

이름: 조명훈

깃허브: <https://github.com/myunghoon96>

1. 네이버 영화 리뷰 분석

분야: 자연어처리

역할: 단독

날짜: 2019-06

개인 프로젝트로 데이터는 네이버 영화 리뷰를 사용하였으며, 전처리 과정에서 KoNLPy를 사용하여 형태소 분석, 불용어 제거등을 진행하였다. 출현 빈도가 높은 단어들은 word cloud로 시각화하였으며, 공기어들은 네트워크로 만들어 gephi를 사용하여 시각화하였다. 감성 분석으로는 0과 1로 라벨링된 리뷰들로 모델을 학습하여 사용하였다.

2. 기본상식을 가진 페르소나 생성

분야: 자연어처리

역할: 데이터 전처리, GPT-2 fine-tuning, NLI 파트

날짜: 2020-09

팀 프로젝트로 3인이 팀을 이루어서 진행하였다. 페르소나는 사용자의 특징 정보를 가진 고유의 정보로 타인에게 비추어지는 외적 성격을 나타내며, 다양한 페르소나를 가진 인공지능 형성을 위하여 프로젝트를 진행하였다. CoMeT의 결과물을 spaCy를 사용하여 문장화하였으며, GPT-2를 fine-tuning 하여 유사하고 다양한 페르소나를 생성한 후 NLI를 활용하여 필터링하였다.

3. 이미지 장소 분류

분야: 컴퓨터비전, 자연어처리

역할: 단독

날짜: 2021-03

개인 프로젝트로 3단계로 나누어서 진행하였으며 데이터로 사진의 이미지, gps 정보, 태그들을 사용하였다. 첫 번째 단계에서 gps 정보들을 기반으로 mean-shift 알고리즘을 사용하여 사진들을 클러스터링 하였으며, 두 번째 단계에서는 bundler, SIFT를 이용하여 사진들의 유사도를 기반으로 그래프를 형성하여 서브 그래프들을 하나의 클러스터로 간주하였다. 마지막 단계에서는 사진들의 태그 텍스트의 유사도를 기반으로 분류하였다.

4. 영상물 등급 판별 자동화

분야: 컴퓨터비전, 자연어처리

역할: 자연어처리 파트

날짜: 2021-03

팀 프로젝트로 3인이 팀을 이루어서 진행하였으며, 영상물 등급 판별 자동화를 위하여 프로젝트를 진행하였다. 데이터는 영상물과 자막 파일을 사용하였다. 먼저 영상물을 down-sampling 하고, object-detection을 사용하여 폭력적이거나 선정적인 물체들을 체크하였다. 자막 파일은 전처리단계에서 불용어를 제거하고, trained svm을 기반으로한 profanity-check을 사용하여 폭력적이거나 선정적인 텍스트들을 체크하였다. 마지막 단계에

서 선정적이거나 폭력적인 장면, 텍스트들의 기록을 기반으로 점수를 산출하여 영상물 등급을 자동으로 판별하였다.