

조동빈

Data Scientist

Birthday	1993.1.5
Email	ken.dbc.career@gmail.com
Mobile	010-4520-7849
Address	경기도 남양주시 오남읍 양지로 240 번길 37

소개

공업/이미지/텍스트/음성/그래프/유저 데이터들을 다루며 Data Scientist 역량을 키워왔습니다. 데이터 분석, 실험 설계, ML modeling, MLOps 까지의 전 과정을 즐깁니다. 현재 추천 시스템 연구/개발자 직무를 구직중입니다.

기술 스택 (숙련도를 기준으로, 가장 높은 기술부터 기입)

구분	Skill
Programing Languages	Python, C, C++, CUDA, JavaScript, Shell Scripting, Make
Framework / Library	PyTorch, TensorFlow, Scikit-learn, Pandas, cuDNN, cuBlas(LT), TensorRT
MLOps / DevOps	Docker, Docker-compose, Kafka, Kubernetes, Nginx
Server	Django, Flask, FastAPI
DB	MySQL, PostgreSQL, MongoDB, DynamoDB
Etc	Ubuntu, AWS

학력 사항

재학 기간	학교명	학과명	연구실	학점	상태
2020.03 ~ 2022.02	한양대학교 대학원	컴퓨터소프트웨어학과	ML System Lab.	4.13/4.5	졸업(석사)
2012.03 ~ 2018.08	강원대학교	산업공학과	기술경영 연구실	3.45/4.5	졸업(학사)
2008.03 ~ 2011.02	보성고등학교	-	-	-	졸업

경력 사항 / 총 1년 2개월

재직 기간	회사명	부서 및 R&R	근무 형태
2022.06.22 ~ 2022.12.31	엔디파인드	개발팀 / AI 개발자	정규직
2019.12.26 ~ 2020.02.29	카카오	추천팀 / 추천 알고리즘 개발자	인턴
2018.11.05 ~ 2019.04.22	한빛소프트	AR 팀 인공지능 파트 / 전임연구원	정규직

기타

자격증 및 교육

ADsP, 데이터분석전문가	2017.12	취득
	2019.07 ~ 2019.12	삼성청년 SW 아카데미 이수

어학

OPIc	2022.01.19	Intermediate High
------	------------	-------------------

URL

<https://github.com/jarvis08>
<https://jarvis08.github.io>

논문

학회명	Accept	논문명
Briefings in Bioinformatics	2022.11	RAMP: Response-Aware Multi-task Learning with Contrastive Regularization for Cancer Drug Response Prediction
ICEIC	2021.12	Quantization training with two-level bit width

프로젝트 목록

진행 기간	기간	소속	주요 기술	내용	성과
2022.11 ~ 2022.11	1m	언디파인드	OCR	경기 결과 자동 기록 시스템 개발	배포
2022.09 ~ 2022.10	2m		EDA, Reco.(CF, MF)	대회 규정 추천 시스템 개발/배포/운영	배포/운영
2022.06 ~ 2022.08	3m		BERT, Starspace	FAQ Chatbot 시스템 개발/배포/운영	배포/운영
2020.03 ~ 2022.11	2.5y	ML System Lab.	Node2Vec, GCN, GTN	Graph Embedding Generation 연구	2022 BIB
2020.03 ~ 2022.02	2y		Quantization(QAT)	DNN Model Quantization 연구 1	논문 작성중
2021.10 ~ 2021.12	3m		Quantization(QAT)	DNN Model Quantization 연구 2	2022 ICEIC
2020.07 ~ 2020.09	2m		Seq2SQL	Korean AI Assistant 연구	-
2020.02 ~ 2020.02	2w	카카오	Word2Vec, Ranking	만화 추천 알고리즘 개발	-
2020.01 ~ 2020.02	2m		MAB, MF, Ranking	자동차 동영상 추천 알고리즘 개발	CTR 상승
2019.01 ~ 2019.04	4m	한빛소프트	Tacotron	(한글) 다중화자 음성 합성 모듈 개발	-
2018.11 ~ 2018.12	2m		Word2Vec, Seq2Seq	(영어) 음성 Chatbot 개발	사내 시연

1. 2022.06 ~ 2022.12 (6개월) 언디파인드 / 개발팀 / AI 개발자 / 사원

* 회사 소개 : Undefined 는 iScrim 이라는 플랫폼을 개발하는 회사이며, iScrim 은 eSports 대회를 쉽게 개최/운영할 수 있도록 돕는 플랫폼입니다.

A. 2022.11 ~ 2022.11 (1개월) / 경기 결과 자동 기록 시스템

기술 스택	Python, Tesseract(OCR module), Docker, Flask, AWS
인력구성 및 기여도	회사 내 유일 AI 개발자 사업/기획/프론트엔드/백엔드 직원들과 협업하여 개발 및 배포를 진행
주요업무 및 상세역할	[전략 수립 -> 데이터 수집 및 모델 학습 -> 배포] eSports 경기 결과 snapshot 으로 부터 username, 경기결과 등을 자동 기록하는 시스템 개발
성과/결과	요구사항 반영 및 지속 가능한 pipeline 개발/배포

B. 2022.09 ~ 2022.10 (2개월) / 대회 규정 추천 시스템

기술 스택	Python, Pandas, Alternative Least Squares, Docker, Flask, DynamoDB, AWS
인력구성 및 기여도	회사 내 유일 AI 개발자 사업/기획/프론트엔드/백엔드 직원들과 협업하여 개발/배포/운영을 진행
주요업무 및 상세역할	[데이터 수집 -> EDA -> 전략 수립 -> 모델 학습 및 최적화(속도) -> 배포/운영] EDA 를 거쳐 데이터 특성을 파악한 후, 적합한 모델을 선택하여 개발/학습/최적화 후 배포/운영
성과/결과	요구사항 반영 및 지속 가능한 pipeline 개발/배포/운영

C. 2022.06 ~ 2022.08 (3개월) / FAQ Chatbot 시스템

기술 스택	Python, BERT, Starspace, Docker, Flask, DynamoDB, AWS
인력구성 및 기여도	회사 내 유일 AI 개발자 사업/기획/프론트엔드/백엔드 직원들과 협업하여 개발/배포/운영을 진행
주요업무 및 상세역할	[전략 수립 -> 데이터 수집 -> 모델 개발/학습/최적화 -> 배포/운영] 적합한 chatbot pipeline 을 설계하고, 자사 데이터를 전처리하여 학습/최적화 후 배포/운영
성과/결과	요구사항 반영 및 지속 가능한 pipeline 개발/배포/운영

2. 2020.03 ~ 2022.02 (2년) 한양대학교 대학원 / ML System lab. / 석사과정

* 회사 소개 : MLsys 연구실은 딥러닝 모델들의 학습/추론 과정을 system 적으로 최적화하는 기법들을 연구합니다. 연구 분야들로 분산학습과 GPU 연산 최적화가 있습니다. 저는 주로 Quantization 과 Graph Embedding 학습과 같은 단일 머신 학습의 최적화를 연구했습니다.

A. 2020.03 ~ 2022.12 (2년 6개월) / Graph Embedding Generation

기술 스택	<ul style="list-style-type: none"> - OS: Ubuntu - Language: Python, Cython - Framework: PyTorch, TensorFlow2 - Library: Scikit-learn, Node2Vec, Pandas
-------	--

인력구성 및 기여도	<ul style="list-style-type: none"> - 항암제와 세포의 반응성을 연구하는 프로젝트를 세 연구실이 협업하여 진행 - 본인의 연구실은 node들의 embedding vector를 학습하는 연구를 수행 - 연구실 내의 참여 연구원은 본인 1명
주요업무 및 상세역할	<ul style="list-style-type: none"> - Negative Sampling(NS)을 개선한 Response Aware NS를 개발 - Embedding vector의 quality를 측정하는 metric인 Embedding Score 개발 - Cython으로 모델 학습 속도 최적화 - Node2Vec, GCN, GTN 등에 개발하여 실험 - 논문 작성 및 revision 진행
성과/결과	2022년 12월, Briefings in Bioinformatics 저널에 published (https://academic.oup.com/bib/advance-article-abstract/doi/10.1093/bib/bbac504/6865135?redirectedFrom=fulltext&login=false)

B. 2020.03 ~ 2022.02 (2년) / DNN Model Quantization - 1

기술 스택	<ul style="list-style-type: none"> - OS: Ubuntu - Language: Python, C - Framework: PyTorch, Darknet - Library: Scikit-learn, Pandas
인력구성 및 기여도	<ul style="list-style-type: none"> - 2021년에는 단독 진행 - 2022년부터 후배 석사과정 학생 2명을 리드 - 개발을 포함한, 모든 작업을 주로 진행
주요업무 및 상세역할	<ul style="list-style-type: none"> - Google's Quantization 기술 follow up - Statistics 관점에서의 개선점 파악 - Framework를 이용한 코드 및 개선점 적용
성과/결과	현재 후배 학생들에게 조언을 주고 있으며, 2023 ICML 논문 제출 예상

C. 2021.10 ~ 2021.12 (3개월) / DNN Model Quantization - 2

기술 스택	<ul style="list-style-type: none"> - OS: Ubuntu - Language: Python - Framework: PyTorch
인력구성 및 기여도	<ul style="list-style-type: none"> - 후배 석사과정 학생 2명을 리드 - 연구에 필요한 코드 개발 - 논문 outline 작성
주요업무 및 상세역할	<ul style="list-style-type: none"> - 논문 분석 및 baseline(Quantization Aware Training, QuantNoise)의 한계점 도출 - 개선점 파악 및 코드 구현 - 논문 outline 작성
성과/결과	2022년 ICEIC에 published (https://ieeexplore.ieee.org/document/9748737)

D. 2020.07 ~ 2020.09 (2 개월) / Korean AI Assistant

기술 스택	<ul style="list-style-type: none"> - OS: Ubuntu - Language: Python, JavaScript - Framework: TensorFlow2
인력구성 및 기여도	<ul style="list-style-type: none"> - Stansford OVAL Lab.과 협업 - 동료 석사과정생과 함께 작업을 진행 - AI Assistant 인 Almond 의 기술들을 분석하고, KR model 연구 과정을 설계
주요업무 및 상세역할	<ul style="list-style-type: none"> - Almond 기술의 논문 research - AI Assistant pipeline 분석 - KR Seq2SQL model 데이터 생성
성과/결과	외부 사정으로 인해 연구 중단

3. 2019.12 ~ 2020.02 (2 개월) 카카오 / 추천팀 / 추천 알고리즘 개발자 / 인턴

* 회사 소개 : 카카오 추천팀은 추천 알고리즘이 필요한 서비스들의 추천 시스템을 총괄하여 개발해 주는 부서입니다.

A. 2020.02 ~ 2020.02 (2 주) / 만화 추천 알고리즘 개발

기술 스택	Python, Word2Vec, Alternative Least Squares, Ranking Algorithm, Docker, Kubernetes, Kafka, Ubuntu
인력구성 및 기여도	<ul style="list-style-type: none"> - 인턴 팀원 2 명과 함께 진행 - 데이터 분석과 Project Manager 역할로 기여
주요업무 및 상세역할	<ul style="list-style-type: none"> - 데이터 수집/분석 - 실험 설계/진행/분석
성과/결과	최종 점검에 대한 PT 제작 및 발표

B. 2020.01 ~ 2020.02 (2 개월) / 자동차 동영상 추천 알고리즘 개발

기술 스택	Python, Multi-armed Bandit, Item2Vec, Word2Vec, Alternative Least Squares, Ranking Algorithm, Docker, Kubernetes, Kafka, Ubuntu
인력구성 및 기여도	<ul style="list-style-type: none"> - 인턴 팀원 2 명과 함께 진행 - 데이터 분석과 Project Manager 역할로 기여
주요업무 및 상세역할	<ul style="list-style-type: none"> - 데이터 수집/분석 - 실험 설계/진행/분석 - 중간/최종 점검에 대한 PT 제작 및 발표
성과/결과	Click Through Rate(CTR) 상승

4. 2018.11 ~ 2019.04 (6개월) 한빛소프트 / AR 팀 / 인공지능파트 / 전임연구원

* 회사 소개 : 당시 사내 AI 도입이 초기단계로, 인프라 역시 초기 단계였습니다. 저는 음성/텍스트 데이터를 이용한 챗봇 개발 및 음성합성 모듈 연구를 진행했습니다.

A. 2019.01 ~ 2019.04 (4개월) / (한글) 다중 화자 음성 합성 모듈 개발

기술 스택	Python, TensorFlow1, Multi-speaker Tacotron, Ubuntu
인력구성 및 기여도	단독 진행
주요업무 및 상세역할	<ul style="list-style-type: none"> - Speech dataset 확보 및 전처리 자동화 (하나의 긴 audio file 을 학습용의 무수히 많은 짧은 file 들로 분할) - 모델 탐색/개발/학습
성과/결과	사내 prototype 데이터셋 및 모델 개발

B. 2018.11 ~ 2018.12 (2개월) / (영어) 음성 Chatbot 개발

기술 스택	Python, TensorFlow1, Word2Vec, FastText, Seq2Seq, Attention Mechanism, Ubuntu
인력구성 및 기여도	<ul style="list-style-type: none"> - 인공지능 파트, 파트장의 리드 하에 작업 - 기존 prototype 모델의 데이터셋과 모델의 퀄리티 개선 model 개서
주요업무 및 상세역할	<ul style="list-style-type: none"> - 기본적인 NLU, NLG 모델을 사용하는 기존 모듈들의 최신화 (W2V -> FastText), (Seq2Seq -> Seq2Seq + Attention Mechanism) - 기존 dialogue 데이터셋의 전처리 과정을 개선(label 재배치 포함) - Metric 변경으로 모델 평가 모듈 개선
성과/결과	NLU/NLG 모델 성능 향상, 사내/사외 시연 진행