



# 구일문

1996.05.05 (만 26세)

Phone 010-2021-2217 | Email mun90505@naver.com | Address 서울시 양천구 목동동로

## 자기소개서

[구체적 원인 분석과 리서치를 통한 결과 정확도 30% 증가]

60%에 불과했던 초기 딥러닝 모델의 정확도를 94%까지 끌어올린 경험이 있습니다. 미국 퍼듀대학교에서 오디오 데이터를 활용하여 드론의 화물 적재 여부를 판단하는 딥러닝 프로젝트를 진행할 때, 드론 소리가 담긴 원시 데이터를 총 4가지로 분류하였습니다. 보다 구체적으로는 노이즈, 화물 비적재 드론, 화물 1개 적재, 화물 2개 적재 드론으로 분류하였는데 초기 정확도가 낮은 원인은 다음과 같았습니다.

먼저 원시데이터 하나로부터 추출하는 피처의 양이 부족했습니다. 다음으로는 데이터 수집 시 드론과 마이크 사이의 거리가 일정하지 않아 균일한 데이터가 수집되지 않았습니다. 따라서 데이터로부터 더 많은 피처를 추출하는 것이 필요했고, 논문을 리서치하던 중 mfcc의 개수를 40개로 설정했을 때 더욱 좋은 결과가 나타났다는 사실을 파악하여 반영했습니다. 또한 드론과 마이크의 최대거리 제한을 15m로 설정하여 오디오 데이터를 재수집했습니다.

그 결과 분류 정확도 CNN 87%, CRNN 87%, LSTM 72%라는 향상된 결과를 얻을 수 있었습니다. 하지만 Confusion matrix 분석 결과 화물 1개 적재 드론과 2개 적재 드론을 잘 구분하지 못하는 문제는 여전히 존재했는데, 두 화물의 무게 차이가 62g에 불과했기 때문이었습니다. 따라서 데이터를 획기적으로 늘릴 수 있는 방안을 리서치하였고, 한 논문에서 제시한 Data Augmentation 방법을 적용하여 데이터를 800여 개에서 2500여 개로 확장하였습니다.

해당 데이터를 바탕으로 모델을 학습시킨 결과 두 화물 적재 클래스를 구분하는 성능이 대폭 개선되었고 전체 성능도 CNN은 87%에서 94%, CRNN은 87%에서 91%로 증가하였습니다. 해당 프로젝트를 바탕으로 작성한 논문은 현재 IEEE-IRC에 제출 준비 중에 있습니다.

[컴퓨터 공학을 시작한 계기]

학부 재학 시절 기술의 발전으로 세계의 많은 사람이 일자리를 잃을 것이라는 제레미 리프킨의 '노동의 종말'이라는 책을 접한 후 4차산업혁명 시대의 AI 기술에 관심을 가졌습니다. AI기술은 새로운 산업을 개척함으로써 역설적으로 많은 일자리를 만들어 줄 수 있다고 생각했기 때문입니다. AI를 다루는 일을 하고 싶다는 생각 아래, 'AI융합전공'을 신청하여 컴퓨터 전공 공부를

시작했습니다. 컴퓨터 과학은 생소한 분야였기 때문에 처음에는 수업을 잘 따라가지 못했고, 내주는 과제도 겨우 완성하기에 급급했습니다. 하지만, 매주 전공과목 교과서를 세 번씩 복습하고 10시간씩 CS 스터디를 진행한 결과 '자료구조'와 '데이터베이스' 과목에서 각각 A+와 A 성적을 받을 수 있었습니다. AI를 제대로 활용하기 위해서는 이론에 그치지 않고 직접 개발을 해야 한다는 생각에 '멋쟁이 사자처럼'이라는 IT 동아리에 가입하여 프로그래밍을 시작했습니다. 동아리에서 주최하는 해커톤에서 백엔드 파트를 맡아 서버 개발에 도전했지만, 초반에는 장고 ORM시스템과 RESTful한 개발의 의미를 전혀 이해하지 못해 개발에 시간이 많이 걸렸습니다. 끝없는 에러에 포기 생각도 했지만, 인내 끝에 사전 개발을 마칠 수 있었고, 해커톤에서 강의평가 시스템을 완성할 수 있었습니다. 프로그래밍을 처음 접했을 때에는 막막하고 어려웠지만, 포기하지 않고 꾸준히 도전한 결과 성장을 거듭할 수 있었습니다.

#### [AI에 대한 관심과 열정을 확인한 인턴생활]

‘모블디’라는 기업에서 운영하는 자체 웹툰 페이지인 ‘모툰이’의 ‘웹툰 추천 시스템 문제 해결 프로젝트’를 수행했습니다. 해당 웹툰 시스템의 문제점은 웹툰을 평가할 적절한 지표를 가지고 있지 않다는 점이었습니다. 클릭률로 알려진 CTR의 경우 해당 웹툰의 내용에 대한 사용자의 실제 선호도를 알기 어렵다는 단점이 있었고, 별점의 경우 평가하는 사람들이 일정하지 않고, 기준도 저마다 달라 신뢰하기 어렵다는 단점이 있었습니다. 웹툰에 대한 종합적이고도 안정적인 지표가 필요한 상황이었고, 리서치를 통해 찾은 카카오의 논문에서 열독률 지표(Deep Reading Index, DRI)를 구현하였습니다. 열독률 지표(DRI)는 사용자가 게시물을 읽은 정도(depth)와 체류 시간을 중심으로 산정하기 때문에 CTR, 별점과 달리 사용자의 콘텐츠에 대한 실제 선호도를 효과적으로 파악할 수 있습니다. DRI를 평가지표에 반영하여 CTR기반 랭킹 대비 DRI기반 랭킹으로 사용자 그룹에 맞춤 추천한 결과 뷰당 평균 체류시간은 3.2%, 인당 평균 체류시간은 2.4% 증가하였습니다. 해당 프로젝트를 위한 리서치를 진행하면서, 최신 기술들이 담긴 원어 논문을 찾아보고, 그 속에 담긴 내용을 검토하여 실제에 구현함으로써 사람들의 실제 반응을 더 정확하게 전달할 수 있다는 점에서 리서치의 매력을 느꼈습니다.

#### [다양성 있는 데이터셋 구축]

퍼듀대학교에서 진행한 위 프로젝트의 목표는 테러 위험지역에서 폭탄으로 의심되는 화물을 드론이 적재하였는지 여부를 사전에 탐지하는 것입니다. 해당 프로젝트에서는 두 개의 드론만을 활용하여 데이터를 수집하였기 때문에, 해당 목표를 충실하게 수행하기 위해서는 다양한 종류의 드론 음성 데이터를 수집하는 것이 필요했고, 추가적으로 10개 종류의 드론, 총 5215초 분량의 드론 음성 데이터셋을 구축하는 추가 프로젝트를 진행했습니다. 현지 교수님과 대학원생이 이끄는 해당 프로젝트에 팀원으로 참여하여 그 중 4개 종류의 드론, 2125초 분량의 데이터셋 수집을 담당하였습니다. 해당 데이터셋을 CNN을 활용해 10가지로 분류한 결과 83% 정확도를 기록하였습니다. 위와 같은 결과를 바탕으로 또 하나의 논문을 작성하였고, 해당 논문은 IEEE-IRC에 제출되어 Short Paper(4 pages)로 승인되어 발표를 앞두고 있습니다.

---

위의 모든 기재사항은 사실과 다름없음을 확인합니다.

작성자 : 구일문