

TRIP REPORT

ASE 2024

24.10.27 ~ 24.11.04 SACRAMENTO, CALIFORNIA, US

흥미로운 발표들.....	2
Paper 1. MR-Adopt: Automatic Deduction of Input Transformation Function for Metamorphic Testing.....	2
Paper 2. Efficient Incremental Code Coverage Analysis for Regression Test Suites	3
Keynote 1. Analysis of Perception Neural Networks via Vision-Language Models	4
나의 발표와 ASE.....	5
만반의 준비와 정신없던 발표.....	5
기대했던 마음과 어수선한 학회장.....	6
아무 것도 없는 도시, 세크라멘토.....	7

흥미로운 발표들

PAPER 1. MR-ADOPT: AUTOMATIC DEDUCTION OF INPUT TRANSFORMATION FUNCTION FOR METAMORPHIC TESTING

연구의 좋은 아이디어는 두 가지 방향이 있다고 생각한다. 첫째로 누구도 풀지 못한 문제를 기동찬 발견으로 푸는 것, 그리고 다른 한 가지는 듣자마자 '나도 할 수 있겠는데?'라는 생각이 드는 것. 이 연구는 후자의 방향을 잘 보여주는 연구였기에 공유하려 한다.

본 연구는 Metamorphic Testing의 효과를 더 극대화 시키는 것이 목표이다. Metamorphic Testing은 Property 기반의 테스트 방법으로 Input Relation이 주어지면 예상되는 Output Relation을 만족하는지 검증하는 기법이다. 예를 들어, sin 함수를 테스트 한다고 생각해보자. Input Relation으로 x 와 $\pi-x$ 가 주어진다면, 예상되는 Output Relation은 $\sin(x) == \sin(\pi-x)$ 이다. 입력을 넣었을 때 해당 출력이 나온다면 적절하게 검증된 것이다.

그런데 실제로 개발자들이 작성하는 Metamorphic Test 메소드를 보면 Output Relation은 명확히 드러나 있지만, Input Relation이 하드코딩 되어있는 경우가 70%로 많았다는 것이다. sin 함수로 예를 계속해서 들자면 Input Relation은 $x1 = 1.234$, $x2 = \pi - 1.234$ 라고 하드코딩 되어있고 Output Relation은 `assert sin(x1) == sin(x2)`로 적절하게 적혀 있다는 것이다. 본 연구에서는 이것을 효과적이지 못한 Metamorphic Testing이라고 본다. 왜냐하면 $x1 = 1.234$, $x2 = \pi - 1.234$ 외의 input에 대해서는 테스트를 하고 있지 않기 때문이다.

본 연구에서는 이런 Method를 입력으로 받아 자동으로 Input Transformation 함수를 만들어 효과적인 Metamorphic Testing을 하고자 한다. 예를 들어, $x1 = 1.234$, $x2 = \pi - 1.234$ 라고 하드코딩 되어있는 경우 `inputTransform(x) = \pi-x`라는 transformation 함수를 자동으로 추출하고자 한다.

본 연구는 이 문제를 풀기 위해 LLM을 활용한다. LLM을 통해 Input Transformation 함수를 만들어내려 한다. 효과적으로 만들어내기 위해 LLM에게 Input Relation을 만족하는 예제 여러개를 던져주려고 한다.

여기서 생기는 Challenge는 도대체 Input Relation을 어떻게 자동으로 추출하냐는 것이다. 예를 들어, input으로 두 개의 시간 값을 $x_1 = 01/01\ 00:00$, $x_2 = 01/02\ 12:00$ 주었다고 해보자. 개발자가 의도한 두 input의 관계는 무엇일까? 하루 차이를 의도한 것일까? 아니면 36시간 차이를 의도한 것일까? 이 부분에서 모호하다는 것이다.

본 연구는 이 문제를 굉장히 쉽게 접근한다. Output Relation을 보면 된다는 것이다. 앞선 예제를 생각해보면, Output Relation이 하루 차이를 검증하면 Input Relation이 그것을 의도한 것이고, 36시간 차이를 검증하면 이를 의도한 것이다. 굉장히 자연스러우면서 당연해 보이고 나도 생각할 수 있을 법 했다.

본 연구의 메소드는 다음과 같다. Output Relation을 기존 Test 메소드에서 추출한다. 이 Relation을 가지고 LLM한테 넘겨 Input Pair를 엄청나게 만들어낸다. Input Pair를 기존 코드의 Output Relation 코드를 통해 검증한다. 검증된 Input Pair들을 가지고 transformation 함수를 LLM한테 작성해달라고 한다. 해당 transformation 함수를 Input Pair로 옳은지 검증한다. 옳은 함수라면 그것이 곧 Input Transformation 함수가 된다.

본 연구의 한계라면 기존 Method가 적절하게 잘 작성되어 있어야 한다는 점이다. 얼마나 테스트들이 명확하게 작성되어 있는지는 확실하게 나와있지는 않지만, 이 부분만 해결하면 나머지는 바로 돌아간다는 점에서 좋은 평가를 받은 것 같다.

PAPER 2. EFFICIENT INCREMENTAL CODE COVERAGE ANALYSIS FOR REGRESSION TEST SUITES

본 연구는 다른 Domain에서 핵심 아이디어를 빌려와 문제를 해결한다는 점에서 좋은 연구라고 생각하여 공유하려 한다. 보통 다른 Domain에서 아이디어를 빌려오게 된다면 Challenge가 생기게 된다. 이 문제를 해결하고자 나오는 아이디어가 곧 논문으로 이어진다. 본 연구는 이러한 방향을 잘 보여주는 연구였다.

본 연구에서는 효과적인 Code Coverage Analysis를 하고 싶어한다. 현재 Code Coverage Analysis 기술은 작은 코드 변화에도 전체 분석을 다시 하기 때문에 비효율적이다. 효과적으로 하기 위해서 작은 코드 변화에 영향을 주는 테스트만 돌려

Code Coverage Analysis를 하면 된다. 이것은 곧 Regression Test Selection 문제와 동일 해진다. Regression Test Selection 문제가 하고자 하는 것이 변화에 영향을 주는 테스트만 골라 테스트 해보는 것이기 때문이다. 여기서 의문이 들었다.

“그러면 그냥 Regression Test Selection 문제를 푸는거 아닌가?”

그렇지 않았다. Regression Test Selection 문제를 푸는 방식으로 Code Coverage Analysis 문제를 풀게 되면 실제 골라야 할 테스트보다 적게 골라진다. 이를 예제와 함께 보여주며 독자로 하여금 Motivating 시켜준다. 그리고 연구의 방향은 기존 Regression Test Selection 알고리즘을 어떻게 확장해서 실제 골라야 할 테스트를 고를 수 있을까로 향하게 된다. 해당 문제를 풀어 논문으로 이어진 것이다.

KEYNOTE 1. ANALYSIS OF PERCEPTION NEURAL NETWORKS VIA VISION-LANGUAGE MODELS

해당 Keynote 발표에서 기존 기술을 다른 관점으로 사용하여 문제를 해결하는 흥미로운 부분이 있어 공유하려 한다.

여기서 풀고자 하는 문제는 Property 기반의 DNN Verification이었다. 특히, 이미지 도메인의 검증을 다룬다. 예를 들어, MNIST 데이터를 분류하는 모델이 있다고 해보자. 만약, '커브가 있으면 1은 절대 아니다'라는 프로퍼티가 있으면 이 프로퍼티를 가지고 분류 모델을 검증해볼 수 있을 것이다.

그런데, 문제는 이렇게 의미 있는 프로퍼티를 어떻게 만들어 내냐는 것이다. 나는 이 질문을 듣자마자 거의 불가능한 것 아닌가 싶었다. 하지만 바로 이어지는 Solution을 듣고 정말 괜찮은 발상이라고 생각했다. 바로 Vision-Language Model (VLM)을 사용하는 것이다. 이 모델은 이미지와 자연어가 얼마나 일치하는지 알려주는 모델이다. 해결책이 바로 떠오르지 않는가? 고양이와 트럭 이미지를 주고 'fur'와 'wheel'이라는 property를 던졌을 때, 고양이와 fur이 높은 점수로 그리고 트럭과 wheel이 높은 점수로 나올 것이다. 가장 높은 점수를 받는 property가 해당 이미지 클래스의 property인 것이다.

이처럼 문제를 푸는 방법이 바로 납득이 되는 연구들이 좋은 연구라고 생각이 들었다.

나의 발표와 ASE

만반의 준비와 정신없던 발표

이번 학회는 논문을 제출하여 발표까지 하는 자리였다. 영예로운 일이지만 그와 함께 출발 전 한국에서부터 너무 긴장됐다. FSE 2022 때 영어듣기 실력이 낮아 질문을 이해하지 못해 어버버 하던 기억이 스멀스멀 올라와 '또 그러면 어찌지'라는 생각 때문이었다. 옆에서는 '그럼 뭐 어때'라고는 하지만, 발표를 유창하게 하지 못하고 떠들떠들 하면 연구가 못 미더워 보이고 자신감도 없어 보이지 않는가. 그래서 이번엔 절대 그런 일이 없도록 만반의 준비를 했다.

첫째로, 영어 실력을 더 높이려고 했다. 특히, 영어듣기 실력을 높이려고 했다. 하지만 짬 내서 공부하기란 여간 쉬운 일이 아니었다. 어떤 자료를 사용해야할지 또 얼마나 공부를 어떻게 효율적으로 해야할지 갈피도 잡지 못했다. 그래서 특단의 조치를 취했다. 내 취미에 영어를 스며들게 하는 것이다. 넷플릭스로 애니메이션을 자주 보곤 했는데, 모든 애니메이션의 음성을 영어로 바꾸어 버렸다. 어색할 줄 알았지만 북미 성우들의 퀄리티가 예전과는 비교할 수 없을 정도로 높아져 굉장히 자연스러웠다. 과장해서 말하자면, 영어 더빙을 전제로 만든 것 같았다. 재밌는 점은 일본 애니메이션이더라도 '넷플릭스 오리지널' 작품이라면 99% 영어 음성 지원이 되는 것이었다. 대규모의 자본이 곧 복지로 이어진 셈이었다. 이렇게 매일 2~3편 보고 자면 자연스레 1시간 영어 공부를 한 셈이었다.

둘째로, 내가 생각할 수 있는 모든 질문에 대해 답변을 준비했다. Review로 왔던 코멘트는 물론 내가 발표 자료를 만들며 시간 문제로 빠트렸던 디테일한 요소들을 모두 질문형태로 만들어 영어 답변을 작성하였다. 사실 첫 발표(FSE'22)에서도 질문을 준비해갔지만 내가 준비해 간 질문 리스트에 있는 질문은 한번도 들어오지 않았다. 그 때 깨달았던 것이 저자가 생각하는 질문 포인트랑 청자가 생각하는 질문 포인트가 다르다는 것이었다. 청자는 처음 듣는 내용이다보니, 나는 당연하다고 생각했지만 다른 사람들에게는 전혀 그렇지 않은 부분들에 대해서 질문들이 들어온다. 나는 이것을 교훈 삼아 이번에는 모든 사소한 부분에 대해 질문을 만들었던 것이다.



덕분에 이번 발표는 마무리는 별탈없이 스무스하게 넘어갔다. 들어온 질문조차도 내가 생각한 질문 리스트에서 나와 마음 편하게 답변을 했다.

사실 발표의 시작에 탈이 있었다. 먼저, Session Chair분께서 정시에 오지 않으셔서 한 2분간 뿔뿔하게 서있었다. 봉사자가 와서 일단 시작은 하되, 시간은 최대한 맞춰보라 해서 시작부터 식은 땀 뻘뻘 흘리며 발표를 시작했다.

또, hdmi to c 젠더가 없어서 흥신 교수님 학생 분께 빌렸었다. 그런데 빔 프로젝터에 연결해보니 이게 웬 걸, 디스플레이 확장은 안되고 미러링만 되는게 아닌가. 그러면 스크립트가 띄워지질 않는 문제가 생긴다. 스크립트 자체는 외우고 있어서 막상 해보면 괜찮았겠지만, 저 때는 발표를 망칠 수 있겠다는 생각에 패닉이 왔었다. 그래서 호다닥 다른 한국인 분들에게 다른 젠더 있냐부터 해서 한번만 도와달라는 등의 도움을 요청했다. 사람들을 불러놓고 보니 갑자기 맥북 터치바가 시야에 들어오면서 순간적으로 '어? 여기에 디스플레이 확장이 뜨나?'라는 생각이 스쳐지나 갔다. 실제로 그랬고 사람들을 다 불러모았지만 1초만에 해결해 버렸다. 발표가 끝나고도 이런 문제들 때문에 정신이 없어 젠더를 빌려준 분께 내가 젠더를 돌려줬는지 아닌지도 헛갈려 우왕좌왕 했다. 심지어는 젠더 빌려준 보답도 못했는데, 다음에 만나게 된다면 꼭 커피라도 사드리리다. 혹은 이 레포트를 우연찮게 보게 되어 메일이라도 주시면 기프티콘을 선물해 드리려 한다.

기대했던 마음과 어수선한 학회장

이번 ASE에는 재밌는 주제들이 생각보다 많았다. 내가 은연 중에 생각하고 있던 문제들이나 이런 게 있으면 좋겠다 싶은 도구들이 발표로 있었기 때문이다. 그래서

굉장히 들뜬 마음으로 전날에 무슨 발표를 어떻게 들어야 효과적으로 들을 수 있을지 동선을 짰었다. 그렇게 기대하는 마음을 가지고 발표장에 임했다.

그런데 분위기가 조금 이상했다. Session Chair가 발표자를 찾는데 발표자가 보이지 않는다. 세션이 시작하기 전까지 모습을 보이지 않아 '시간 딱 맞춰 오려는건가?' 싶었다. 하지만 시간이 다 되었는데도 발표자는 오질 않아, 그 발표 자체가 없어져 버렸다. 나는 매우 이상하다고 생각하면서 이렇게 진행이 되어도 괜찮은가 싶었다.

진실은 중국의 비자(VISA) 문제였다. 미국에 입국하려면 반드시 비자가 필요한데, 이게 발급이 안되었다고 한다. 그래서 발표의 1/4은 영상 대체가 되었고 1/4은 노쇼 형태로 진행되었다. 이런 상황은 모두가 처음 겪어봐서 참가자들도 당황하고 Chair들도 진행을 어려워하는 모습이 보였다.

이러한 어수선한 모습 때문에 기대했던 마음은 큰 상실감으로 돌아왔다. 이 시기의 국제 정세를 보면 이 상황은 더 심각해질 것 같은데, 앞으로 학회는 이 문제를 해결하기 위해 어떤 해결책을 낼지 궁금해진다.

세크라멘토

세크라멘토는 놀랍게도 캘리포니아 주의 수도지만 아무 것도 없는 게 특징이었다. 나중에 들어보니 가난한 도시여서 노숙자들만 많고 인프라들이 매우 낙후되어 있다고 한다. 더 심했던 것은 호텔 WiFi를 이용하는데 돈을 지불해야 했었다. 나는 다행히 무제한 데이터로 신청해서 큰 문제는 없었지만, 나름 학회에서 지정한 4성급 호텔이었는데 이래도 되는가 싶었다.

그래도 아무 것도 없는 덕분에 얻은 것도 있었다. 주변에 할 게 없다보니 내가 갔던 그 어떤 학회보다, 사람들을 많이 만나 이야기를 할 수 있었다. 이야기를 해보며 느낀 점은 한국의 위상이 굉장히 올라갔다는 점이다. 연구 분야 쪽으로도 많이 올라왔지만 최근 문화적으로 한국이 그렇게 인기가 많다고 한다. 어느정도였냐면 우리가 군대를 다녀온다는 사실까지 알고 있었다.

주변에 뭔가가 없어 사진을 남기지는 못했지만, 그래도 덕분에 사람들과 많은 이야기를 할 수 있어 꽤나 좋은 경험이었다. 이렇게 값진 경험을 할 수 있게 도와주신 교수님께 감사의 인사를 드리며 글을 마치려 한다.

감사합니다, 교수님.