



Instagram Hashtag Generator

2021-1 YBIGTA 교육세션 신입기수 프로젝트



김연재 김채형 박다현
윤성현 주우진



① 프로젝트 소개

② 프로젝트 과정

③ 프로젝트 계획

프로젝트 목표

입력 이미지에 대하여 자동으로 해시태그 생성하기

! 레 고 레 고 !





입력 이미지와 유사한 이미지 선정

- 입력 이미지와 유사한 이미지 K 개 선정
- 태그 토큰화 : 서로 다른 태그에 공통으로 포함된 단어를 토큰으로 처리
- 선정한 K개의 이미지들이 공통으로 가지고 있는 태그/토큰들을 선별하여 공통 토큰으로 선정
- 공통 토큰을 기준으로 토큰 전환 및 토큰 확장 수행

토큰/태그 전환

- 워드임베딩 (Word2Vec or Fastext) 학습
- 태그/토큰 간 유사도 계산 위해 필요
- 토큰 전환 : 워드 임베딩 벡터 이용
- 토큰 t 와의 유사도가 일정 수준 이상인 토큰 t' 을 연관 토큰으로 추가하는 것
- 예 : 개 => 강아지

토큰/태그 확장

- 토큰 확장 : 조건부 확률 이용
- 조건부 확률 = 기존 토큰인 t_{curr} 이 들어왔을 때 다음 토큰으로 t_{next} 가 올 확률
- 조건부 확률이 일정 수준 이상이면 t_{curr} 와 t_{next} 를 연결하는 것
- 예 : 개 + 스타그램 => 개스타그램



지금까지 한 것

<데이터 수집>

- 인기 해시태그 Top 100 기준으로 해시태그 한 개당 100개씩 포스트 수집
- 인스타그램 포스트로부터 이미지 url, 이미지 파일, 장소 태그, 본문 해시태그, 댓글 해시태그 수집

<모델링>

- ImageNet으로 학습된 VGG19의 피쳐맵을 디스크립터로 활용하여 입력 이미지와 유사한 이미지 K개 선정하는 모듈 ok

앞으로 할 것

<데이터 전처리>

- 외국어 포스트 제외
- 해시태그가 너무 적거나 너무 많은 포스트 제외
- 해시태그 중 출현빈도가 낮은 것 제외
- 태그 토큰화

<모델링>

- 워드임베딩 학습
- 토큰 전환 모듈화
- 토큰 확장 모듈화

Thank You :)

Q&A