



FSE 2022 in Singapore

이석현

2022.11.13. - 11.19.

초록

FSE'22는 올해 두번째 학회 참여 기회였다. 완화된 COVID-19 상황 속에서 학회에서 다양하고 재미있는 새로운 연구들을 접할 수 있었고, 편하게 다양한 싱가포르의 모습을 느낄 수 있었다. 이 글을 통해 이번 FSE'22와 싱가포르에서 보고 얻게된 점을 짧게 공유하고자 한다.

1 개요

FSE 2022에 참여하기 위해서 2022년 11월 13일부터 29일까지 6박 7일간 싱가포르에 다녀왔다. 정말 많은 연구 발표가 진행된 만큼 다양하고 재미있는 연구들을 접해 볼 수 있었다. 한국에서는 한파가 기승을 부리기 시작하던 시기에 날씨가 더운 싱가포르에서 싱가포르의 다양한 모습도 함께 볼 수 있다.

2 FSE 2022

ESEC/FSE(The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering)는 소프트웨어 공학 분야의 저명한 학회로, 매년 많은 연구자들이 본인의 연구를 발표하고 새로운 아이디어를 얻기 위해 이 학회를 찾는다. 2019년 말부터 이어져오던 COVID-19의 상황이 완화되면서 3년만에 다시 오프라인 학회가 진행되었고, 지난 2년간 FSE에서 발표됐던 논문들 모두 오프라인에서 다시 발표할 수 있는 기회가 있어 학회에서 더 많은 연구를 접할 수 있었지만, 병렬로 진행되는 발표의 수도 많아 흥미가 생기는 연구들 중 일부를 선택해야 하는 경우도 있었다. 그럼에도 불구하고 학회에서 소개되는 논문이 많아서인지, 학회에 있는 시간 중 많은 시간을 흥미로운 연구를 들으며 보낼 수 있었다. 또한, 발전하는 기계학습 기술에 따라, 정말 많은 연구가 기계학습 기술에 의존하는 것도 흥미로웠다. FSE'22에 소개된 여러 연구들 중 가장 인상 깊었던 연구 몇 가지와 기초 연설 하나를 짧게 소개하고자 한다.

1. NatGen: Generative Pre-training by “Naturalizing” Source Code [2]

최근 몇년간 인공신경망을 이용한 자연어처리 기술이 비약적으로 발전하면서, 많은 분야에서 인공신경망 기술을 이용하여 기존에 이뤄내지 못한 성능향상을 이뤄내고 있다. 그 중 한가지 분야가 프로그램 합성 분야인데, 프로그램을 작성하는 것을 자연어를 사용하는 것과 비슷하게 해석하여 기계학습 모델이 프로그램을 작성할 수 있도록 한다.

이 연구는 수집한 데이터에 변형을 가해 학습 성능을 높이는 연구이다. 인공 지능 모델이 특정 구문(혹은 표현식)의 정확한 의미를 파악하는데 도움을 주기위해 수집된 데이터에 인위적으로 같은 의미의 다른 구문(혹은 표현식)으로 대체하는 변형을 가한 후 이를 복원하는 학습을 진행하여 기계학습 모델이 그 의미를 더 정확하게 파악 할 수 있도록 했다. 이 연구를 들으면서 기존 기계학습 분야의 성공적이었던 다른 연구들(오토인코더, 확산 모델)이 생각나 더욱 흥미로웠다.

2. All You Need Is Logs: Improving Code Completion by Learning from Anonymous IDE Usage Logs [1]

Jetbrains는 세계적으로 많은 곳에서 사용되는 IDE(통합 개발 환경)들을 개발하는 회사이다. IDE의 여러가지 기능들 중 코드 조각 추천 기술은 IDE를 사용하는 가장 중요한 기술 중 하나로 많은 개발자들이 실제로 코드 조각 추천 기술을 활용하여 개발의 속도를 더한다. 이 연구는 Jetbrains에서 코드 조각 추천 기술의 향상을 위해서 사용한 다양한 기법들(논문의 제목에도 나와 있는 익명의 로그를 활용하는 방법을 포함하여)이 나와 있는데, 한번도 생각해보지 않았던 협업에서 활용될 도구를 개발하기 위해 고민해야 할 다양한 정보를 얻을 수 있었다. 다만, 산업 트랙에 발표되는 연구여서 그런 것인지는 모르겠지만, 많은 예시를 보여주지 않아 아쉬웠다.

3. TypeWriter: Neural Type Prediction with Search-based Validation [5]

동적 타입을 지원하는 언어(Python, Javascript 등)을 분석하는 것은 정적 타입을 지원하는 언어를 사용하는 것보다 어렵다. 동적 타입을 지원하는 언어를 분석하기 어렵게 만드는 가장 큰 이유 중 하나는 변수의 타입을 정적으로 유추하기가 어렵기 때문인데, 이는 정적으로는 특정 변수의 값은 물론이거니와 그 타입 또한 알기 어려워 기존의 분석 기술들이 해당 변수를 꼭대기로 분석해야하기 때문이다.

이 연구는 기존에 정적으로는 알기 어려운 정보를 인공신경망 기술을 활용하여 얻고자 하는 연구이다. 정적인 타입 정보를 계산하기 위한 식을 세우고, 정적으로 알기 어려운 정보는 인공신경망 모델에 질의하여 가장 그럴듯한 타입을 활용하여 타입을 계산해 보는 것이다. 타입을 계산하는 데 성공하면 타입을 반환하고, 실패하면 다시 인공신경망 모델에 질의하여 다른 타입을 받아 다시 시도해 보는 것을 반복하는 것이 이 연구의 아이디어이다.

당시, 동적 타입 시스템을 갖는 언어(특히, Python)에서 타입을 추론하는 연구를 준비하고 있었기 때문에 이 연구의 저자가 자신의 연구를 소개하는 것과 동적 타입 시스템 언어에서 타입을 추론하는 것에 대한 생각을 직접 들을 수 있어서 좋았다.

4. AI-Assisted Programming: Applications, User Experiences, and Neuro-Symbolic Techniques [4]

최근 프로그램 합성 기술이 많은 주목을 받고 있는 것 같다. 올해 참석한 두 학회 모두에서 프로그램 합성 기술들(특히, 기계학습 분야의 기술을 활용하는)에 관한 이야기를 학회 전체의 기조 연설로 채택했다.

이 기조 연설에서는 기존의 Flash-Fill[3]을 이용의 다양한 활용 방법과 한계, 그리고 최신의 인공신경망 기술을 이용한 더 나은 Flash-Fill을 소개했다. 이번 기조 연설에서 소개한 다양한 활용 예시 중, 학생들의 컴파일 오류를 수정해주는 것도 있었는데, 현재 진행 중인 연구와 그 동기가 비슷하여 같이 간 연구실 동료들과 함께 웃기도 하였다.



그림 1: 왼쪽: 싱가포르의 유명한 조형물인 머라이언 동상, 오른쪽: 싱가포르의 밤에 찍은 아트 사이언스 뮤지엄과 야경

3 싱가포르

FSE'22는 싱가포르의 싱가포르 국립대학에서 열렸다. 북위 1.290° 에 위치한 싱가포르는 한국에서는 겨울이 시작되는 11월 중순임에도 여전히 날씨가 더웠다. 바닷가에 위치한 도시 국가인 만큼 해안선을 따라 건물이 높게 서있었다. 해안선을 따라 펼쳐지는 건물들의 풍경은 흔히 생각하는 해안 도시 그 자체였다. 특히, 해가 저물고서 볼 수 있는 해안선을 따라 펼쳐진 싱가포르의 야경은 밤의 싱가포르가 왜 유명한지를 알려주었다. 또한, 싱가포르는 아시아의 국제 무역의 중심 도시 중 하나로 여러 다양한 문화(중국, 인도, 유럽, 아랍 등)를 다양하게 접할 수 있어 좋았다(다양한 문화를 한번에 체험해 볼 수 있었지만, 싱가포르만의 특색이 강하게 느껴지지는 않아서 아쉽기도 했다).

COVID-19 싱가포르와 한국 모두에서 COVID-19 관련한 제한이 거의 해제되면서 이제는 편하게 다녀 올 수 있었다. 양국에 입국할 때 간단한 문진표만 작성하면 돼, 지난 1월 미국에서 열린 POPL'22에 다녀 올 때 수 많은 고생을 했던 것과 비교하면 제한이 없는 것 같았다. 싱가포르 내부에서도 대중교통을 이용할 때와 병원에 방문 할 때를 제외하고는 마스크 착용도 의무가 아니라 더욱 편하게 다닐 수 있었다.



그림 2: 싱가포르 플라이어의 최정상에서 찍은 싱가포르의 풍경, 원쪽에서 가든즈 바이 더 베이, 마리나 센스 베이를 포함하여 다른 높은 건물들이 함께 보인다.

4 마치며

많은 연구실 동료들과 함께 FSE'22에 다녀올 수 있어서 싱가포르에 있는 모든 순간이 재밌었다. 특히, 학회에서 다양한 연구 발표를 들으면서 새로운 지식을 얻어 갈 수 있어서 즐거웠다. 마지막으로, 좋은 기회를 지원해주신 오학주 교수님께 감사드린다.

참고문헌

- [1] BIBAEV, V., KALINA, A., LOMSHAKOV, V., GOLUBEV, Y., BEZZUBOV, A., POVAROV, N., AND BRYKSIN, T. All you need is logs: Improving code completion by learning from anonymous ide usage logs. In *The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (2022), FSE'22.
- [2] CHAKRABORTY, S., AHMED, T., DING, Y., DEVANBU, P., AND RAY, B. Natgen: Generative pre-training by “naturalizing” source code. In *The ACM Joint European*

Software Engineering Conference and Symposium on the Foundations of Software Engineering (2022), FSE'22.

- [3] GULWANI, S. Automating string processing in spreadsheets using input-output examples. In *Principles of Programming Languages* (2011), POPL'11.
- [4] GULWANI, S. Ai-assisted programming: Applications, user experiences, and neuro-symbolic techniques (keynote). The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE), 2022.
- [5] PRADEL, M., GOUSIOS, G., LIU, J., AND CHANDRA, S. Typewriter: Neural type prediction with search-based validation. In *The ACM Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering* (2020), FSE'20.