

박 아 정(Ahjeong Park)

다양한 데이터를 학습해 여러 갈래의 서비스에 적용할 수 있는 딥러닝에 매료되어 연구를 진행했습니다. NLP 기반 서비스(예: 기계 번역, 자연어 생성, AI 비서 등)의 한계점을 찾아내고 실용성을 향상해 세상을 바꾸는 AI Service을 제공하고 싶습니다.

함께 성장하는 것을 지향합니다. 이를 위해 프로젝트 진행 시 에러 해결 과정과 구현한 코드를 꼼꼼하게 기록해 팀원에게 전달하는 역량을 쌓고자 노력해 왔습니다. 또한, 소통은 문제 해결의 Key point라고 생각하기에 다양한 구성원과 적극적으로 커뮤니케이션했습니다.

PERSONAL DETAILS

Phone	010-7448-8798
Email	ajeongi59@gmail.com
Github	https://github.com/aaajeong
Linked In	https://www.linkedin.com/in/ahjeong-park-5822b4245
Blog	https://azeomi.tistory.com/

EDUCATIONS

2021.03 ~ 2023.02	숙명여자대학교 일반대학원 IT공학과 석사
2016.03 ~ 2021.02	숙명여자대학교 IT공학과 학사

RESEARCH INTERESTS

- LLM(Large Language Model)
- NLP(Natural Language Processing)
- Deep Learning, Ensemble
- Auto Code Comment Generation

PATENTS

2022.12.12	재귀 신경망 모델의 양상블 방법 및 시스템 Ahjeong Park , Youngmi Park, Chulyun Kim
2022.12.12	트랜스포머 기반의 자연어 주석 자동 생성 방법 및 장치 Youngmi Park, Ahjeong Park , Chulyun Kim

PUBLICATIONS

석사 학위 논문	REGEN: Recurrent Ensemble Methods for Generative Models Ahjeong Park , Youngmi Park, Chulyun Kim
IEEE Access 2023	ALSI-Transformer: Transformer-Based Code Comment Generation with Aligned Lexical and Syntactic Information Youngmi Park, Ahjeong Park , Chulyun Kim
KCC 2022	코드 주석 생성 품질 개선을 위한 AST 순회 방법에 관한 연구 Youngmi Park, Ahjeong Park , Chulyun Kim
KSC 2021	RNN의 새로운 앙상블 기법을 통한 Seq2Seq 모델 성능 개선 Ahjeong Park , Chulyun Kim

RESEARCH

석사 연구 1 2021.07 ~ 2023.12	<p>생성 모델에 적합한 새로운 앙상블 알고리즘 제안(REGEN)</p> <p>Descriptions</p> <ul style="list-style-type: none">• Problem<ul style="list-style-type: none">• 기존 앙상블이 딥러닝 모델을 블랙박스로 간주해 최종 결과만을 취합한다는 점은 길이와 범위에 제한이 없고 출력의 다양성이 높은 생성 모델에 적합하지 않은 한계점 존재• Contribution<ul style="list-style-type: none">• 반복적으로 다음 단어를 예측하며 생성 작업에 적합한 Decoder 모델에 적용 가능한 새로운 앙상블인 REGEN 제안• 앙상블 구성 모델로 Seq2Seq, Transformer를 활용해 각 모델의 Decoder가 생성 단계마다 다음 생성의 영향을 미치며 모델 간 합의 진행하는 알고리즘 제안 - Consensus, Survival Ensemble• Method<ul style="list-style-type: none">• Bagging, Voting(Soft Voting, Hard Voting)• Result<ul style="list-style-type: none">• 성능 평가 : 생성 Task로 기계 번역 실험을 진행, REGEN은 단일 구성 모델뿐만 아니라 기존 앙상블보다 성능이 우수함을 확인• Seq2Seq 모델 적용 : Single(19.56 BLEU) < 기존 앙상블(20.64 BLEU) < REGEN(22.71 BLEU)• Transformer 모델 적용 : Single(24.93 BLEU) < 기존 앙상블(24.02 BLEU) < REGEN(25.15 BLEU)• 적용 가능 모델 : Auto-Regressive Model(GPT, ELMO 등) <p>Achievement</p> <ul style="list-style-type: none">• 석사 학위 논문• 국내 특허 출원 - '재귀 신경망 모델의 앙상블 방법 및 시스템'
------------------------------	--

석사 연구 2	딤러닝 기반 소스 코드 자연어 주석 자동 생성 연구
2022.01 ~ 2023.02	<div><div>Introduction</div><div><ul style="list-style-type: none">한국콘텐츠진흥원 정부연구과제 - 공동 연구개발기관으로 참여과제명 : '공정한 SW 저작권 거래 및 유통 생태계 지원을 위한 저작권 응용 기술 개발'SW 규모가 커짐에 따라 소스 코드 자연어 주석 자동 생성 기술의 필요성 증가</div><div>Descriptions</div><div><ul style="list-style-type: none">Problem<ul style="list-style-type: none">기존 소스 코드 주석 생성에 필요한 syntactic 데이터는 정보손실 및 소스 코드와 Aligned 되지 않는 한계점 존재Contribution<ul style="list-style-type: none">소스 코드 자연어 주석을 자동으로 생성하는 딤러닝 기반 모델인 ALSI-Transformer와 소스 코드의 syntactic 정보 데이터인 CAT(Code-Aligned Type)을 새롭게 제안Method<ul style="list-style-type: none">Transformer, CNN, AST(Abstract Syntax Tree)Result<ul style="list-style-type: none">소스 코드 주석 생성에 필요한 기존 데이터 세트의 한계점을 보완하고 모델 구조를 개선해 정확성이 높고 크기가 작은 주석 생성 Task의 SOTA 모델 완성</div><div>What did I do</div><div><ul style="list-style-type: none">공동 연구개발기관 실무 담당자 및 참여 연구원주석 생성 관련 연구 조사 및 SOTA 달성 모델 분석새로운 데이터 타입 CAT 설계 및 구현해 총 465,812개 데이터 추출소스 코드의 멀티 모달 (Lexical & Syntactic)을 결합해 1개의 인코더로 처리하기 위한 6가지 방법을 설계 및 실험<ul style="list-style-type: none">Average(50.75 BLEU) → Concat(51.41 BLEU) → Addition(53.21 BLEU) → Gate Network(53.80 BLEU)Baselines와 비교 실험 진행해 SOTA 확인 : Baseline(49 BLEU) → ALSI-Transformer(53.80 BLEU)논문 작성</div><div>Achievement</div><div><ul style="list-style-type: none">정부 과제 1차 연도 연구 성과 달성IEEE Access, 2023 게재국내 특허 출원 - '트랜스포머 기반의 자연어 주석 자동 생성 방법 및 장치'</div></div>

PROJECTS

2021.08 ~ 2021.11	NH투자증권 주식 보유기간 예측 및 서비스 제안
참여인원: 3명	
	Descriptions
	<ul style="list-style-type: none">제공된 국내 고객 및 잔고 이력을 분석해 고객의 주식 보유기간을 예측하고 AI/빅데이터 분석 기술을 활용한 서비스 아이디어를 제안하는 대회입니다.
	What did I do
	<ul style="list-style-type: none">국내 고객 및 주거래 계좌 정보 정형 데이터 분석하여 연령, 자산 구간 라이프스타일 등에 따른 보유 기간 특성 확인국내 주식 잔고 이력 정형 데이터 분석하여 고객별 주식 판매 횟수, 매수 및 매도 금액 평균, 과거 해당 주식 보유기간 feature 추출고객의 투자 성향별 주식 보유 기간이 다름에 기반하여 투자 성향별 보유기간 feature 추출해 feature importance 증가, 최종적으로 RMSE를 97 → 60으로 감소외부 데이터(뉴스심리지수, 공휴일 정보 등)를 활용한 주식 관련 feature 추출머신 러닝: XGBoost(XGBRegressor)로 총 49개 feature 활용해 학습서비스 아이디어 제안: 고객의 투자성향 및 보유기간 등 feature를 기반으로 하는 투자 MBTI 생성문서화: Notion 활용매주 금요일 화상회의 진행
	Tech
	<ul style="list-style-type: none">Python, Pandas, Github
	Achievement
	<ul style="list-style-type: none">2021 NH투자증권 빅데이터 경진대회 '본선' 진출

PROJECTS

2019.11 ~ 2020.07	<div><div>Samsung partnership Bixby Project</div><div><div>Descriptions</div><div><ul style="list-style-type: none">(주) 라이온브리지테크놀로지스코리아 Lionbridge AI에 근무하면서 진행한 프로젝트입니다. Bixby의 연락처 기능을 담당하는 Contact Capsule에서 발화에 따른 빅스비 동작 모델링을 담당했습니다.</div><div><div>What did I do</div><div><ul style="list-style-type: none">연락처 기능 관련 학습 데이터를 관리 및 생성 (ex. 홍길동 연락처 저장하고 문자 보내줘)Problem<ul style="list-style-type: none">자주 발생하는 발화 오류 데이터 패턴 확인빅스비에 별명(nickname)으로 호출했을 때 저장된 별명을 검색하지 않고 별명과 동일한 문자열을 가진 다른 연락처 호출하는 오류 발생Contribution<ul style="list-style-type: none">별명을 태깅할 수 있는 발화 데이터 추가 및 별명 호출 동작 모델 추가 (ex. 홍길동에 내 단짝이라고 별명 추가해 줘, 내 단짝에게 전화해 줘)Device에 따른 발화 동작 모델 수정</div><div><div>Tech</div><div><ul style="list-style-type: none">Javascript(기능구현), GitLab, Bixby Developer Studio</div><div><div>Achievement</div><div><ul style="list-style-type: none">빅스비 마켓플레이스에 연락처-별명 캡슐 배포</div></div></div></div></div></div>
2019.09 ~ 2019.10 참여인원 20명	<div><div>2019 공개 SW 오픈소스 컨트리뷰톤</div><div><div>Descriptions</div><div><p>비전공자도 딥러닝에 쉽게 접근할 수 있는 Keras용 원스톱 객체 감지 API 제작 프로젝트입니다. 오픈 소스에 기여를 하는 경험을 할 수 있었습니다.</p></div><div><div>What did I do</div><div><ul style="list-style-type: none">Problem<ul style="list-style-type: none">Tensorflow Object Detection API의 결과 bounding box 생성 시 이미지 사이즈 미고려로 인한 가독성 저하Contribution<ul style="list-style-type: none">Output 후처리 - 이미지 픽셀 크기 값에 따른 bounding box 두께 및 크기 조절 기능 추가해 가독성 향상</div><div><div>Tech</div><div><ul style="list-style-type: none">Python, Tensorflow, Keras, Docker, Github</div><div><div>Achievement</div><div><ul style="list-style-type: none">2019 공개SW 컨트리뷰톤 '최우수상(정보통신산업진흥원장상)'</div></div></div></div></div></div>

PROJECTS

2019.03 ~ 2019.08	딥러닝 이미지 인식을 활용한 축구 하이라이트 영상 추출 프로그램 제작
참여인원 2명	
	Descriptions <ul style="list-style-type: none">• 학사 졸업프로젝트로, Tensorflow Obejct Detection API를 활용해 축구 경기에서 하이라이트 장면들을 뽑아 시간과 함께 시청자가 볼 수 있는 서비스를 제공하는 프로그램입니다.
	What did I do <ul style="list-style-type: none">• 문제 정의 : 축구 영상 속 중요한 장면(골, 프리킥, 좋은 패스 등)은 하이라이트로 판단되며 로고 사이에 리플레이된다는 특징 확인• 데이터 수집 : OpenCV 활용해 크롤링 된 중계 영상의 로고 이미지가 속한 프레임 추출 & 1,399개 로고 이미지 라벨링• 이미지 학습 : pretrained 모델(SSD) 사용해 학습 진행• 하이라이트 추출 프로그램 제작 : PYPI, MoviePY를 활용해 직접 하이라이트를 추출할 수 있는 서비스 제작
	Tech <ul style="list-style-type: none">• Python, Tensorflow, PYPI, OpenCV

OTHER EXPERIENCE

2022.08 ~ 2022.11	NLP(Natural Language Processing) 논문 스터디
	Descriptions <p>자연어처리에서 중요한 논문을 읽고 논문 리뷰 및 코드 구현과 발표를 기록하는 스터디입니다.</p>
	What did I do <ul style="list-style-type: none">• 스터디 진행 및 자연어처리 중요 논문 정리 담당
	Achievement <ul style="list-style-type: none">• 논문 스터디 발표자료 공유 https://www.notion.so/NLP-Study-65051e80de9a4d05919d6740f1f4931a?pvs=4• 자연어처리 논문 스터디 동영상 유튜브 업로드 https://youtube.com/playlist?list=PLFYmifvO4WT8iqxv1cMqdNH1-xU7Dej9v

SKILLS

Language

- Python : 주로 사용하는 언어입니다. 파이썬의 문법을 잘 활용할 수 있으며 제공되는 라이브러리 문서를 참고해 목표하는 기능을 구현할 자신이 있습니다. Python을 활용한 크롤링, 딥러닝, IoT 프로젝트 경험이 있습니다.
- Javascript : 간단한 기능을 구현할 수 있습니다. Javascript를 활용하는 빅스비 프로젝트에 참여한 경험이 있습니다.
- Java : 간단한 기능을 구현할 수 있습니다. 해커톤 참가 및 학사 프로젝트를 통해 모바일 애플리케이션 제작을 경험했습니다. 하지만 멀티스레드 및 캡슐화 등에 대해 능숙하지 않습니다.

Deep Learning Framework

- Pytorch : 공식 문서를 보고 모델링 할 수 있고 텐서플로우 및 케라스 언어를 파이토치로 바꾸는 프로젝트를 진행한 경험이 있습니다. 석사 연구를 파이토치로 진행했습니다.
- Tensorflow : 간단한 모델(CNN, Seq2Seq)을 구현할 수 있습니다.
- Keras : 간단한 모델(CNN)을 활용한 딥러닝 앙상블 프로젝트를 진행한 경험이 있습니다.

Database

- MySQL : Java와 MySQL을 연동해 간단한 렌터카 프로젝트를 진행한 경험이 있습니다.

ETC

- Git : Pull Request와 Pull/Push/Merge를 자유롭게 할 수 있습니다.

ETC

- **개인의 성장**
 - 학과 성과와 CS 개발 역량을 키우기 위해 여러 해커톤에 도전했고 새로운 기술을 공부하기 위해 AI 컨퍼런스에 적극적으로 참여합니다. 최근 LLM 응용 관련 공부를 하며 이를 통해 개인의 삶과 커리어를 성장하고자 하는 열정을 가지고 있습니다.
- **봉사 정신**
 - 인도네시아 현지 교육 ICT 봉사단 및 일본 규슈대 ICT 프로그램에 자발적 참여를 통해 내가 배운 언어 및 개발 기술을 해외 학생에게 가르치고 혼자가 아닌 함께 성장하는 봉사 정신을 함양하고자 합니다.
- **책임감**
 - 프로젝트 진행 시 능숙한 분야가 아닌 상황에서 주저하지 않고 관련 기술 조사 및 스터디를 적극적으로 주도해 좋은 성과를 얻었습니다.
- **커뮤니케이션**
 - 연구 및 논문 작업을 했던 연구원, 기업과의 협업을 통해 소통의 중요성을 배웠습니다. 2주마다 중간 점검을 통해 참여 기관끼리 진행 상황을 확인했고 노선으로 업무 분배, 화상회의를 적극 활용했습니다. 특히 참여 기관이 가지고 있는 지식 도메인이 다르다는 점에 집중하여 작업물을 최대한 구체적, 시각적으로 표현하여 의견을 전달했고 객관적인 지표를 제시해 연구원을 설득했습니다.