프로그램 합성 분야를 중점적으로 봤던 SIGPL 겨울학교

이도윤

February 5, 2023

프로그램 합성 관련 과제의 마감 기한이 다가와 그것에 몰두하고 있었고, 앞으로의 연구도 그 쪽으로 생각하고 있었기 때문에 프로그램 합성에 관한 내용에 집중해서 듣는 자세를 선택했다. 이번 겨울학교에서는 초청강연의 시간 비중이 여느 때보다도 상당히 높았다. 그러나 초청강연들이 애초에 PL분야라고 보기 힘든 강연들이 많았기 때문에, 내 현 상황에 도움이 되는 지식들만 뽑아서 이해하는 것이 나을 것으로 생각했다. 번개 발표는 주제만 보고 어떤 포스터를 보러 갈지 참고하는 정도로만 들었으며, 포스터 세션에서 유익한 대화를 나눌 수 있었지만 시간이 짧아 아쉬웠다.

그나마 프로그램 합성과 연관이 있었던 초청강연은 초청강연1로 "딥 전이학습 기반 프로그램 자동 수정: 산업 적용 사례를 중심으로"라는 주제의 강연이었다. 수행 중인 과제도 오류 자동 수정이라는 테마에서 시작한 연구이기 때문에 흥미를 가지고 들었지만, 그 접근 방식이 LLM을 사용하는 것이었기 때문에 방식에 대해서는 별 도움을 얻을 수 없었다. 문법이 아닌 생성된 설명 기반으로 코드를 분류하고 접근하기 때문에 이론적인 뿌리가 다르다고 느꼈다. 그렇지만 이 LLM이 프로그램 합성과 비슷한 일을 해내고 있고, 좋은 성능을 내고 있기 때문에 제대로 알고 배울 점을 찾아낸다 생각하며 들었다. 첫 번째로, 설명 기반으로 코드를 분류하는 것이 왜 도입되었는지를 생각했을 때, 그것이 꽤 일리가 있다고 생각했다. 프로그래머가 무엇을 하고자 했는가를 파악하면 코드 자동 생성 문제에 방향이 잡히기 때문이다. 프로그램 합성 문제는 원숭이 백만 마리에게 시킨다는 표현처럼 가능한 모든 프로그램을 만들어가는 방향으로 풀어나가서 이 자체로는 딱히 문제가 잘 풀릴 이유가 없다. 그래서 싹수 분석 같은 가지치기 기술을 통해서 문제가 잘 풀릴 이유를 따로 준비해줘야 한다. 내가 다룬 상황들은 입출력기반 프로그램 합성이 대부분이었는데, 사용자가 무슨 의도였는지 밝혀지지 않고 그저 입출력 데

이터에 프로그램을 짜맞추게 된다. 사용자의 의도를 미리 파악할 수 있으면, 어떤 프로그램 조각을 사용해야 할지, 어떤 식으로 이들을 합성할지에 대한 가이드라인을 얻을 수 있을 것이다. 이런 방식은 프로그램의 최소성을 보장하지는 못하여 한정된 재료 크기 내에서의 안전성을 잃게 되지만, 프로그램 합성 속도는 더 향상시킬 수 있다. 두 번째로, edit script를 사용하여 코드를 수정했는데 이런 식으로 어떠한 종류이든 템플릿을 사용하는 것이 참고할 만했다. 내가 겪은 프로그램 합성 문제에서는 항상 기본 문법에서 시작하여 기본 구조만을 이용해 합성해 나갔는데, 이러면 정답 프로그램에 상당히 가까워지더라도 바로 정답 프로그램으로 점프하는 것은 불가능하고, 재료 크기에 따라 지수적으로 증가하는 그 많은 스텝들을 하나하나 다 거쳐야만 한다. 특정 템플릿을 재료로 사용하거나, 혹은 미리 탐색할 수 있는 지름길을 제공하는 것이 합성 성능 향상에 도움이 될 것으로 여겨졌다.

그 외의 초청 강연에서는 마지막 날에 있었던 "코드 속성 그래프의 부분 그래프 매칭을 통한 대규모 취약점 탐지"라는 주제의 강연에서 배울 점을 여럿 찾을 수 있었다. 첫 번째로 CFG에 대해서 배울 수 있었다. 프로그램 합성 문제에 따라서 주변 맥락이나, 수정해야 할 코드 등 어떠한 형태로든 코드가 주어지는 경우가 있다. 이들을 이해하여 표현하는데 그래프 요소를 활용하는 방식도 익혀야 하겠다는 생각이 들었다. 내가 이를 잘 익힌다면 요약 해석에서의 요약 도메인으로 이러한 그래프 요소를 쓸 수 있게 될 것이다. 두 번째로 github에서 코드를 긁어와서 벤치마크로 사용했는데, 사실은 제대로 된 코드들 자체가 별로 없어 많은 비율이 버려졌고, 거기서 더욱 일부만을 실험 대상으로 해야 했다는 경험을 얘기해주신 것을 나중에 꼭 염두에 두어야 할 것이라 생각했다. 하려는 연구에 대해서 벤치마크가 제대로 된 것이 없는 상태라면, 그 벤치마크를 구성하는 것도 만만치 않은 일이라는 것을 최근 들어 느끼고 있다. 그래서 벤치마크 구축 작업도 연구의 영역에 넣는 것만 것 옳은 것은 아니고, 그냥 그 연구는 아직 하기 이른 것이라는 판단을 내릴 때도 필요할 것이라는 생각이 들었다. 세 번째는 듣다 보니 든 생각인데, 그래프 요소에 대한 싹수 분석도 연구 주제로 괜찮을 거라는 생각이 들었다. 그래프 요소는 프로그램의 기본 자료형은 아니지만, 대부분의 알고리즘에서 사실 기초적으로 사용되고 있는 것이 이러한 노드와 간선으로 이루어진 그래프 요소들이다. 프로그램 합성이라는 분야가 프로그래머가 정말 필요로 하는 프로그램을 만들어내게 하기 위해서는, 프로그래머들이 알고리즘을 배우며 기초적으로 배웠을 그래프 요소관련 연산들을 문법으로 정리하여 이를 분석하는 것이 꽤 도움이 될 것이라는 생각이 들었다.

포스터 발표가 특이하게도 저녁 이전에 있었는데, 시간은 더 짧아서 보통은 관심 있든 없든 모든 포스터를 살펴보기는 했는데, 이번에는 그러지 못했

다. 내가 관심 있는 주제들만 찾아가서 얘기를 나누었다. 그 주제들은 포인터 분석, rust 매개변수의 타입, 퍼징 순서, LLM 솔루션 참고 프로그램 합성, 유사한 버그의 패치, 양자 프로그램 합성과 같은 주제들이었다. 각 주제에 대해서 여기에 정리하지는 않겠지만, 프로그램 합성 분야에 대해서 상당히 심각한 말을 들었다. 프로그램 합성 분야에 나온 논문들이 재현하려고 하면 불가능한 것들 것 상당히 많다는 것이고, 특히 중국 쪽에서 나온 논문들이 그러하다는 것이다. 나는 논문을 읽을 때 저자의 국적까지 신경 쓰며 읽지는 않았기 때문에 그 지식들이 분리되지 않은 상태로 머릿속에 들어있는 상황이다. 다시 논문들의 저자를 확인해야 하는 수고를 해야 하게 되어버렸다. 덕분에 프로그램 합성 분야를 계속하는 것이 맞을지 고민하는 시각을 갖게 되었다. 좁게 보면 이미 발을 꽤 담근 거지만, 계속 연구하는 인생을 살 거라고 넓게 보면 아직 늦지 않았을지도 모르기에 천천히 고민해 보려고 한다.

학문적인 것은 프로그램 합성을 하는 사람의 시야에서 본 것을 위에서 설명했고, 그 외에는 역시 대전의 명물 성심당이 기억에 남는다. 항상 겨울학교 같은 워크샵에 갔을 때 먹는 것을 풍족하게 먹고 있다. 첫 날 저녁을 운영진이 정해 준 조별로 먹게 되어 있었는데, 이번 겨울학교에 추가되었던 활동인 조별활동에 대해서는 의도는 좋았으나 유익하지는 않았다고 생각된다. 차라리 포스터 시간이 많은 것이 관심 분야의 사람들과 더 깊은 얘기를 나눌 수 있어 서로에게 유익할 것이다. 어쩌면 내가 너무 시야를 좁게 가지고 참여를 해서 이런 판단을 한 것일 수 있으나, 인턴 때와는 달리 대학원생이 되고 나서는 자기 분야 챙기기도 바빠지는 것이 다들 어쩔 수 없을 거라고 생각한다. 다음 여름학교에서는 과연 어떤 시야를 가지고 참여하는 상황일지 기대된다.