관성 또한 다른 요소에 비해 높은 것으로 나타났다.