SIGPL 2024 겨울학교 Trip Report

KAIST N1 117, 2023.1.31 - 2023.2.2

오규혁

ghoh@ropas.snu.ac.kr

SIGPL 2024 겨울학교

SIGPL 2024 겨울 계절학교는 대전 KAIST에서 열렸다. 이번 계절학교의 주제는 "사이좋게 지내자"로 자주 못 오셨던 분들의 발표를 듣고, 학생끼리 조를 만들어 다양한 학생들이 조별활동을 통해 친해지고자 하는 취지였다.

1월 31일 (수요일)		
12:00 ~ 13:00	등록	
13:00 ~ 13:10	**개회사**	강지훈 (KAIST)
13:10 ~ 13:20	전체 프로그램 설명	이성호 (충남대)
13:20 ~ 14:00	(초청강연 1) 딥 전이학습 기반 프로그램 자동 수정: 산업 적용 사례를 중심으로	김미수 (전남대)
14:00 ~ 14:30	휴식	
14:30 ~ 15:30	**번개 발표 1**	대학원생
15:30 ~ 15:40	휴식	
15:40 ~ 16:30	**번개 발표 2**	대학원생
16:30 ~ 18:00	**포스터 발표**	대학원생
18:00 ~	저녁 및 조별 활동	
2월 1일 (목요일)		
10:30 ~ 11:10	(초청강연 2) 임베디드 시스템을 위한 컴파일러 디자인	허선영 (경희대)
11:10 ~ 11:30	휴식	
11:30 ~ 12:10	(초청강연 3) 근사 컴퓨팅의 자동 적용을 위한 컴파일러 최적화 기법	박영준 (연세대)
12:10 ~ 12:20	사진촬영	대학원생
12:20 ~ 13:40	점심식사	**교수회관**
13:40 ~ 14:20	(초청강연 4) LeakPair: 단일페이지 웹 어플리케이션내의 메모리 누수를 고치기 위한 선제적 디버깅	김동선 (경북대)
14:20 ~ 15:00	(초청강연 5) 다중 노드 환경에서 빅데이터 스트리밍 엔진의 성능 평가 방법	양신형 (연세대)
15:00 ~ 15:20	휴식	
15:20 ~ 16:00	(초청강연 6) 소프트 오류로부터의 신뢰성 향상을 위한 비트 단위 정적 분석 기법	고유선 (연세대)
16:00 ~ 18:00	조별활동	
18:00 ~	뱅킷	**교수회관**
2월 2일 (금요일)		
10:30 ~ 11:10	(초청강연 7) 코드 속성 그래프의 부분 그래프 매칭을 통한 대규모 취약점 탐지	위성일 (UNIST)
11:10 ~ 11:40	조별활동 콘테스트 및 시상 / 포스터 시상	박지혁 (고려대)
11:40 ~ 11:50	폐회사	강지훈 (KAIST)

Sigpl Schedule

기존과 비슷하게 대학원생들의 연구들을 볼 수 있는 번개 발표와 포스터 발표, 다양한 연구자 분들의 연구에 관한 강연에 이번에 새롭게 추가된 학생들간 교류를 하는 조별활동으로 구성되어 있었다.

번개 및 포스터 발표

첫날 진행된 번개 및 포스터 발표에서 다룬 연구 중 'LLM이 생성하는 코드의 문법적 올바름 보장하기', '언어 모델을 이용한 엄밀 증명 합성', '파이썬 API Migration의 자동화', 'Synstiller: 3분 완성! 언어 모델 추출, 합성, 이식' 등 거대 언어 모델을 활용한 연구들이 눈에 띄었다. 아직은 부족한 점이 존재하지만 최근 빠르게 발전하고 주목받았던 언어 모델이 만드는 결과물을 합성에 이용하거나 반대로 내놓은 결과물을 수정하는 방식이 주로였다.

이러한 연구 외에 개인적으로 재밌게 보았던 연구 주제는 양자 프로그램 합성에 관한 연구, C-Rust 번역 시 출력 매개변수를 대수적 타입으로 대체하기, TEE API 정형 명세와 응용 프로그램 검증 등이 있었다.

포스터 발표 중에는 여러 종류의 맥주가 제공되어, 참석자들은 이를 마시면서 발표를 듣는 점이 특이했다. 술을 곁들여서 그런지 사람들의 텐션도 좀 더 좋아 발표에서 좀 더 활기가 돌았던 것 같고 가끔식 얼굴 톤도 술톤으로 변한 사람들을 보는 재미도 있었던 것 같다.

강연

3일 동안 진행된 강연은 다양한 연구를 접해볼 수 있어 신선했다. 각 강연마다의 내용을 간략하게 정리해보았다.

김미수, 딥 전이학습 기반 프로그램 자동 수정: 산업 적용 사례를 중심으로

김미수 교수님의 연구는 초기에는 "정보검색 기반 버그 추적 자동화"에 중점에 두었다. 버그 리포트를 받아 해당 내용을 검색하여 결함 위치를 특정하는 "정보검색 기반" 방법을 사용하였고, 이를 자동으로 추적하는 "추적 자동화" 기술을 개발하여 이를 통해 다수의 버그 리포트를 분석하고 결함 위치를 추출하는 연구를 수행했다고 했다.

최근에는 "자동 프로그램 수정" 분야로 연구 주제를 변경하였습니다. 특히, 패치 생성 기법을 통한 자동 프로그램 수정에 집중하고 있습니다. 이 방법은 코드 패치 틀을 만들고, 해당 틀에 맞춘 수정을 자동으로 수행하는 것으로 최근에는 딥러닝과 딥전이학습을 활용하여 코드 패치 틀을 생성하는 방법에 초점을 둔 것 같다.

딥러닝을 활용한 방법에서는 이미 있는 패치에 대한 훈련을 통해 패치 생성 모델을 만드는 과정에서 "잡음 토큰"을 처리하는 것이 성능 향상에 핵심적인 부분이라고 강조하셨다. 또한, 딥전이학습을 이용하는 방법에서는 이미 존재하는 딥러닝 파이프라인에 패치 생성 능력을 장착하는 방식을 사용하며, 이는 통합이 용이하다는 장점이 있다.

마지막으로, 현재 강연자의 연구는 실제 버그를 다수 분석하고, 개발자들에게 효과적인 결과를 제공할 수 있는 기술을 개발하는 방향으로 진행되고 있는 중이라고 하셨다.

허선영, 임베디드 시스템을 위한 컴파일러 디자인

임베디드 시스템에서 머신러닝 모델을 실행할 때 컴파일러 디자인에 대한 연구에 대한 강연을 해주셨다.

임베디드 시스템이 가지는 하드웨어 한계로 동적 메모리 최적화, 양자화, 불필요한 코드 제거 등을 통해 최대한 최적화를 해주지만 이러한 최적화 과정에서는 라이브러리에 의존적이고 사람이 일일이 해줘야 하는 경우가 많아 개발에 있어서 난이도가 높은 상황을 해결하고자 하는 연구였다.

이를 위해 모델을 높은 수준에서 최적화를 하는 컴파일러를 개발하여 라이브러리의 의존도를 낮추고 동적 메모리 계획이 아닌 정적 분석으로 통해 더 좋은 성능을 얻을 수 있었다고 했다.

강연 중 언급한 임베디드 시스템의 하드웨어 제약이 너무나도 빡빡하여 이러한 하드웨어에서 머신러닝을 돌리기 위한 최적화라는 것이 놀라웠다.

박영준, 근사 컴퓨팅의 자동 적용을 위한 컴파일러 최적화 기법

GPU를 사용하는 근사 컴퓨팅에서의 연구와 관련된 발표였다. GPU의 구조적 특성을 잘 활용하기 위한 코드가 필요하지만 프로그램 실행 특성이 동적으로 변하기 때문에 동적 최적화가 필요하다고 한다.

이를 위한 컴파일러 최적화 기법은 패턴 기반으로 여러 패턴 중 적합한 패턴을 골라 적용하는 방식이지만 같은 의미지만 다른 표현으로 패턴을 적용하기 힘들고 적합한 패턴을 고르는데에도 어려움이 있을 수 있는 문제가 있다고 했다.

이를 위해 어셈블리 수준에서 CFG를 가지고 최적화가 가능한지 파악하는 방식으로 의미 기반의 최적화를 진행할 수 있는 점이 유리하며 최적화가 입력에 영향을 받기 때문에 이를 고려하여 영향을 적게 받는 최적화부터 적용하는 방식으로 성능을 높였다고 했다.

입력을 고려한 최적화에서의 어려움과 이를 해결하는 방식이 인상깊었던 것 같다.

김동선, LeakPair: 단일페이지 웹 어플리케이션내의 메모리 누수를 고치기 위한 선제적 디버깅

싱글 페이지 웹 애플리케이션은 필요한 내용을 그때마다 받아오며 렌더링하는 방식으로 여러 페이지로 구성된 경우보다 효율적이지만 메모리 누수가 생길 수 있다는 문제가 있는 것을 해결하기 위한 연구이다. 이를 위해 개발자들의 실수를 패턴 기반으로 감지하여 패치를 제안하는 방식이다.

메모리 누수는 찾기 어려운 문제로, 개발자들은 즉시 해결이 필요하지 않은 경우에는 수정을 꺼린다는 것을 고려하여 연구자는 메모리 버그가 기능적 버그가 아니라는 점을 강조하며, 개발자들에게 패치가 비즈니스적으로 중요하다는 것을 설득하는 것의 중요성을 설명했다. 실제 연구가 활용되기 위한 고민과 방안이 인상적이었다.

양신형, 다중 노드 환경에서 빅데이터 스트리밍 엔진의 성능 평가 방법

빅데이터 스트리밍 서비스의 성능을 측정하기 위한 연구로 스트리밍 서비스의 특성상 여러 서버의 시간을 동기화하는 방법을 제안했다. NTP와 같은 프로토콜은 동기화의 정확도가 떨어져 신뢰하기 힘들며 금융이나 항공 우주와 같은 정밀도를 요하는 서비스를 위한 PTP의 경우에는 GPS, 오실레이터 등이 필요해 너무 비용이 크다는 문제가 있다고 했다.

컴퓨터 하나를 선택하여 ms가 아닌 오실레이터의 진동수를 이용하여 타입 유닛을 설정하여 시간을 측정하는데 생기는 문제와 이를 해결하는 방법들을 소개했으며 생각해보지 못한 주제로 배경 지식도 부족하여 구체적인 방법들을 이해하기 쉽지 않았지만 성능 측정 후 성능 개선을 하는 것이 연구의 목적은 성능을 측정하고 개선하는 것이었지만, 측정 단계에서 발생한 문제로 인해 예상치 못한 어려움이 생겨 이를 해결하는 연구를 하게 되었다는 것이 인상깊었다. 연구 과정에서 생기는 예상치 못한 문제와 이를 해결해가는 과정에서의 어려움을 엿볼 수 있었던 것 같다.

고유선, 소프트 오류로부터의 신뢰성 향상을 위한 비트 단위 정적 분석 기법

우주 방사선 등의 다양한 이유로 생기는 하드웨어 오류가 있으며 이로 인해 하드웨어가 신뢰할 수 없고 이를 해결하기 위한 방법으로 회로 단위에서 보완해주는 방법과 컴파일러가 바이너리를 생성할때 보완해주는 방법에서 소프트웨어 단계에서 해주는 방법을 연구한 내용을 강연하셨다.

세가지 무결성(Control flow, Memory flow, Memory Data)을 위해 카운터, 레지스터, 명령어 캐시, 메모리 캐시가 어떤식으로 영향을 주는지 잘 분류를 하고 비트 레벨 단위를 하향식으로 각 비트의 값을, 상향식으로 각 비트가 프로그램에서 어떤 영향을 줄지 분석하는 방법이었다.

실험에서 소프트웨어에서 보완해주는 방법으로는 오버헤드로 인해 늘어난 프로그램 길이를 고려해주지 않은 기존 대부분 연구의 실험 방법과 다르게 이를 고려해주는 실험 방법을 통해 기존 연구에서 오히려 결과가 나빠지는 경우도 찾을 수 있었다고 한다. 평가 방법을 잘 고려하는 것이 연구를 위해 중요하다는 것을 느낄 수 있었으며 슈퍼마리오와 같이 재밌는 일화로만 알고 있었던 하드웨어 오류를 고려한 컴파일러라는게 색다르고 신기한 강연이었던 것 같다.

위성일, 코드 속성 그래프의 부분 그래프 매칭을 통한 대규모 취약점 탐지

웹서버에서 사용자에게 입력을 받을 때 생길 수 있는 보안 취약점을 위한 연구로 Abstract Syntax Tree, Control flow Graph, Program Dependency Graph을 가지고 있는 Code Property Graph의 부분그래프가 이미 알려진 취약점과 유사한지 찾는 방식이지만 비용이 큰 작업으로 큰 프로그램에 적용하기 어려움이 있다고 했다. 이를 위해 가지치기로 최적화를 하였고 그래프를 잘 추상화하여 비슷한 프로그램인데 다르게 표현 될 수 있는 것들을 요약을 잘 하는 것이 다른 어려운 점이었다고 했다. AST, CFG, DPG를 하나로 표현한 Code Property Graph라는 것이 신기했던 것 같다.

조별활동

조별활동은 첫번째 날과 두번째 날에 진행되었으며 첫번째 날은 조끼리 저녁식사, 두번째 날은 골든벨로 구성되어 있었다.

조는 다양한 학교와 연구실로 구성되어 있으며 구성원의 연차도 고르게 구성하여(박사 졸업부터 졸업 직전의 박사분, 1년차, 석사과정, 학부생 인턴 등) 다양한 사람과 “친하게 지내자”를 할 수 있었던 것 같다. 저녁으로는 대부분 사람한테 호불호가 없는 치킨으로 다양한 연구실에서의 일화와 고민거리들을 나누며 긴 저녁식사를 보내는 재밌고도 유익한 시간이었다.

골든벨은 여러 구성으로 첫번째는 번개 및 포스터 발표에서 발표자료를 일부를 매우 크게 확대한 것을 맞추는 것이었다. CS관련 퀴즈에서는 Lisp 개발자나 Java에서 람다식이 추가된 버전이 몇이라든가 하는 퀴즈들이 있었으며 관련 키워드를 Dall-E를 이용하여 생성한 이미지를 보고 키워드를 맞추는 단어로 Dall-E의 능력이 의심스러움과 동시에 사람들의 센스들이 좋다는 생각을 했다. 퀴즈를 많이 맞추지 못한게 약간은 아쉬웠던 것 같다.

성심당

3일동안 성심당에서 케이터링을 해주었는데 참 맛있었고 먹기도 편리해서 좋은 기억에 남았다. 학회가 끝나고 여유시간이 있어 성심당 본점에 들렀다. 사람이 굉장히 많아 고르는게 쉽지 않았으며 케익만 따로 파는 케익부띠끄가 본점바로 옆에 있는 것이 상당히 인상적이었다. 이렇게 많은 유동인구가 있다는 것이 놀라웠고 대전하면 항상 나오는 성심당의 위엄을 느낄 수 있었다. 빵을 먹어본 가족들의 평도 굉장히 좋아서 누군가가 대전에 대해 물으면 나도 첫번째로 성심당을 말할 것 같다.

마치며

포스터 및 번개 발표에서 다양한 학원생들의 연구에서 항상 동기부여를 받을 수 있는 것 같다. 나도 좋은 연구로 발표의 한 부분을 했으면 하는 마음이 들었던 것 같다.

이번에는 자주 못 왔던 분들의 강연이 상당히 색다르게 다가왔던 것 같다. 평소 접하지 못한 내용들이었지만 학회에 맞게 중요한 통찰을 제시해준 강연들이서 좋은 영감들을 얻을 수 있는 귀중했던 시간이었던 것 같다.

마지막으로 “사이좋게 지내요”라는 모토와 조별활동이 처음에는 걱정이 조금 들었지만 재미있는 조별활동 구성과 다양한 사람들과의 저녁식사는 재미있었던 것 같다. 마지막으로 활발한 연구자들과 교류하며 서로의 연구를 발전시킬 수 있는 학회를 참석할 수 있게 도와준 교수님께 감사하다는 말을 하고 싶다.