

빠른 습득력과 기본기를 탑재한 박홍준입니다.



박홍준 1996년 (27세/만 25세) | 남

✉ howay96@gmail.com

☎ 010-4337-0365

☎ -

🏠 (06635) 서울 서초구 서초대로64길

학력사항 대학교(4년) 졸업	경력사항 신입	희망연봉 회사내규에 따름	희망근무지/근무형태 서울전체 정규직, 인턴직, 해외취업	포트폴리오 http://hpark46.github.io/
---------------------------	-------------------	-------------------------	---	---

학력

최종학력 | 대학교 **4년** 졸업

재학기간	구분	학교명(소재지)	전공	학점
2015.08 ~ 2021.12	졸업	University of Rochester대학교 (북-중미)	컴퓨터과학	2.92 / 4.0
2011.09 ~ 2015.05	졸업	Georgetown Preparatory School고등학교	이과계열	-

경력

신입

대외활동

기간	구분	기관/장소	내용
2018.08 ~ 2020.05	사회활동	서울중앙지방검찰청	사회복무요원

보유기술

보유기술명/수준/상세내용
HTML / 초급 HTML과 PHP는 local 서버에서 제작한 Query 시각화용 모의 보험 데이터베이스 웹페이지 제작시 처음 학습과 사용하였으며, 추후에 포트폴리오와 같은 개인 웹페이지 제작/업데이트 경험이 있습니다.
PyTorch / 중급 PyTorch는 Pandas, NumPy, TensorBoard 등의 다른 Library들과 함께, Transformer기반 딥러닝 프로젝트 진행시 필수로 사용하였으며, NLP분야 알고리즘과 논문 구현 연습 과정에서 꾸준히 사용하였습니다. NLP독학과 함께 사용하기 시작했으며, 학습에 필요한 데이터를 전처리부터 Hyperparameter Search하는 과정까지 pipeline을 제작할 수 있습니다.
R Studio / 초급

MY-SQL / 초급

데이터베이스 제작을 통해 MySQL을 학습했으며, 원하는 Query 작성 또는 제 3자의 Query 를 읽고 이해할수 있습니다. 테이블 구축/관계 생성/데이터 로드 등의 것들이 가능합니다

Python / 중급

지난 4년간 모든 대학 과제, 개인 학습과 프로젝트 진행시 파이썬을 주 언어로 사용하며 언어에 대한 역량을 키워 갔습니다. 자바 또는 C++로 구성된 frame work을 Python으로 재구현 한 경험이 많으며, 필요한 알고리즘을 표현하거나 제 3자의 코드를 읽고 이해하는데 어려움이 없습니다.

Swift / 중급

Swift는 Python 다음으로 사용하는데 가장 불편함이 없는 언어로, Xcode에서 CoreData, UI/UX 디자인, Reality Composer, API 호출등 다양한 기능들을 여러 게임/기능성/AR 앱 제작 프로젝트에서 사용한 경험이 있습니다.

Java / 초급

Java는 2년간 대학 과제를 자주 사용하였으며 object oriented programming을 연습 하는데 가장 유용했던 언어 입니다. 다양한 Library을 사용해 보진 않았지만 원하는 알고리즘을 구현하는데에는 어려움이 없습니다.

비즈니스영어

취업우대사항

병역대상	군필 2018.07 ~ 2020.05 공익 이병 소집해제
------	---

포트폴리오/기타문서

파일 구분	파일명
포트폴리오	http://hpark46.github.io/
기타	http://velog.io/@howay96

경력기술서

1) 개인 프로젝트명 : √한국어 Hate Speech Detection 모델

- 기술 스택 : Python, √PyTorch, NumPy, Pandas

- 업무 기간 : 2022.03

- 개발 인원 : 1명

- 상세 내용 : Pre-trained 된 KcELECTRA-base모델을 Korean Hate Speech Dataset을 사용하여 Fine-Tuning 하였습니다. Multi-class classification을 위해 Discriminator의 CLS 토큰들을 Classification Head Layer로 3개의 Label (Hate/Offensive/None) 로 분류할 수 있게 Fine-Tuning 하여, 모델은 Validation Set 기준 74%의 정확도를 보여줍니다. 모델의 학습시 Hyper parameter Tuning 보다는 preprocessing, parameters, pre-trained 모델이 학습에 미치는 영향을 알아볼 수 있는 기회가 되었습니다.

2) 개인 프로젝트명 : 한국어 Semantic Textual Similarity Rest API

- 기술 스택 : Python, √PyTorch, NumPy, Pandas

- 업무 기간 : 2022.04

- 상세 내용 : √Klue benchmark paper에 기재되어 있는 KLUE-ST5를 구현하기 위해 klue/roberta-large pretrained-model을 klue-sts-v1.1 데이터 셋을 사용해 Fine-Tuning 하였습니다. Roberta를 Siamese 구조로 사용하여 두개의 output을 mean pooling 한 값을, Cosine Similarity을 통해 문장의 유사도를 유추

하였습니다. Klue paper 에서 명시된 hyper parameter범위 내에서 random search을 진행 하였으며, 결과물은 0 에서 5의 점수를 부여하는 모델로 Pearsonr 값은 0.8817이며, 3에 threshold을 추가해 binary classification 으로 다루었을때 F1 score은 0.8353을 기록하였습니다. 추후에 KorSTS 데이터 셋도 추가로 Fine-Tuning 시킨 모델로 1:1, 1:N 문장 유사도 점수를 return 하는 Rest API로 구현하였습니다. Siamese structure, hyper parameter search 과 파인 튜닝된 모델의 API화를 중점적으로 학습할 수 있는 기회였습니다.

3) 개인 프로젝트명: MatchSum 문서요약

- 기술 스택 : Python, PyTorch, NumPy, Pandas

- 업무 기간 : 2022.04

- 상세 내용 : Google Colab에서 MatchSum 논문 그대로 Training을 진행하는데에 어려움이 있어 Model Framework와 데이터 셋 preprocessing에 중점을 두어 진행하였습니다. 학습의 핵심인 Candidate 추출 과정에서, 기사의 Pruning을 논문에서 사용한 BertExt로 점수를 부여해 진행하는 방식이 아닌, BertSum에서의 Orade Summary 생성에 사용되는 Greedy-Selection과 Combination-Selection을 변형하여 사용하였고, 각 기사당 20개의 Candidate Summary을 생성하였습니다. 모델은 기사와 Gold Summary/Candidate Summary을 매칭하기 위해 CLS 토큰 벡터가 사용되며 Loss function으로는 MarginRankingLoss가 사용됩니다.