

언어 및 데이터지능 연구실 인턴 지원서

예연희

저의 관심 분야는 자연어 이해(Natural Language Understanding)와 자연어 이해의 성능을 높이는 것입니다. 덴마크에서 교환학기를 보내던 시기에 네이버에서 개발한 번역기인 '파파고'와 영어를 교정 · 교열하는 플랫폼인 'Grammarly' 모두 딥러닝을 통한 자연어 처리에 기반한다는 것을 알게 되었습니다. 전쟁, 교역, 도서 등을 넘어 기술도 자연어의 변화에 개입할 수 있다는 점이 놀라웠습니다. 파파고를 통해 제 해석보다 더 자연스럽게 영어 텍스트를 이해하고 Grammarly를 통해 제 문장보다 더 올바르게 전치사를 쓸 수 있었습니다. 이 과정에서 제가 지금까지 알고 있던 것과 다른 방법으로 영어 문장을 바라보고 쓰게 되었습니다. 만약 더욱 많은 사람들이 저와 비슷한 경험을 한다면 우리는 조금씩 실수를 줄여 나가며 언어를 사용해 나갈 것입니다. 이처럼 기술에는 비교적 짧은 시간 동안 다양한 배경을 갖고 있는 사람들에게 긍정적인 영향을 끼칩니다.

더 나아가 기술의 또 다른 장점은 아이디어를 실현하는 힘이 있다는 것입니다. 워드 임베딩(word embedding)을 하는 데에 있어 언어학자 존 루퍼트 퍼스(J.R. Firth)의 이론은 핵심적인 아이디어입니다. 퍼스는 '한 단어와 함께 쓰이는 단어들로 그 단어의 의미를 알 수 있다'고 말합니다. 예를 들어 '시간이 많다'와 '두 시간이 걸린다'에서 모두 '시간'이 쓰였지만, '시간'의 의미는 주변 술어나 명사에 따라 '여유'가 되고 '시간의 단위'가 됩니다. 우리는 역으로 '시간'과 같이 쓰인 단어를 분석하여 '시간'의 의미를 파악할 수 있습니다. 분석철학자 비트겐슈타인은 한 단어의 의미는 문맥 내에서 그 단어가 사용되는 방식으로 인해 결정된다고 했고, 저는 비트겐슈타인의 이론과 퍼스의 주장이 상통한다 생각했습니다. 이와 같이 철학에서 던지는 물음이 형이상학적인 논의에 그치는 것을 넘어 현실에서 이루어지는 것을 바랬고, 이는 제 전공을 자연어 처리로 바꾸고 싶은 계기가 되었습니다.

이를 위해 스탠포드(Stanford) 대학의 '딥러닝을 통한 자연어 처리(CS224N: Natural Language Processing with Deep Learning)' 수업을 바탕으로 자연어 처리에 관해 배우고 있습니다. 그리고 이 수업의 과제를 풀어가며 자연어 처리에 필요한 수학과 코드를 익히고, 프로그래밍을 할 수 있는 플랫폼인 Codecademy의 예제를 활용하여 머신러닝의 모델을 코드로 구현하는 연습을 하고 있습니다. 그렇지만 혼자 공부하면서 제가 미처 보지 못하는 부분을 채우고 싶다는 갈망이 커졌습니다. 딥러닝으로 자연어 처리를 활발히 연구하는 곳에서 실제 자연어 처리 연구는 어떻게 이루어지는지 직접 경험하고 싶었고, EMNLP, ACL, NAACL에 논문을 꾸준히 올리는 연구실은 바로 '언어 및 데이터 지능 연구실'이었습니다. 언어 및 데이터지능 연구실의 연구 분야를 찾아 보면서 자연어 처리의 성능을 높이기 위해서는 인터넷에 흩어진 수많은 데이터를 먼저 효과적으로 처리할 수 있어야 한다는 것을 알게 되었습니다. '호랑이를 잡으려면 호랑이 굴에 들어가야 한다'는 말처럼 제 관심 분야인 자연어 이해를 깊이 연구하는 사람이 되기 위해 자연어 처리 연구 현장에서 현 연구자들과 함께 논문을 읽고 연구 동향을 파악하고 싶습니다.

철학을 전공하면서 철학자들의 다양한 논의뿐 아니라 문제를 제기하고 그에 대한 답을 논증을 해 나가는 법을 배웠습니다. 이는 컴퓨터 과학을 배울 때도 주요한 방법이란 것을 알게 되었습니다. 처음 수강한 융합 소프트웨어의 수업에서는 높은 성적을 받지 못했습니다. 제 자신에게서 시작하는 문제가 없었기 때문입니다. 그 다음 학기에 수강한 자바(java)로 프로그래밍을 하는 수업에서는 문제를 해결하기 위해서 어떤 코드를 써야 하고 그 코드는 왜 필요한지 등 제가 궁금한 점에서 시작하고 그 점에 하나씩 답을 했습니다. 그 결과 프로그래밍 실력이 점차 느는 것을 느끼고 A+을 받을 수 있었습니다. 이러한 마음 가짐으로 교수님의 연구실에서 대학원 연구생들의 지도를 받는다면 부족한 부분을 채우고, 더 나아가 자연어 처리에서 핵심적인 문제를 제기하는 능력 있는 연구자로 거듭날 거라 확신합니다.