

이 력 서

기본정보

- 이름 : 김영훈
- 생년월일 : 1985 년 5 월 1 일(만 38 세)
- 성별 : 남
- 주소 : 서울시 강서구 마곡중앙 5 로 6
- 이메일 : genholy77@gmail.com
- 연락처 : 010-9004-6355



핵심역량

- AI 관련 데이터 수집, 분석
- Python
- Linux
- Excel
- English : TOEIC 775 점(취득일자: 2023.09.24)
- Github link : <https://github.com/realhoon>

학력사항

송실대학교 정보과학대학원 인공지능학과 졸업 (본교:야간) 석사	2021.03 ~ 2023.08
한국외국어대학교 경제학과 졸업 (본교:주간) 학사	2005.03 ~ 2013.02
부산중앙고등학교 인문계 졸업	2001.03 ~ 2004.02

경력사항

디플렉스씨앤씨 (KT DS 협력사)	2023.07~2023.12 (5 개월)
에스맨파워 (KT DS AICC 서비스 1 팀)	2021.08~2023.07 (1 년 11 개월)
CJ 바이오사이언스 (재경팀)	2017.11~2021.03 (3 년 5 개월)

병역사항

- 만기제대(2006.08~2008.07, 의무경찰)

교육과정 및 기타사항

교육과정명	교육기관	교육기간	교육시간	수료여부
파이썬 빅데이터 분석	코드랩아카데미학원	2022.03 ~ 2022.06 (4 개월)	112 시간	정상수료

디플렉스씨앤씨

2023.07~2023.12 (5 개월)

프리랜서, KT DS 협력사, 퇴사사유 : 계약만료

KT - NH 투자증권 지식구축업무(TA, Text Analysis 엔진 데이터 구축업무)

고객의 의도 파악을 위한 녹취 기반의 상담데이터 분석 기술, 문장/발화 단위의 상세의도분석 제공 및 자연어 처리를 위한 음성데이터 자산화를 통한 비즈니스 인사이트로 연계

1. **업무 환경** : 딥러닝 기술 기반의 음성인식 이후 자연어처리 단계의 TA(Text Analysis) 엔진 시스템
2. **담당 업무** : 증권사 이용고객의 녹취 기반 상담 데이터 분석 및 자연어 처리 모델 학습을 통해 고객 의도 분류 및 주요 키워드 인식을 향상으로 상담 서비스 품질 제고
 - 상담유형 중 학습량이 부족한 문장 추가
 - 원활한 개체명 인식을 위한 토큰화, 형태소 분석 및 추가 작업
 - 표제어, 유의어, 사용자 사전 추가 작업
 - Linux 로그 데이터 분석 : 미추출문장 분석 및 콜 ID 확인, 자연어 처리 모델 학습 과정 확인
 - 주차별 학습데이터 기반으로 모델링 진행 후 검증 및 테스트 성능 수치 확인 및 배포(웹환경)
 - 모델 검증 및 성능 테스트에 필요한 데이터셋 구축
3. **분석 도구** : Visual Studio(md), Linux, Excel

에스맨파워

2021.08~2023.07 (1 년 11 개월)

파견직 사원, KT DS AICC 서비스 1 팀, 퇴사사유 : 계약만료

KT AI Contact Center 보이스봇 음성기반 대화 인터페이스

AI 와 음성대화로 상담업무를 효율화하고자 하는 고객사를 대상으로 문의 고객의 콜인입에 대해 예약 및 안내 업무를 지원하는 음성통화 기반의 AI 가상 상담 솔루션

1. **업무 환경** : 음성인식 단계의 LM(Language Model) 언어모델(통계-확률 기반의 N-gram)
2. **담당 업무** : 다양한 고객이 보이스봇 시스템을 통하여 해당 서비스를 편리하게 이용할 수 있도록 음성인식 및 처리 기술에 관련한 언어모델 생성(데이터 보정) 및 서비스 품질 제공
 - 서비스 개시 이후, 녹취 음성파일(pcm)과 정답지(STT 결과)를 바탕으로 인식을 측정
 - 오인식/미인식 어휘 및 문장 추출 후 언어모델 학습
 - 고객 VOC 발생 시, 데이터 축적 경로 파악하여 오인식/미인식 사례 분석 후 대응 지원
 - 카테고리, 패턴 작업을 통해 말뭉치(코퍼스) 생성
 - 숫자, 영문 인식을 위한 이중전사 작업으로 학습데이터 생성
 - 모델 서빙 및 배포 : 음성인식 서버 시스템 구성도 이해를 바탕으로 프로세스(PNR) 작업을 통한 배포 및 도메인 모델 별 연결 서버 관리
3. **분석도구** : Linux, Excel

정규직 대리, 재경팀, 퇴사사유 : 업직중 전환

재무회계, 자금관리, 결산보고

- 자금관리 : 외부 투자금 사용용도별 관리 및 보고자료 작성
- 매입관리 : 재고자산/매출원가, 기타 매입비용 관리
- 회계감사자료 작성 : 분기/반기/연간 감사자료 작성
- 한국거래소 상장 심사자료 작성 : 연도별 비교식 자료 작성

프로젝트**KT NH 투자증권 음성데이터 자산화 사업****2023.07~2023.12 (5 개월)**

프로젝트 기관 : NH 투자증권

수행역할 : TA 엔진 데이터 구축

1. 주요 성과 : 의도 추출율 89% > 95%
2. 협의 내용 : KT 연구원과 모델 성능 및 추출율 향상을 위한 회의 진행 후, 의도 분류시 confusion(혼동이 오는 부분)과 fallback(미추출) 부분을 집중적으로 학습하기로 방향 설정
3. 주요 역할
 - 고객 상담 유형 분류 시, AI 의 Confusion 부분이 있는 문장들을 대거 수집하여, 본래 고객 의도에 맞게 재분류
 - 일별 미추출 fallback 문장 수집 후 의도별로 문장 추가
 - 학습 데이터 생성 후 모델 학습 및 배포 진행

KT AI CONTACT CENTER B2B / B2G 구독형 서비스**2021.08~2023.07 (1 년 11 개월)**

프로젝트 기관 : VIPS 외 다수

수행역할 : LM 언어모델 데이터 고도화

1. 주요 성과 : VIPS 보이스봇 대화 인터페이스 서비스 전국 지점 확대(서비스 전 인식률 75% > 서비스 개시 이후 안정화 97%)
2. 협의내용 : 대화 시나리오 저작팀과 협의 후 필요한 학습 문장 확인
3. 주요역할 : 대화 시나리오 분석 후 식당 서비스에 관련된 발화문장(말뭉치:corpus) 추가적 수집 후 학습데이터 생성, 언어모델링 작업 및 배포
4. 그 외 서비스 회사
 - A/S(린나이, 현대 HT)
 - 차량번호 및 차종(미래플러스, 다산콜센터, 하모니렌트카)
 - 엘리베이터 비상신고 및 점검(TKE)
 - 요식업(라그릴리아, 풀바셋, 부베트) 등

자기소개

프로젝트 수행 내용 및 본인의 역할

1. 프로젝트 : KT AICC B2B 빙스 보이스봇 대화 인터페이스 서비스

- 성과 : 인식률 75% > 97%, 빙스 보이스봇 서비스 전 지점 확대
- 역할 소개

KT DIGICO AICC(Artificial Intelligence Contact Center) B2B 구독형 서비스 중 VIPS-빙스 고객센터 상대로 식당 예약과 관련된 보이스봇을 구축하는 프로젝트가 있었습니다. 저는 음성인식 (STT : Speech to Text) 부분에서 LM(Language Model : 언어 모델) 학습을 맡았습니다. LM 은 하이브리드 음성인식기에서 음향모델인 AM(Acoustic Model 음향모델)과 함께 음성인식 (STT) 단에서 중요한 역할을 담당합니다. AM 이 소리 자체를 인식하는 역할을 한다면 LM 은 언어적인 면에서 인식능력을 담당한다고 할 수 있습니다. 예를 들어, 사람이 '카' 라는 소리를 들었을 때 LM(언어 학습적 측면)에서 'car'라는 단어가 학습되지 않았다면 자동차를 떠올리지 못합니다. LM 은 이러한 역할을 담당하면서 보이스봇의 성능을 향상 시킵니다.

보통 상용으로 서비스를 제공하려면 인식률은 90% 초반은 나와야 합니다. 서비스 초창기라 제가 LM 인식률을 측정했을 때는 70% 대로 나왔습니다. 저는 인식률을 향상시키기 위해 고객센터에서 제공한 시나리오를 분석하고 저작팀에서 요청한 문구들을 수집했습니다. 시나리오 분석을 통해 서비스가 제공되는 흐름을 파악하고, 저작팀에서 요청하지 않은 예상 고객 발화도 추가하며 코퍼스(말뭉치)를 생성했습니다. 아울러 첫 서비스로 제공된 매드포갈릭이 빙스와 도메인이 유사하다고 판단하여 매드포갈릭에 사용된 언어모델의 코퍼스로 수집하여 함께 모델링을 진행하였습니다. 기본적인 예약 관련 문구, 날짜, 시간, 인원수 등은 매드포갈릭 코퍼스로 커버가 가능했고, 기타 기존에 없던 대표메뉴/아기의자/유모차 등 관련 문구들은 저작팀 요청 및 시나리오 분석을 통해 인식이 가능하게끔 하였습니다. 결국 빙스에 초기에 제공된 서비스 인식률은 90 점대 초반을 이루어 기존보다 높은 인식률로 서비스를 제공하였습니다. 이후 품질유지를 위해 실제로 인입되는 음성 및 전사파일을 분석하며 오인식 개선 작업을 꾸준히 하여 인식률을 97%로 끌어 올렸고, 다른 팀의 적절한 시나리오 구성변경과 음성합성 (TTS) 발화속도 조절을 통해 보다 신속하고 간결하며 정확한 서비스를 제공할 수 있었습니다. 빙스 고객센터는 서비스에 만족하고 하나의 회선을 50 개로 늘리고 전국 지점에서 이 서비스를 이용하였습니다. 제가 생성한 언어모델은 '예약 01'이라는 LM 으로 구독형 서비스에 대표적인 기본 LM 으로 자리 잡았고, 이후 이 서비스를 기점으로 다른 도메인에도 LM 을 보다 확장하게 되는 출발점이 되었습니다.

2. 프로젝트 : KT NH 투자증권 음성데이터 자산화 사업

- 성과 : 의도 추출율 89% > 95%
- 역할 소개

KT 의 또다른 사업으로 NH 투자증권 음성데이터 자산화 프로젝트로, 음성인식 (STT) 다음 단계에서 이루어지는 자연어처리 (NLP : Natural Language Processing) 부분에서 고객 발화의 의도를 분류하고 개체명 인식을 향상시키는 지식구축 역할을 담당하게 되었습니다. 예를 들어, 고객이 '삼성전자 주식 300 주

매수해주세요' 라고 발화한다면 'stock/종목명-삼성전자', 'share/주식수-300 주'라는 개체명 인식과 더불어 '매수_요청'이라는 의도분류 작업이 자연어처리 (NLP)단계에서 필요합니다. 하루에 15 만 정도의 콜이 인입되는데 프로젝트 수준은 이 모든 콜들을 전수로, 그리고 정확하게 의도에 맞는 추출은 못하는 상황이었습니다. 우선은 KT 연구원과의 회의 결과를 바탕으로, 분류 추출률을 향상시키기 위해 fallback 으로 떨어지는 미추출 문장들을 의도에 맞게 태깅하여 학습데이터에 추가하였습니다. 그리고 검증 및 테스트 결과(precision, recall, f1-score)의 수치와 에러 결과를 분석하여 수치가 떨어지는 의도유형과 개체명들을 로그데이터에서 추출하여 학습데이터로 활용하였습니다. 프로젝트를 인수 받았을 때 데이터 추출률은 89%이었는데, 이후 의도분류에 혼동이 오는 부분을 제대로 분류하고 미추출된 문장들을 대거 수집하여 학습을 진행하여 95%로 추출률을 끌어 올렸습니다. 문제점의 정확한 원인을 데이터를 통해 분석하고 발견하여, 그 해결점을 찾아 학습을 진행하였더니 품질이 좋아지는 결과를 얻을 수 있었습니다. 무작위로 학습데이터를 추가 시키기 보다는 검증 결과 분석을 통해 근거 있는 데이터들을 추가함으로써 유의미한 성과를 이루었습니다.

본인의 직무 강점

저는 데이터 정제와 분석을 통해 문제점을 발견하고 해결책을 찾는데 강점이 있습니다. 앞서 소개 하였던 프로젝트 경험들은 인식을 향상이라는 수치적인 요소와 이 결과를 만들어 내기 위해 분석하고 생성한 자료들이 저에게는 의미 있는 성과로 남아있습니다. 업무를 수행하면서 발생하는 여러 과정의 기록들을 결과물에 치열하게 남기려고 노력하고, 모델의 성능을 수십 번 테스트하여 원하는 성과를 이루기 위해 끊임없이 도전하였습니다. 이러한 업무적인 강점과 더불어 대학원에서 공부하고 연구한 AI 기술에 대한 직관력이 AI 개발에 기여할 수 있는 부분이라고 생각합니다. AI 기술의 직관력은, 결국 AI는 인간을 모방하는 것에서부터 시작한다는 통찰력으로, 이러한 생각의 방향은 대학원에서 교수님들의 수업을 수강한 것과 더불어 논문을 작성하며 깨닫게 되었습니다. 이 분야에 획기적인 성과를 이루어 낸 딥러닝도 인간의 신경망 뉴런을 모방하는 것에서부터 시작하였습니다. 마찬가지로 AI의 기술 또한 인간의 인지 과정을 잘 관찰하고 비슷하게 구현한다면 인간과 비슷한 성능을 보일 수 있다고 생각합니다. 저는 이러한 강점을 통해 당사가 고객에게 더욱 편안하고, 편리한 AI 서비스를 제공 할 수 있도록 기여하겠습니다.

아울러, 저는 프로그래밍 언어로 Python을 공부하였습니다. 우선은 대학원 커리큘럼 상 인공지능 알고리즘을 구현하기 위해 처음부터 Tensorflow, Keras 등의 딥러닝 framework로 접근하기 보다 파이썬만을 이용하여 단계별로 직접 알고리즘을 구현하여 동작 원리를 기초부터 자세히 마스터 하였습니다. 다만, 딥러닝의 CNN, RNN 동작 알고리즘은 상대적으로 이론이 복잡하고 코드가 많기 때문에 Tensorflow로 구현하는 연습을 하였습니다. 또한 학부에서 경제수학과 통계학을 심도 있게 학습하여 인공지능 이해에 필요한 확률 및 통계, 선형대수, 미분 등의 개념을 수월하게 이해할 수 있었습니다. 이러한 인공지능과 머신러닝, 딥러닝의 전반에 대한 이론 및 코드 이해를 기반으로 AI연구개발의 업무를 담당하고 싶습니다.

아울러, 저는 대학원 졸업을 위해 논문 작성을 완료한 경험이 있습니다. 업무 여건 상 회사 서비스 문제점 파악 및 개선방안 기획의 형태로 논문 작성을 완료하였습니다. 이 기간동안 AI 상담센터에 대한 전체적인 구조(STT-NLP-TTS)를 넓은 시야로 파악할 수 있었고, 각 단계별로 타 회사 서비스 및 관련 논문들을 찾아가며 현 기술 수준 및 신 기술에 대한 탐색 활동을 열심히 하였습니다. 이러한 경험을 바탕으로 AI신기술 기반 응용 모델 연구개발 및 AI 신기술 대응에 필요한 인력이 되도록 하겠습니다.

부가적으로 이전 경력인 재무팀에서 쌓은 수치를 기반으로 한 엑셀 보고서 작성 능력과 영문으로 의사소통하고 문서를 읽는데 적합한 영어 실력은 원활한 업무 흐름에 좋은 요소로 작용할 것입니다.