

딥러닝을 활용한 Facebook 대나무숲 ‘공감 AI’ 제작

KU-BIG (문선미, 김강우, 박기찬, 이세희)
지도교수: 산업경영공학부 강필성교수님

연구의 필요성 및 목적



연구방법 및 과정

1. 데이터 수집

1. 대학교 대나무숲 Grouper 이용
2. 서울,고려,연세,성균관,경희 총 5곳 에서 12만개이 post추출
3. 동적 웹페이지를 Crawling할 수 있는 Selenium 사용

2. 전처리 및 Word Embedding

1. Crawling후 기본 전처리
2. Mecab을 통해 형태소 추출
3. Word2vec을 이용하여 형태소에 벡터값 부여
4. 각 형태소의 유사도를 계산

크롤링 데이터 기반으로 word embedding
Word2vec의 skip-gram모델 이용

3. SeqGAN 모델링

Fake Data를 생성하기 위해 독립적으로 학습하는 GAN 모델의 일종
GAN을 강화학습(Reinforcement Learning)을 통해 개선한 모델

연구결과

생성 텍스트 예시

GoodCase

그렇게 어느 날은 언저 잘하고 싶은 친구도 아니었어요. 연애하고 싶다. 작년 이맘때 까지 감사하고 그렇게 하고 싶었어요. 모두 좋은 공간이 너무 힘들어. 그런데 나가고 싶은 후배들은 좋아하지!

BadCase

대술! 우리는 부당하게 하게 수백우유해서 유전자에서 살고싶은 현신 같다. 생각합니다. 진짜로 갈수록 감사하고 부끄럽어. 내가 너를 수술하고 싶었어.

Figure 14: Good / Bad cases of Output sentences

- Good Case의 경우 자연스러운 문장 생성
- 커뮤니티 내의 언어적 특징이 반영된 결과물
- 한 문장 내의 어구 수가 많지 않음
- 문장들 간의 연결이 원활하지 않음

Optimal한 결과를 위한 Experiments

1. An effectitveness of Adversarial Training
2. An effectitveness of Pre-Training
3. Hyperparameter Tuning
4. The ratio in Adversarial Training

=> 네가지 주제로 BLEU-4 score를 기준으로 Optimal Model 선정

결론 및 제언 / 참여자 소감

- 커뮤니티의 언어적 특성을 반영한 한글 텍스트 생성 가능성 확인
- 웹 페이지 상의 포스트를 python의 bs4, Selenium을 활용해 크롤링
- 딥러닝 기반 모델인 Word2Vec, SeqGAN 모델 적용해 텍스트 생성
- 학습 단계 재조정, 다양한 모델 적용, Hyperparameter 조정 등의 다양한 실험을 통해 최적화된 결과 도출
- 방대한 데이터의 크기, 한글 텍스트의 전처리 과정과 모델의 Computational Expensiveness로 인한 한계 봉착
- 이외 상용화를 위해 전체 단계의 경량화를 위한 방법 구상중

문 선미: 친구들과 카페에서 제너레이터로 대나무숲 글 작성하면 좋아요 몇 개나 받을까? 하고 무심결에 했던 농담이 실제 프로젝트로 이어졌고 두 학기에 걸쳐 이제가 마무리가 되었다. 프로젝트를 하며 스스로 공부하고 적용해 보고 수업에서 배운 것 이상으로 많이 배울 수 있었던 것 같다. 후배들과 친구들에게 CCP를 많이 추천해 주고 싶다.

김 강우: 솔직히 '이걸 진짜 해낼 수 있을까'라는 막연한 두려움으로 팀원들을 모았고 기획서를 제출했다. 지금에 서야 돌아보면 무모한 도전이었지만 CCP 프로그램이라는 이름으로 팀원들이 함께 달릴 수 있었던 것 같고, 정말 멀리까지 온 것 같다. 학교를 다니면서 했던 가장 좋은 경험 중 하나로 남을 것 같고 이런 기회를 얻게 된 것에 너무나도 감사하다.

이 세희: 처음에 텍스트마ining, 딥러닝이라는 분야에 전혀 익숙하지 않은 상태였기 때문에, 막연하게 프로젝트를 진행하였다. 하지만 1년간의 활동을 통해 프로젝트를 마무리할 수 있어 큰 성취감을 느낀다. ccp를 통해 이룩한 학업적 성취를 바탕으로, 앞으로 더 심도있는 프로젝트 및 연구를 할 수 있을 것이라 생각한다.

박 기찬: 프로젝트를 진행하면서 정말 많은 것을 배울 수 있었다. 학업적인 부분 외에도 앞으로 인생에 있어 더 소중한 가치관들을 배울 수 있었던 것 같다. 나에게 이런 기회를 준 우리 팀원들에게 진심으로 감사하고 학부일 때 꼭 다양한 프로젝트와 다양한 경험들을 할 수 있도록 친구들과 후배들에게 CCP를 추천하고 싶다.