

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

stringr: Simple, Consistent Wrappers for Common String Operations

[Python Libraries]

Hugging Face: Datasets, Transformers.

Pandas: Python Data Analysis Library

원활한 분석을 위해서는 충분한 양의 댓글을 수집할 필요성이 있었고 이를 위해서는 댓글 순으로 기사들을 정렬해주는 서비스를 제공하는 사이트에서 댓글을 수집할 필요가 있었으며 이를 만족하는 사이트로써 nate를 찾게 되었고 해당 사이트에 맞게 Python의 Requests와 BeautifulSoup를 이용해 스크래퍼를 개발하게 되었다.

... 하지만 future가 reticulate를 지원하지 않아 rvest로 다시 작성했다.