

# 학술대회 참관기

2012010922 이영섭

작년 11월부터 시작했던 자동채색 프로젝트의 논문이 한국정보과학회에 통과되어 학술대회 발표를 위해 6월 26(수) 부터 28일(금)까지 3일간 제주도 중문 ICC에서 열리는 정보과학회 학술대회(KCC 2019)를 참가하게 되었다. 사전등록 기준으로는 경상대학교에서 혼자 참가한다고 되어있었기 때문에 혼자 다니기 아쉬워 이번 기회를 통해 네트워크를 늘려보고자 학술대회 시작 전 Facebook 페이지인 tensorflow-kr 커뮤니티를 통해 KCC 참가하는 사람을 조사했다. 개인적으로 처음 참가하는 학술대회라 어떤 일을 해야 하는지 학회에서는 어떤 일을 하는지 정보가 없어 기대 반 걱정 반이라 다른 사람들은 어떻게 학회를 참관하는지 보려는 생각도 있었다. 일단 포스터 세션을 위주로 인공지능 및 네트워크 모델 경량화를 통한 성능 최적화 관련 논문들 위주로 찾아가며 질문하고 힌트를 얻었다. 공통으로 질문했던 내용은 첫째 데이터수집 방식, 둘째 데이터 증식(argumentation) 셋째 이미지의 경우 평가 방식"을 질문했다. 대부분 크롤링을 통해 수집하는데 드물게 기업과 연계된 연구실의 경우 기업 제품 테스트를 하며 데이터를 만들거나 처음부터 기업에서 데이터를 제공한 곳이 몇몇 있었다.



숙소는 학회가 열리는 중문 제주 ICC에서 400m 떨어진 곳을 예약했다. 가격도 저렴하고 위치도 부영호텔을 제외하고 나머지 호텔에 비해 가까웠기 때문이다. 문제는 첫날이 였는데 폭우가 쏟아지는 탓에 10시 이후 도착했다. 폭우 탓인지, 늦은 탓인지, 접수 데스크에 사람이 거의 없어 빠르게 접수를 하고 김진현 교수님을 만났다. 교수님과는 현재 창원경상대병원과 진행하고 있는 의료데이터 처리방식, 향후 계획 그리고 도메인설명을 위한 미팅 일정에 관해 얘기했다. 이후 포스터 섹션에서 흥미를 끄는 논문들을 찾기 시작했다. 첫날 유전알고리즘을 사용해 강화학습을 진행한 포스터가 인상 깊었는데 자세한 설명은 뒷장에서 다시 설명하려고한다. 아쉬웠던 점은 첫날 저녁쯤이 돼서야 커뮤니티에서 참가하는 사람이 있다고 연락이 와 첫날을 혼자 다녔다. 점심 이후 튜토리얼은 참가 신청하지 않았기

때문에 다음날 있을 발표를 위해 4층에 있는 발표장소와 마지막 포스터 섹션을 확인 후 다음날 발표 준비와 호텔에서 근무하는 친구를 만나기 위해 숙소로 돌아갔다.



둘째 날 다행스럽게도 비가 그쳐 단정한 차림으로 행사장에 도착할 수 있었다. 발표는 4층 발표장에서 진행되었으며 같은 섹션에서 초청논문을 제외하고 6개 중 5개의 논문이 GAN을 사용한 것으로 보아 발표 섹션은 비슷한 주제(or 도메인)나 네트워크 모델을 사용한 논문을 기준으로 정해졌다고 생각된다. 초청 논문 때문인지 25명 정도 발표장에 있었는데 발표 순서가 초청 논문 다음 발표였기 때문에 평소보다 더 긴장되었다. 발표와 받은 질문에 관한 자세한 설명은 뒷장(6장)에서 설명한다. 아쉬운 점은 김진현 교수님 발표, 그리고 커뮤니티에서 연락이 온 Naver Clova 인턴인 지영채님의 발표 모두 같은 시간 섹션에 발표를 진행해서 모두 다 볼 수 없었다. 발표 섹션 진행 중에 본인 발표 후에도 섹션 사람들이 한 명도 나가지 않아 세션 끝까지 남아야 한다고 생각했는데 3일 차가 돼서야 본인 발표만 하고 나가도 된다는 것을 알게 되었다. 관심 있던 다른 섹션 발표를 보지 못한 점과 오전 발표 중간에 김성훈 교수님(모두를 위한 딥러닝 제작자)을 만날 기회를 놓친 점이 매우 아쉬웠다.



지영채님과 식사 후 요즘 진행하는 토이프로젝트 및 tflite 모델 변환(모바일용)을 얘기하고 김성훈 교수님(현재는 Naver Clova 리더)의 발표를 듣기 위해 같이 이동하였다. 교수님 발표에서, 시작 후 10분간 본인의 합성음을 사용해 발표를 진행 한 부분이 인상적이었다. Clova에서 만들고 있는 새로운 TTS (text-to-speech) 모델을 사용해 생성한 합성음이고 놀랍게도 학습 데이터양이 40분, 500음절 사용했다고 하였다. 지난 방학 중 20시간의 JTBC-손석희 앵커 데이터를 사용해 학습시킨

tacotron 모델을 받아 돌려본 적이 있었는데 그보다 훨씬(압도적으로) 자연스러운 음을 합성하는 것이 놀라웠다. 설명으론 시간이 날 때마다 조금씩 작업하고 감기에 걸린 상태였기 때문에 다시 데이터를 만들어 학습시키면 더 자연스러운 결과가 나올 것이라고 설명을 덧붙였다. 이후 인상적인 부분은 네이버에서 최근 발표한 논문(논문인 거 같습니다.)은 17년 제주도에서 했던 인턴쉽(인턴쉽인지 별개의 프로젝트인지 잘 모르겠습니다. 네이버로 간 시점이 17년 3월로 알고 있긴 한데 17년 초반에 진행했다고 했던 프로젝트였다고 했고, 타 전공 학생뿐만 아니라 1~2년 개발 경력을 가지고 있는 분도 참가해서 매일 3~5분씩 피어 리뷰를 했다고 합니다. 실험적인 프로젝트라고 언급했기 때문에 정규 인턴쉽이 아닐 수도 있습니다.) 프로젝트팀에서 시작된 결과물이 38%라는 점이였다. 이외 여러 흥미로운 정보가 있었는데 자세한 정보는 뒷장에서 언급하겠다. 발표 직후 바로 앞자리에 김성훈 교수님이 앉으셔서 잠시 이야기 할 기회가 있었다. 이후 뱅킷은 표가 없어서 참가하지 못하고 마지막 포스터 섹션을 확인하고 바로 숙소로 돌아갔다..

마지막 날은 짐을 정리하고 들고 다녀야 했기 때문에 다른 날 보다 늦게 11시에 출발 했다. 이날은 발표 섹션 위주로 둘러보았는데, 전날 발표를 보러 와 질문을 했던 "식질머신" 제작자 고우람님의 발표가 오후에 있었습니다. 발표 직전까지 단일 신경망에서 End2End로 진행할 것으로 생각했지만, 전혀 다른 접근 방법이였기 때문에 흥미로운 내용이었다. 발표에 대한 자세한 내용은 뒷장에서 다시 설명하겠다. 오후시간에는 흥미를 끌었던 몇 가지 발표를 참관하고 블록체인 경진대회를 중간부터 지켜봤다. 대회에서는 주유소 사기(주유기 조작 및 성분 조작)를 방지하기 위해 센서 정보를 블록체인으로 관리하는 제품을 개발한 팀이 가장 사업화에 가깝게 구체적으로 진행되어 있어 인상 깊었다. 블록체인에 대한 지식이 없어 자세한 설명은 생략한다. 경품 추천은 재미로 참가했는데 운 좋게 2등에 당첨되어 라인 캐릭터 노트북 가죽 슬리브를 받았다. (1등은 Clova 인공지능 스피커!) 이후 공식적인 학술대회 일정은 모두 끝났으며 공항 근처의 숙소로 이동해 참관했던 내용을 정리했다.

## 참관 목적

공식적인 목적은 KCC2019에 통과된 Oral 섹션 논문(신경망과 사용자 힌트를 사용한 애니메이션 스타일 일러스트 자동채색) 발표를 위해 참관했다. 몇 달 전까지만 해도 논문을 내고 학회를 간다는 것은 다른 사람들의 얘기로밖에 생각되지 않아 무엇을 해야 하고 어떤 것을 봐야 하는지 잘 알 수 없었다. 논문발표 이외의 목적으로는 1. URP 및 개인 프로젝트를 위해 quantization을 비롯한 네트워크 모델 경량화 기법, 2. 다른 사람 무슨 아이디어를 가지고 있는가?, 3. 다른 사람들의 Hyperparameter 최적화 작업 및 데이터 수집 및 관리 방법을 알아보기, 4. 커뮤니티 사람들과의 네트워킹 등이 있었다.

발표 및 포스터 논문의 대부분이 pytorch를 사용하여 진행해서 모바일 내장을 위한 경량화 기법은 찾을 수 없었지만 Hyperparameter를 찾아가는 방식은 대부분 비슷하게 많은 실험으로 얻어진 결과라는 것을 알 수 있었다. 기업과의 연계로 프로젝트를 진행하는 것 중에서 신발에 압력 센서를 설치하여 압력 패턴으로 사람의 행동을 예측하는 모델을 만든 논문이 있었다. 압력 패턴을 1차원으로 만들고 1D convolution을 통해 예측하는데 커널 사이즈를 20으로 확정하기 위해서 많은 실험을

거쳤다고 했다. 작년 말부터, 네트워크가 중요하다고 생각되어 관심을 갖고 커뮤니티 활동을 하였는데 많은 사람을 만난 것은 아니지만 커뮤니티의 도움과 생각지도 못했던 한 분을 만나게 되어 성공적으로 목적을 달성 했다고 생각한다.

## 만남 사람

### 고우람: 식질머신: 딥러닝을 이용한 만과 글자 제거 시스템, 고우람, 조환규(부산대)

다른 친구를 통해 고우람님의 Discord 연락처를 받아두었지만, 시간이 없어 연락을 못 하고 있었다. 고우람님은 이미 GitHub와 다른 친구를 통해 내 프로젝트를 알고 있는 상태였고 학회 책자에 논문 제목과 학교를 보고 발표장으로 왔다고 하였다. 답하기 곤란한 질문을 계속해서 누군지 궁금했는데 발표 직후 그쪽에서 먼저 인사를 해 서로 알아보게 되었다. 내 발표에 관해서, 왜 크기를 고정해서 진행했는지 계속 궁금해했고 학습을 시킨 다음에는 크기에 상관없이 진행해도 결과물이 채색이 되지 않는냐고 물어봤는데 이전 실험 결과 그런 방식으로 진행해도 선화 특징이 변경되는 탓인지 초안 모델에서 채색이 안 된다고 알려주고 문제를 해결하기 위해 여러 방식을 얘기했다. 결론은 학습은 1차: 256Px, 2차: 512 Px 이후 모델을 사용할 때는 256 Px 에서 초안을 생성하고 2차 모델에서 크기에 상관없이 진행하는 것을 생각해봤다. 또 한 발표한 모델의 총 학습 시간을 어떻게 3일로 단축했는지 궁금해했는데 이 부분은 딱히 근거가 없이 실험으로 만들어진 결과라서 답을 하지는 못했다. 그의 경우 원하는 Loss가 나올 때 까지 3주간 무한으로 돌려 만들어진 결과라고 한다.

### 지영채 (Naver Clova 인턴) : 청각장애인을 위한 UNet-LSTM 기반의 자동 음성-수어 번역 모델, 김윤기(동탄고), 지영채(한양대), 하정우(네이버)

tensorflow-kr 커뮤니티에서 학회 참석한다고 올려둔 분이 있어 연락했다. 워낙 커뮤니티 활동이 활발한 분이라 이전부터 김윤기 학생과 같이 이름은 알고 있었다. 직접 만나기 전 당연히 4학년(혹은 대학원) 인줄 알고 있었는데 1학년이라는 사실에 많이 놀랐다. 김윤기 학생의 경우 3월 GDG 판교 tensorflow Dev Summit에서 한번 만났는데 이번에는 학교 시험 때문에 KCC2019에 참석을 못 했다고 했다. 세션 시간이 같은 시간이라 끝나고 식사나 하자고 연락을 했었는데 나오는 도중에 만났다. 본인은 미리 나와 김성훈 교수님과 얘기를 하고 있었다고 한다. 식사 후 어떻게 네이버 인턴을 하게 됐는지 물을 기회가 있었는데 대부분의 머신러닝 관련 행사를 참가하면서 커뮤니티 활동을 하다 보니 김성훈 교수님 쪽에서 먼저 연락이 왔다고 했다. 인턴을 하고 싶다면 정규 인턴십을 통과하는 건 어려운 일이고 직접 메일을 쓰는 방법이 의외로 가능성이 있다는 정보를 얻을 수 있었다. 이후 요즘 하는 프로젝트나 구상 중인 프로젝트에 관해 의견을 주고받았다. 발표한 모델은 네이버 인턴소속으로 발표한 논문이며 스펙트로그램 형태로 뉴스 음성을 입력으로 U-Net으로 압축된 정보를 생성한 다음 이를 LSTM의 입력으로 한다. 출력값은 수어 표현으로 위해 124개 2차원 좌표 영상 프레임을 생성한다고 한다. 모델 학습을 위해 5시간의 수어 영상이 포함된 공개 뉴스를 사용했다고 하며 Naver의 지원으로 50개의 GPU를 사용해 진행했다고 한다.

## 새롭게 배운 것

### 1. Network Pruning

성능을 유지하면서 네트워크를 줄이는 최적화 기법이라고 하는데 모델 내부 값의 정밀도를 조절하는 Quantization과는 달리 정해진 기준에 따라 변수를 줄인 뒤 재학습 시켜 성능은 비슷하게 유지하면서 사이즈를 감소시키는 방법이라고 한다. 해당 논문을 쓴 정종훈 씨는 슈퍼 샘플링 분야에 대해 복원도를 유지하면서 변수의 양(모델 크기)을 71% 정도 줄이고 인퍼런스 시간(실행 시간? 한 장을 모델이 생성하는 시간)은 기존 시간 대비 60% 증가했으며 정확도는 PSNR 기준으로 94%로 유지되었다고 한다. 해당 기법이 채색모델에서는 얼마나 효율적일지는 실험을 해봐야 알겠지만, 새로운 기법을 알 수 있었기 때문에 의미가 있었다. hyperparameter 최적화 과정에서 레이어 수를 줄이거나 비슷한 방법을 사용해 병목 레이어를 만들어 경량화를 진행하는 사례는 봤는데 프루닝과 같은 사례는 처음 알게 되었다. 문제는 본인 구현체는 pyTorch를 사용했기 때문에 Tensorflow에 적용하기 위해서는 조사가 필요하다.

### 2. TCNN: TEMPORAL CONVOLUTIONAL NEURAL NETWORK

지영채 님과 대화 중 얻은 힌트인데 유비트 채보 자동생성 프로젝트에 관한 얘기를 진행 중 배웠다. 4x4를 1/16 박자로 진행되는 리듬 게임인 유비트 특징상 채보가 올라오는 이미지 형태도 중요하고 이전 상태에 따른 이어지는 현재 패턴도 중요했기 때문에 학습시킬 데이터 형태를 고민 중이라고 하니 TCNN을 추천했다. 자세한 내용은 아직 조사를 덜 진행해 설명할 수 없지만 residual block을 사용해 여러 스택을 만들어 데이터 경로는 skip 하는 방식으로 모델을 구성하는 것 같다. 구현체는 이미 찾았으며 향후 실험할 데이터가 만들어지면 진행해볼 생각이다.

### 3. TensorWatch

프레임워크 의존성이 없는 AI 디버깅 툴 그래프, 차트, 테이블, 이미지 등 다양한 포맷의 출력이나오며 최근 오픈소스화되어 올라온 tensor\_board의 사용자화가 조금 힘든 부분이 있었는데 이번 프로젝트를 통해 TensorWatch를 사용해볼 생각이다. 지영채 님과 대화 중 머신러닝 디버깅 툴에 관해 말하다 얻은 새로운 정보다. MS에서 만들고 있는 오픈소스 머신러닝 디버깅 툴이며 간단한 사용법으로 프레임워크를 가리지 않고 쓸 수 있는 실시간 디버깅 툴이라고 한다. tensorflow 2.0 정식 버전 출시 전까지 사용해 보려 생각하고 있다.

### 4. 이미지의 정량적 평가 방식(SSIM, PSNR)

영상 분야에 관심이 있는데 기법의 평가를 위해 정량적으로 사용하는 평가 방식에 대한 지식이 많이 없어 이와 관련해 많은 질문을 하고 다녔다. 다수의 논문이 MSE, 오차 함수, L1\_loss 등의 방법을 사용하는데 SSIM, PSNR 등의 방법도 사용하는 논문이 몇몇 있어 질문 결과 FID 와 같이 자주 사용하는 평가 지표라고 하여 이후 프로젝트에도 평가방식으로 적용해 보려 한다. SSIM은 두 이미지의 구조 정보(밝기, 대비, 구조)를 사용해 비교하는 평가 방식이라고 한다. PSNR은 MSE 값을 사용해 이미지 신호의 최대 전력에 대해 잡음의 전력을 계산한다고 한다. (구글 검색 결과)

## 포스터 및 발표 세션



가장 먼저 본인의 발표 리뷰를 시작으로 포스터 및 발표 섹션 리뷰를 하려 한다. 첫 학회 참석이라 많이 긴장했었는데 발표 자체는 무사히 진행됐다. 받은 질문으로는 좌장 질문을 포함해 총 4개의 질문을 받았는데, 좌장 질문을 다시 해석해 보니 발표 때 프로젝트를 진행했다는 단어와 논문이라는 단어를 혼용해서 썼던 것이 문제였던 것 같다. 첫 질문은 해상도를 늘리는 방향으로 진행한 이유를 물었다. 업스케일링을 하면 정보가 손실되어 정확도가 떨어질 텐데 왜 그런 방향으로 진행했냐는 것이다. 사실 기존에 몇 가지 공개된 모델이 있었고 작년 11월 기준으로는 총 3개의 서비스(그중 paintschainer 는 상용화)가 있었는데 모든 서비스가 최대 해상도 512x512를 벗어나지 못해 만들었다고 답했다. 두 번째 질문은 “식질 머신”을 만든 고우란 님의 질문이었는데 Full CNN을 사용했는데 왜 굳이 모델 사용 시에도 크기를 고정해서 사용하고 업스케일링으로 크기를 늘렸냐는 것과 제시한 학습 데이터로 얼마나 학습을 했느냐는 것이다. 먼저 크기를 고정한 이유는 크기를 고정하지 않는다면 선화의 질(두께로 예상)이 바뀌는데 그러면 모델의 채색 안정성이 너무 떨어져서 고정해서 사용한 것이라고 답했다. 학습 시간은 시연한 영상의 모델의 경우 3일이 걸렸다고 답했다. 세 번째 질문은 비슷한 화풍의 이미지만을 사용해 학습을 진행한 것 같은데 다른 화풍이나 이미지를 사용한다면 채색이 되는지 궁금하다고 했다. 선을 이해하고 채색을 진행하는 것은 아니라고 생각되었고 내 이미지를 선화로 만들어 채색했을 때 채색이 진행됐던 것을 생각하면 선화이기만 하면 칠을 한다고 답했다. 마지막 좌장 질문은 답을 하지 못했는데 기존 기법을 잘 정리했지만, 기존 기법과 비교하는 평가 방식이 공평한것인지 모르겠다고 했다. 평가에 사용한 데이터도 다르고, 결과물 사이즈도 모든 모델이 동일하지 않고, 힌트를 반영했을 때 생성되는 이미지에 대한 평가가 없다고 한다. 좋은 논문을 쓰기 위해서는 이후 작업에서 공평한 평가 방식을 만들어 사용하는 것이 좋을 거 같다는 의견을 마지막으로 내 발표는 끝이 났다.

첫날 GA(유전자알고리즘)와 GAN으로 문장을 생성하는 논문<sup>1</sup>이 있었다. 학습을 강화학습으로 진행했다고 해서 어떻게 진행되는 건지 감을 잡을 수 없었고 관련 지식이 없고 GA를 데이터 증식에 사용한 거 같아 보이지 않아 GA를 어떤 식으로 GAN에 적용했는지 질문을 했다. 에폭마다 변이 부여 여부와 어절의 일부를 랜덤으로 정의한다고 한다. (메모했는데도 이해를 못 하겠습니다.) 이후 질문은 GAN도 안정적인 학습이 어려운데 GA를 사용하게 되면 학습을 진행하면서 변이 전에 만들어진 학습 기울기 정보들이 대부분 의미가 없어서 지속하게 되면 학습이 어렵지 않냐고 질문을 했는데 자연스러운 문장을 생성하는 부분에 있어 어려움이 있기 때문에 보완하고 있다고 했다.

두 번째 흥미로운 논문<sup>2</sup>은 Twitch.tv 에서 편집하지 않은 전체영상에서 자동으로 사용자들의 흥미를 유발하는 인기 구간(클립) 추출을 자동화하는 모델이었다. 흥미로운 점은 메타데이터뿐만 아니라 동영상의 썸네일을 CNN 블록으로 특징을 추출해 합쳐 사용했다는 점이였다. 처음엔 왜 영상이 아니라 썸네일을 사용했는지 질문했는데 답변으로는 보통 사람들이 영상을 클릭할 때 제목을 보지만 많은 경우 영상의 대표 썸네일도 같이 보기 때문이었다. 유튜브에서 자동으로 영상의 썸네일을 추천해 주는 걸 생각하면 당연한 이유였다. 모델의 구조는 메타데이터, 썸네일, 클립 제목을 각각의 네트워크를 사용해 특징을 추출하고 뒷단에서 합쳐 인기 클립, 비인기클립을 예측하는 분류 문제로 정의했다. 테스트 셋을 사용한 정확도 측정 결과 0.859의 정확도를 얻었다고 한다.

마지막은 "식질머신"이다. 초청 논문을 제외하고 참가했던 세션 중 가장 반응이 좋았던 발표였다. 단일모델로 어떻게 깔끔하게 제거했지 라는 생각과는 달리 1단계로 CNN을 사용해 텍스트 영역에 Segmentation 작업을 거치고 U-net에 집어넣었다. 시연 영상에서 1300x600 크기 이미지를 6.3s 만에 처리하는데 Segmentation으로 처리할 영역을 정하고 네트워크에 집어넣기 때문에 이미지 크기나 상태와 관계없이 진행할 수 있었던 걸로 생각된다. 질문으로는 그러면 최대 얼마나 크게 가능한지 물어봤는데 분할-정복 기법을 사용한다면 원하는 대로 크기를 키울 수 있다고 한다.



<sup>1</sup> 유전 알고리즘을 이용한 텍스트 GAN, 오진영, 차정원(창원대학교)

<sup>2</sup> 온라인 스트리밍 동영상 인기 구간 예측, 송현호(KAIST), 박건우(카타르 컴퓨팅 연구소), 차미영(기초과학연구원 데이터 사이언스 그룹)