

윤도연



생년월일 : 1989.01.26

Email : ydy89899@gmail.com

Phone : 010-3219-9511

Github : <https://github.com/ydy8989>

Github Blog : <https://ydy8989.github.io/>

Notion : <https://www.notion.so/whydo/Doyeon-Yoon-05603016086c4b3ca954cf2b6c64e46f>

안녕하세요. 수학을 좋아하며 AI를 공부하는 윤도연입니다. 현업에서 필요한 수학적 사고와 공학적 사고를 구분하여 활용할 수 있습니다. 다양한 도메인(의료, 금융, 자연어, 추천시스템)에 관심이 많으며, 몇몇 프로젝트 경험을 통해 해당 도메인에서의 Key considerations를 이해할 수 있었습니다. 현재는 NLP와 추천 시스템(GNN)에 관심을 갖고 공부하고 있으며, 궁극적으로 ML/DL Engineer로 성장하는 것이 목표입니다.

경력

(주)에이치비

2019.6 - 2020.3

AI팀 / AI research engineer

- 반도체 공정 스마트 인터락 진/가성 분류 모델 개발 2019.9 - 2020.3

[AI team. machine learning engineer]

"SAMSUNG Smart Interlock true/false classification project (Pilot)"

- 역할 : 공정 센서의 인터락 타입 자동화 분류 프로젝트의 AI 모델 개발 담당
- 데이터 Type : 반도체 공정 데이터 (Image data + Time series sensor data)
- Model architecture : Semi supervised GAN, Auto-Encoder + Linear regression ensemble
- Skills : Flask, Keras, Tensorflow, Tensorflow serving
- 성과 : Phase 1 pilot project 성공적 종료 후 전 공정에 적용.

포항공과대학교 정보통신연구소

2018.11 - 2018.12

연구부/인턴 연구원

- 아산병원 "HeLP Challenge 2018 Contest" 참가 2018.11 - 2018.12

[AI Intern researcher]

1. Medical image Segmentation 경진대회 참가 :

기존 MRI 촬영으로만 진단했던 심장질환을 CT 영상만으로 진단하기 위한 의료이미지 Segmentation 대회. 심장의 세 부위를 정확하게 segmentation하는 contest.

- 역할 : 모델 구현 및 preprocessing 담당
- 데이터 : 3D CT-images (.mha format)
- Skills : Docker, Keras, Tensorflow
- Model architecture : 3D-UNet
- 성과 : Top 5(수상 X)
- github : www.github.com/ydy8989/Cardiac_Segmentation

2. 인근 중소기업 기술 애로사항 해결을 위한 기술지원 참여 :
지역 중소기업 AI 기술지원 및 사업 활성화의 일환으로, 간접적 솔루션 제안 및 프로젝트 지원.
- 역할 : 논문 survey 및 seminar 발표
 - subject : semi-supervised learning
 - review paper : learning by association

| | | |
|----|--|-----------------|
| 학력 | 광운대학교 대학원 수학/응용수학 석사 졸 Combinatorics(A+), Graph Theory(A+) 졸업논문 : The anti-bandwidth of triangular grid graph | 2015.3 - 2017.2 |
| | 광운대학교 수학과 학사 졸 Linear Algebra(A+), Algebra, Graph Theory(A+) etc. | 2013.3 - 2015.2 |
| | 서울과학기술대학교 기계시스템디자인공학 중퇴 공학수학, 재료역학, 정역학, 동역학 등 | 2008.3 - 2013.2 |

| | |
|-------|--|
| 기술 스택 | 언어 Python 기타 SW Pytorch, Keras, Docker, Flask |
|-------|--|

| | | |
|-------|--|--------|
| 교육 이수 | Naver Boostcamp AI Tech 기간 : 2021.01 ~ 2021.06 교육 내용 - 부스트캠프 AI Tech 1기 수료 - AI 기초 지식 및 응용분야 최신 논문 학습 - U Stage(핵심이론)과 P Stage(대회 형식)를 통한 CV / NLP / 정형 데이터 AI 모델 구현 및 개선 프로젝트 경험 자세한 내용 - 교육 정리 : https://github.com/ydy8989/boostcamp | 2021.1 |
| | 포항공과대학교정보통신연구소'AI·Big Data 인재양성 심화과정' 이수 기간 : 2018.09 ~ 2018.10 (2개월 합숙 교육) 교육 내용 - 인공지능 개론, 컴퓨터 비전, 머신러닝 개론, 자연어 처리, 빅데이터 분석1,2 - 팀 프로젝트 진행(뉴스 감성분석을 통한 익일 추가 상/하락 예측) | 2018.9 |

머신러닝과 딥러닝을 활용한 빅데이터 분석전문가 양성과정 수료

2018.2

교육기간 : 2018.02 ~ 2018.06(544시간)

교육 기관 : 에이콘 아카데미 종로

교육 내용

- Python, R, SQL 기초
- Data Analysis(Scikit-learn, pandas, numpy, matplotlib 등)
- Django, AWS

Toy project

1. 음식 추천 챗봇 "꿀꿀봇"

- 모델 : 협업필터링 + seq2seq 앙상블
- Skills : Django, tensorflow serving, 카카오 플러스친구 api

2. 단순 시계열 주가 예측 및 백테스팅 시스템 구현

- 모델 : LSTM
- Skills : tensorflow

프로젝트

뉴스 감성분석을 통한 익일 주가 상/하락 예측

기간 : 2개월

프로젝트 개요 :

같은 뉴스도 기업의 주가에 각각 다르게 영향을 미칠 것이라는 점에서 프로젝트 진행.

특정 기업의 뉴스 기사를 통해 다음날 주가 상승 하락을 예측하는 프로젝트.

역할 : project manager 및 전처리(feature engineering), 모델 구현

데이터 : 뉴스 데이터 크롤링 5년치 (Bigkinds.or.kr)

Skills : Tensorflow, LIME, NLP

Model architecture : 1DCNN

github : [www.github.com/ydy8989/senticle-proj](https://github.com/ydy8989/senticle-proj)

Demo : <https://youtu.be/syQfQGFAAZ0>

대회 참여 경험

마스크 착용 상태 분류(부스트캠프 AI Tech)

기간 : 2021.03.29 ~ 2021.04.09(2주)

리더보드 순위 : 152/223등

대회 개요 : 마스크를 쓴 사람의 성별, 나이, 마스크 유무 총 18개 Class로 이미지 분류

기술 개요 :

- 데이터 : 384 X 512 images
- 모델 : EfficientNet b0 - 4
- 전처리 : EDA를 통한 labeling 조정, Augmentation, Label filtering,
- 학습 전략 : data sampler, Stratified K-Fold
- Github : <https://github.com/bcaitech1/p1-img-ydy8989>

Relation Extraction(부스트캠프 AI Tech)

기간 : 2021.04.12 ~ 2021.04.23 (2주)

리더보드 순위 : 46/135등

Github : <https://github.com/bcaitech1/p2-klue-ydy8989>

대회 개요 :

- 관계 추출(Relation Extraction)은 문장의 단어(Entity)에 대한 속성과 관계를 예측/분류하는 문제입니다. 관계 추출은 지식 그래프 구축을 위한 핵심 구성 요소로, 구조화된 검색, 감정 분석, 질문 답변하기, 요약과 같은 자연어처리 응용 프로그램에서 중요합니다.

기술 개요 :

- 데이터 : train set 9000문장, test set 1000문장
 - input : 문장, 단어1, 단어2
 - output : 단어1과 단어2의 41개 관계 Class
- 모델 : Roberta (hugging face pretrained model), kobart, base-bert 앙상블
- 전처리 :
 - EDA(Easy Data Augmentation)
 - Back Translation
 - TEM(Typed Entity Marker)

Dialogue State Tracking(부스트캠프 AI Tech)

기간 : 2021.04.26 ~ 2021.05.21 (4주)

리더보드 순위 : Public 1위 | Private 1위

Github : <https://github.com/bcaitech1/p3-dst-teamed-st>

대회 개요 :

- 유저와 시스템 간의 대화에서 미리 정의된 시나리오에 의해 정의된 정보(Slot)과 그에 속할 수 있는 Value의 집합인, 대화 상태(Dialogue State) 를 매 턴마다 추론하는 Task

기술 개요 :

- 데이터 :
 - train : 7000개 대화, public 1000개 대화, private 1000개 대화
 - input : (시스템 발화, 유저 발화)로 이루어진 하나의 turn 중, 유저 발화 문장
 - output : 유저 발화에서 얻을 수 있는 정보(Domain-Slot-Value으로 이루어진 pair)
- 모델 : SUMBT, TRADE, SOM-DST 모델 앙상블
- 전처리 :
 - MLM Pretraining
 - COCO Augmentation

역할 :

- 팀에서 SOM-DST 모델 구현 및 implementation 담당

Deep Knowledge Tracing(부스트캠프 AI Tech)

기간 : 2021.05.24 ~ 2021.06.15(3.5주)

리더보드 순위 : Public 2위 | Private 4위

Github : <https://github.com/bcaitech1/p4-dkt-decayt>

대회 개요 :

- KT는 개인에 맞춤형 피드백을 받기 힘들고 성적을 올릴 수 있을 지 판단하기 어려운 상황에서 유저별 "지식 상태"를 추적하는 task입니다. 유저별 시험데이터를 통해 지식 상태를 추적하고 이를 통해 다음 문제의 정답 여부를 예측하는 task입니다.

기술 개요 :

- 데이터 :

- Input : 7442명에 대한 문제 번호, 문제지 번호, 문제 유형 별 시계열 데이터와 정답 유무가 포함된 230만 rows X 6 columns

- Output : 특정 유저의 특정 시점에서의 정오답 예측

- 모델 : LightGBM, SAINT 모델 앙상블

역할 :

- Feature engineering

- EDA

- Model implementation