

허건혁 이력서

010-****-****

girjsgur@naver.com

GeonHyeock

[2024-09-15 수정]

Career

현대 오토에버 - 데이터 엔지니어 (정규직)

2024.09 ~ Now

2024년 09월 09일 데이터분석구축팀 입사

네이버 커넥트재단 - 교육 코치 (프리랜서)

2024.01 ~ 2024.02

부스트클래스 AI 엔지니어 기초 다지기 교육 코치로 활동하면서 20명의 담당 훈련생을 전담하여 학습 코치 및 과제의 코드리뷰 진행

Research Publications

Hyung-Jun Lim, Gye Wan Kim, **Geon Hyeock Heo**, Uidon Jeong, Min Jeong Kim, Dokyung Jeong, Yoonsuk Hyun, Doory Kim†, "Nanoscale single-vesicle analysis: High-throughput approaches through AI-enhanced super-resolution image analysis", Biosensors and Bioelectronics, 263, 116629, 2024

Awards

KSIAM 학술대회 포스터 우수상

2024.05

KSIAM(한국산업응용수학회) 인공지능 연구회 분과에서 Graph Nerual Network를 주제로 포스터 발표를 하여 수상

주제 : Project - Planar Graph Classification : Graph Nerual Network for Graphs

주관 : KSIAM(한국산업응용수학회)

2023 인하 인공지능 챌린지 - 대상

2023.08

추천시스템을 주제로 학부 및 대학원 부문에서 모델링 성능 종합 1등을 하여 수상

주제 : competiton - 2023 인하 인공지능 챌린지 Multi-modal Recommender System

주관 : 인공지능융합연구센터, BK 산업융합형 차세대 인공지능 혁신인재 교육연구단

Projects

Planar Graph Classification : Graph Neural Network for Graphs

2023.12 – 2024.05

(MI Lab 학술 프로젝트 : 개인)

- **문제 정의**
추천 및 화학 구조 등 그래프 구조의 데이터를 분석하는데 큰 성과를 보여준 그래프 신경망이 그래프의 본질적인 특성을 얼마나 잘 이해하는지 의문을 갖고 그래프 이론에서 관심 있는 주제 중 평면그래프 판별에 대하여 실험을 통해 성능을 평가하고자 했다.
- **실험 설계**
노드별 특징으로 그래프 노드의 차수를 활용해 두 가지 임베딩 방법에 대하여 그래프 신경망을 학습해 정확도를 비교 실험하였다.
임베딩 방법1은 노드의 차수에 대하여 랜덤하게 임베딩 하였다.
임베딩 방법2는 노드의 차수와 주기함수를 합성해 임베딩 하였다.
- **가설 검증**
가설 : 임베딩 방법 2는 방법 1과 다르게 주기함수를 통해 차수 간의 관계를 학습하여 학습하지 않은 차수의 노드를 갖는 그래프에 대하여 일반화 성능이 더 좋을 것이라고 생각했다.
검정 : 6가지의 그래프 신경망을 통하여 Accuracy를 측정하여 Paired t-test를 진행해 가설을 검증하였다. P-value 값이 0.026으로 학습하지 않은 새로운 대형 그래프를 판별할 때 임베딩 방법 2가 더 적은 학습파라미터로 더 좋은 일반화 성능을 낼 수 있음을 입증하였다.
- **결론**
추가로 그래프 신경망을 통하여 그래프의 성질을 파악하는 것은 기존의 알고리즘보다 약 1000배 더 빠른 속도로 성질을 파악할 수 있었다. 따라서 다항 시간 안에 해결하지 못하는 채색 문제, 해밀턴 경로 문제 등 NP-Complete 들에 대하여 노드의 임베딩을 주기함수의 합으로 표현되는 푸리에 급수의 형태로 표현력을 키워 기존의 NP-Complete 문제들을 극복하는 시도로 확장이 가능하다고 생각한다.

CDV(나노 규모의 소포체) Detection using MLflow

2023.08 – 2024.03

(MI Lab 외주 프로젝트 : 개인)

- **문제 정의**
한양대학교 화학과 연구실에서 나노 규모의 소포체 위치를 결정하는데 사람이 직접 하기에 많은 시간과 노력이 들어 DBSCAN을 활용하여 소포체를 검출하려 하였으나 서로 다른 물질이지만 하나의 물질로 인식하는 한계가 있었다. 해당 문제를 해결하고자 외주 프로젝트를 진행하게 되었다.
- **수행 과정**
MLflow를 활용하여 변화하는 데이터에 대하여 객체 탐지 모델을 버전관리 하여 학습하고 유지 보수 하였다.
Docker를 활용하여 모델 학습, 모델 배포, Streamlit UI 환경의 3가지 Container를 생성하여 모델을 배포했다.
- **성과**
Streamlit을 활용하여 UI 환경에서 Confidence값과 IOU값의 임계값을 조절함으로써 DBSCAN 알고리즘의 서로 다른 물질을 하나의 물질로 인식하는 문제를 해소하였다.
Ray를 활용한 병렬처리로 단일 이미지 분석 대비 여러 장의 이미지를 한 번에 추론할 때 약 2.25배 빠르게 추론하였다.
YOLO모델을 기준으로 평가 산식 mAP/0.5에 대하여 0.73을 달성하였다.

- **문제 정의**
학원에서 근무하는 조교 및 선생님들이 다양한 학생들의 시험지를 직접 채점하는 것은 많은 시간과 노력이 필요하다는 사례를 접했다. 이를 AI를 활용해서 자동화 하고자 했다. 객관식 문제에 대하여 문제와 정답의 위치를 객체 탐지 모델을 활용하여 결정하고 이를 IOU 기반으로 매칭하여 문제를 채점하였다.
- **수행 과정**
저작권 문제를 피하고 충분한 양의 데이터를 확보하기 위하여 합성데이터를 활용해 데이터를 구축하고 레이블링 하였다. 학습과 검증 데이터로 각 2644, 2712장을 사용하며 데이터 레이블링 가이드를 수정하면서 변하는 형식의 데이터에 대하여 DVC를 활용하여 데이터를 지속적으로 버전 관리하였다. 또한 실제로 학생이 풀이한 데이터를 확보해 210문제에 대하여 테스트 데이터로 활용했다. Docker를 활용하여 Front, Back 부분의 Container를 생성하고 이를 Docker Compose를 활용해 GCP환경에서 모델을 배포하였다.
- **성과**
Onnx를 활용하여 학습한 모델을 932mb -> 319mb로 경량화 하였다.
모의고사 시험지의 경우 좌우 대칭으로 문제가 구성된 특징이 있어 Cutmix 증강에 영감을 받아 Half Cutmix를 개발해 시험지의 좌우 문제를 새롭게 합성하는 데이터 증강과 기존의 데이터 증강을 활용해 객체 탐지 모델의 성능을 mAP 0.780 -> 0.852로 성능을 올렸다.
객관식 문제에 대하여 94%의 정확도로 문제를 채점할 수 있었다.

Education

BS, Inha University Mathematics & Statistics.**2018.03 – 2024.08****Major GPA: 4.37 / 4.5 & GPA: 4.06 / 4.5**

(수석 졸업)

장학금 수혜 내용 : 성적 우수 장학금, 학업 우수 장학금, 국가 우수 장학금(이공계), 김을형 자연과학사랑 장학금

Experience

Mathematical Intelligence Lab 학부연구생**2023.03 – 2023.08**

학술 및 외주 프로젝트, 인공지능 모델링 대회 참가 및 연구실 정기 세미나를 통한 지식 교류

- 수행 프로젝트
 1. Planar Graph Classification : Graph Neural Network for Graphs
 2. CDV(나노 규모의 소포체) Detection using MLflow
- 경진대회 참가
 1. 2024 인하 인공지능 챌린지 : Machine Reading Comprehension
 2. 2023 Samsung AI Challenge : Image Quality Assessment
 3. 2023 인하 인공지능 챌린지 : Multi modal Recommender System
- 교육 조교 활동
 1. 데이터사이언스학과 이산수학 과목 TA
 2. 신입 학부연구생 기초 파이썬 및 AI 모델링 멘토

네이버 부스트캠프 AI Tech CVTrack 연수생**2022.09 – 2023.02**

데이터로부터 문제를 정의하고 AI 모델을 설계하며 서비스화하는 전체 문제 해결 과정을 경험


- 수행 프로젝트
 1. 모의고사 자동 채점 시스템
- 경진대회 참가
 1. 재활용 품목 분류를 위한 Segmentation
 2. OCR 데이터 제작
 3. 재활용 품목 분류를 위한 Object Detection
 4. 마스크 착용 상태 분류


ETC


● Language

2024.06  TOEIC Speaking - Intermediate High

● Certificate

2021.07  컴퓨터활용능력 1급 - 대한상공회의소

2020.07  ADSP(데이터분석준전문가) - 한국데이터산업진흥원

2019.12  지게차 운전 기능사 - 한국산업인력공단