

박우석 Wooseok Park

Junior Data Scientist



SUMMARY

협업 능력과 꾸준함이 강점인 데이터 과학자, 박우석입니다. 8명의 공저자들과 함께 관측 데이터 분석 결과를 SCIE급 저널에 1저자로 기고하였으며, 5인 팀 프로젝트로 추천 시스템 서비스를 구축, 실사용자 확보 및 A/B 테스트를 수행한 경험이 있습니다. 현재는 추천 시스템에 관심을 갖고 공부하고 있으며, 지속적인 Kaggle 대회 참여로 실전 데이터 분석 및 예측에도 계속 도전하고 있습니다.

✉: hoen5373@gmail.com

🌐: [hoen](https://github.com/hoen)
(<https://github.com/hoen>)

Projects

Jan 2023 - Feb 2023

길동국사 - 한국사능력검정시험 기출문제 개인화 추천 서비스 [repo]

(https://github.com/boostcampitech4lv23recsys1/level3_product-serving-level3-recsys-03)

SUMMARY

효율적인 한국사능력검정시험 학습을 제공하고자 유저 개인별 풀이 기록을 학습하여 유저에게 틀릴 가능성이 높은 문제를 추천하는 서비스입니다.

- 백엔드 및 추천 모델링을 담당하여 추천 시스템 패키지인 RecBole을 이용, SASRec, EASE 를 user-based 추천 모델로 서빙
- 문제 해설에서 형태소를 추출 및 임베딩하여 item-based 이웃기반 협업필터링 모델을 구현하였고, 이를 통해 신규 사용자에게도 합리적인 추천을 제공하여 Cold-Start 문제를 해결
- 프로젝트 기간 1달간 누적 유저 190명, 누적 문제풀이 수 12000건 이상을 달성
- A/B test를 통해 모델 추천 문제의 사용자 오답률이 높음을 확인하여 틀릴 문제를 추천하는 모델의 효율성을 검증함

May 2023 - present

Kaggle - Predict Student Performance from Game Play 참여 [repo]

(<https://github.com/hoen/student-gameplay-prediction>)

SUMMARY

Kaggle에서 주최하는 Predict Student Performance from Game Play 대회에 참여하여 학습용 게임 로그 기록을 바탕으로 유저의 정답 여부를 예측하였습니다. 유저의 게임 기록을 순서대로 추적하는 것이 학습 상태 파악에 적합하다고 판단, sequential model을 이용해 예측을 시도하였습니다.

- LSTM을 이용하여 사용자의 게임 로그를 embedded sequence data로 변환, 정답여부를 예측하는 시계열 모델을 구현
- 기존 LightGBM 모델 대비 LSTM으로 CV f1-score를 0.687 -> 0.76으로 향상

Skills

Data processing and Visualization :

★★★

Pandas Matplotlib

ML optimization & ensemble :

★★☆

Pytorch

Recommendation System : ★★★

Recbole

Version control, Co-op : ★★★

Git Notion Slack

Education

Sep 2022 - Feb 2023

Naver Boostcamp AI Tech 47 | from Naver Connect

- Pytorch, AI Math
- Memory-based / Model-based Collaborative Filtering
- Data Visualization
- Prototyping and Servicing AI

Mar 2019 – Aug 2021

**M.Sc from Space Science
Research, Kyung Hee
University**

Publications



Oct 2021

Quantifying Variability of Young Stellar Objects in the Mid-infrared Over 6 Years with the Near-Earth Object Wide-field Infrared Survey Explorer (https://github.com/hoeen/WISE_YSO_variability_analysis) by **Wooseok Park**

SUMMARY

6년간 축적된 원시성(탄생 초기의 별) 밝기변화 sequence data를 정제 후 통계적으로 분석하여 별의 진화 단계에 따른 밝기변화 특징을 밝혀내고 분류하였습니다. 연구 결과로 초기의 원시성일수록 밝기변화가 불규칙하며 그 강도가 큼을 밝혀내었습니다.

- 신호에서의 주기성 탐지 기법 중 하나인 Lomb-Scargle Periodogram으로 밝기변화의 주기성 여부를 탐지
- 밝기변화를 수치화하여 새로운 feature를 생성 및 이를 이용해 원시성 밝기변화의 형태를 분류
- 장기적으로 축적된 raw 데이터 속에서 유의미한 인사이트를 발견하고 새롭게 정제하여 데이터 재배포
- 1저자로 미국 천체물리학 저널에 논문 투고, 20회 이상 인용 및 1000회 이상 다운로드 성과

- Radio / Infrared signal processing
- Data Processing
- Statistical Analysis