



하동준 (Ha DongJun)



Birth | 1998.02.11

Email | tnfodd02@gmail.com

Phone Number | 010-6478-6194

Address | 서울시 동작구 현충로 151 한강현대아파트

GitHub | <https://github.com/djh0211>

Velog | <https://velog.io/@djh0211>

Introduction

성격

- 꼼꼼하고 세심한 성격으로 매사에 **자료와 정보 등을 분류하고 기록, 저장** 하는것을 생활화하고 있습니다.
- 타인에 대한 **배려와 커뮤니케이션의 중요성**을 배워와 타인과의 원만한 관계를 잘 쌓아오고있습니다

관심 영역

- 전반적인 AI 분야에 관심이 많아 현재 네이버 부스트캠프 AI Tech 3기 추천시스템 트랙에 참여 중이며 논문과 기술 학습을 하고 있습니다.
- 다양한 분야의 문제를 AI 로 해결해보고자 타자와 투수, 경기 내용을 반영한 야구 구종 추천, 주식 개인화 추천, 지역별 업종별 상권 추천 등의 프로젝트를 진행 예정입니다.

지원동기

- **검색엔진 및 추천시스템 분야에** 관심이 많아 대학원 진학을 목표로 하고 있습니다. 현재 재학중인 대학교에는 관련 분야에 관한 커리큘럼이 없어 방학 기간 동안 관심분야에 대한 학습 방법 및 기회를 찾던 중 본 프로그램에 지원하였습니다.

학력

- 2013.03 ~ 2016.02 세화고등학교 졸업
- 2017.03 ~ 2020.12 한국산업기술대학교 컴퓨터공학부 중퇴 (3.46/4.5)
- 2021.03 ~ 현재 건국대학교(서울) 컴퓨터공학부 재학 (3.7/4.5)

병역사항

- 2018.11 ~ 2020.06 대한민국 의경 병장, 만기전역

외국어

- 토익 835점 (2020.10.25, ETS)

※ 영어 역량 개발을 위하여 지속적으로 학습 중 입니다.

주요 역량

Machine Learning & Data Science

- python3, pytorch, Tensorflow, scikit-learn, WandB, Tensorboard

Collaboration

- Git, Github, Notion, Slack

Mobile Application

- Kotlin, Androidstudio

교육

네이버 부스트캠프 AI Tech 3기 추천시스템 트랙

NAVER connect 재단에서 진행하는 Ai Engineer 양성 교육 과정

기간 | 2022.01 ~ 2022.06 (6개월)

주요 내용 | Python, PyTorch, DL 이론 학습 및 미니 프로젝트 참여 : Image classification, Movie recommend, DKT 등

주요 활동 내역

Face mask image classification

카메라로 촬영한 사람 얼굴 이미지의 마스크 착용 여부를 판단하는 AI model을 개발 (mask Wear/ Not wear/ Incorrect, age, gender)

주관 | 네이버 부스트캠프 AI Tech 3기

기간 | 2022.02.21 ~ 2022.03.07

자료 | [Report](#), [Github](#)

Contribution

- **EfficientNet-b2** | Swin-transformer | Convnext large
- EfficientNet-b7 10 epoch 학습 시간 기준으로 b2, b4, b5 도 테스트 했고 성능과 시간을 고려하여 b2 를 최종 선정하였다.
- 학습에 얼굴이 가장 중요할 것으로 판단해 centercrop 사용. 범위가 충분히 크지 않으면 얼굴이 잘려 성능 향상에 도움이 없었다. 따라서 얼굴만 crop 한다가 보단 네 모서리를 자르는 느낌으로 crop 진행했다. 또한 colorjitter 로 다양한 변화와 15도 미만의 Randomcrop 등을 이용하여 기울어진 얼굴들을 학습하도록 노력했다.
- 30세 미만, 30세 이상 60세 미만 데이터가 1000 ~ 1200개 인 반면 60세 이상 데이터가 200개 정도로 imbalanced 하였고 이 때문에 성별, 마스크 상태에 비해 연령 예측 정확도가 현저히 떨어졌다. 따라서 opencv 활용하여 연령을 별도 예측, focal loss 사용, 60세 이상 데이터는 5회 이외 데이터는 1회 re-sampling 등을 시도했고 이 중 효과가 있었던 60세 이상 데이터 1회 re-sampling 을 사용했다.
- 상위 모델 3개 Voting | 60 대 이상 이미지 예측 능력이 낮아 60대 이상 으로 예측한 모델이 한 개라도 있으면 60대 이상으로 분류했다.
- f1 score : 0.7527 | accuracy : 80.1905

Movie recommender system

MovieLens 10M 데이터셋을 일부 전처리 하고 implicit data 로 만든 유저 별 시퀀스 데이터에서 dropout 된 부분을 예측하는 AI 모델 개발

주관 | 네이버 부스트캠프 AI Tech 3기

기간 | 2022.03.21 ~ 2022.04.14

자료 | [Report](#), [Github](#)

Contribution

- BertRec | SASRec | BPR | Mult-VAE | RecVAE | **EASE**
- 유저, 아이템 이외의 feature 들에 이상치 등 학습에 사용시 성능 향상을 기대하기 어려워 대회 task 인 last sequence 예측 보다 top-k 추천 으로 자체 변경하였다.
- 유저가 평가한 Timestamp 가 영화 개봉일 이전 이거나 유저가 10 초 이내 수십개 영화 평가를 한 이력이 다수 존재하여 sequential model 보다 AutoEncoder 를 사용하였다.
- 상위 모델 3개 Voting 및 단일 모델 최고 성능을 낸 EASE 를 k-fold Cross Validation 진행하여 overfitting 을 해결하였고 private 데이터셋에서 좋은 성능을 기록하였다.
- Recall@10 0.1602

Deep Knowledge Tracing

Iscream Edu 데이터셋에서 유저 별 마지막 시퀀스 데이터의 정답 여부를 예측하는 AI 모델 개발

주관 | 네이버 부스트캠프 AI Tech 3기

기간 | 2022.05.17 ~ 현재

Contribution

- Model : LGBM | LSTM | LSTMATTN | BERT
- 데이터의 양이 부족하여 Transformer 보다 LGBM 등 트리 구조가 효과적
- LGBM Feature Engineering

추천 시스템 관련 논문 리딩

추천 시스템 및 AI 논문을 읽고 있습니다.

기간 | 2022.02 ~ 현재

자료 | 블로그 리뷰, 논문리스트

태블릿 전용 이야기책 제작 어플리케이션 개발

건국대학교 산학협력프로젝트로 아동을 대상으로 한 이야기책 어플리케이션 제작

기간 | 2021.10 ~ 2021.12

자료 | 최종발표 ppt

2022.05.10

지원자 하동준(서명) 