딥러닝 기반 비속어 필터링

- 6주차 -

2021202057 황세영 2021202058 송채영

Doing 진행사항

[6주차]

- 원래 sst변환을 통해 단어의 유사성을 구분하려고 했지만, 많은 시간이 소요될 것이라 예상됨
- sst 변환 과정을 생략하고 형태소 분석을 통해 유사한 단어들을 비속어로 변환하는 과정을 사용하기로 함
- 띄어쓰기가 포함된 비속어 데이터를 수집함 (ex, 아저씨 발냄새 나요, 해바라기씨 발아, 무지개 같은 사장님, 곧 휴가 철이네 등)
- 네이버 금칙어 데이터, 인스티즈 금칙어 데이터를 수집함
- -> 네이버는 19금과 관련된 단어들을 금칙어로, 인스티즈는 (ex, D쥐고, D지고, jonna, jot같, mi쳤, tlqkf, wlfkf, 씨1리, 씨1리)과 같은 영어, 특수

기호를 사용한 단어들을 금칙어로 설정해둠

- 딥러닝 모델의 입력값으로 사용하기 위해 임베딩 기법으로 분류하기를 시도해봄

To Do 앞으로 할 것

딥러닝 기반 비속어 필터링 에브리타임, dcinside, Instagram 게시물, youtube, 뉴스 등 데이터 크롤링으로 예외처리로 비속어로 분류될 수 있 자주 방문하는 사이트 내 댓글을 통해 단어(data) 획득 으나 비속어가 아닌 데이터 찾기 욕설의 표준, 기준이 되는 것들을 얻을 예정 sst 변환을 통해 유사 텍스트 변환을 통해 유사성 비교와 단어 학습(라벨링) 비속어인지 아닌지를 비교함 단어를 임베딩 기법으로 분류해 Word2Vec or fastttext 딥러닝 모델의 입력값으로 사용 비속어로 추출된 어절을 중심으로 앞 뒤 단어를 trigram으로 반환해 벡터화함 2 class로 나누어 벡터화한 데이터를 비속어인 경우 0. CNN 모델에 넣음 비속어가 아닌 경우 1 그래드 캠과 shap을 데이터를 시각화함

[7주차]

- 획득한 데이터를 임베딩 기법으로 분류
- 중간 계획서 작성 후 제출
- 발표 영상 찍기

[8주차 ~ 12주차]

- 벡터화, CNN모델에 넣기, 출력, 데이터 시각화 진행

[13주차]

- 프로젝트 최종 보고

Done 한 것

- 주제선정
- 데이터 획득

(에브리타임, dcinside, Instagram 게시물, youtube, 뉴스 등 자주 방문하는 사이트 내 댓글을 통해 욕설의 표준, 기준이 되는 것들을 얻음)

- + 데이터 획득(5000개 이상)
- -> 띄어쓰기가 포함된 비속어 데이터를 수집함 (ex, 아저씨 발냄새 나요, 해바라기씨 발아, 무지개 같은 사장님, 곧 휴가 철이네 등)
- -> 추가 데이터 네이버 금칙어 데이터, 인스티즈 금칙어 데이터를 수집함
- + 딥러닝 모델의 입력값으로 사용하기 위해 임베딩 기법으로 분류하기를 시도해봄
- -> 완성 x