

개인 회고

들어가며

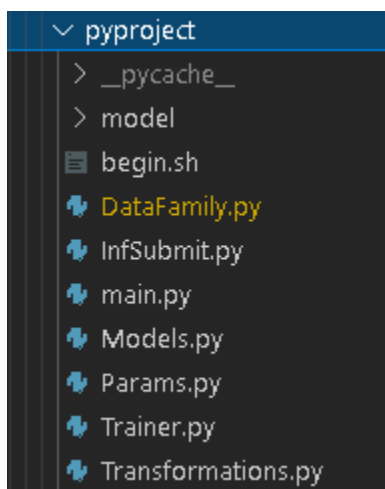
이번 첫번째 대회는 사뭇 낯선 문제를 다루는 문제였습니다 .

마스크를 쓰고 안쓴 것만을 판단하는 대회였다면 이야기가 수 배는 간단했을 텐데, 그렇지 않고 나이와 성별을 가려야 했기에 쉽지 않은 대회일 것이라는 두려움에 주눅든 채 시작했던 대회였습니다.

할 수 있던 것을 다 했다면 후회가 없으리라 생각하지만, 대다수의 사람들이 그렇듯 저는 회고를 쓰려하는 지금 제 최선을 다 하지 못한 기분이 듭니다.

조금은 버겁게 시작된 첫번째 P 스테이지는, 다행히도 훌륭한 강의와 친절한 팀원들 덕에 큰 문제 없이 시작할 수 있었습니다. 정리를 어떻게 시작해야 할지 크게 감이 오지 않는 중이지만, 일단은 강의의 흐름에 따라 제가 대회동안 팀에서, 또는 저 개인이 행했던 일에 대해 회고하는 것이 옳을 것으로 생각합니다. 강의를 지나고 난 뒤, 일반적으로 셀 단위 실행의 강력한 기능을 활용한 노트북 파일에서 파이썬 IDLE 프로젝트 방식으로 전환하는 작업을 했습니다.

저희 팀은 순위가 아닌 배움이 첫째라고 여겼고, 협업을 시작하기 전까지 각자의 방식으로 베이스라인 코드에서 벗어나 프로젝트를 구성해 보는 것을 목표로 했습니다.



템플릿에서 벗어나 전체적인 프로젝트를 짜보는 것을 목표로 했던 저이기에, 대회가 끝나고 1등이나 2등의 결과는 받지 못했지만 충분히 만족하였습니다.

```

main.py x
teamrepo > kbs > code > pyproject > main.py > ...
24 '''
25 if __name__ == '__main__':
26
27     hp = Parameters() #Loading hyperparams...
28     dataframe = DataFrame().dataframe #creating dataframe...
29
30     trainset = DataCluster(0, hp.img_dir, dataframe).set #loading trainset...
31     validset = DataCluster(1, hp.img_dir, dataframe).set #loading validset...
32     testset = DataCluster(2, hp.test_img_path).set #loading testset...
33
34     train_loader = DataLoader(trainset, batch_size=hp.batch_size, num_workers=4, shuffle=True)
35     valid_loader = DataLoader(validset, batch_size=hp.batch_size, num_workers=4, shuffle=True)
36
37     model = MyModel(hp.model_name, pretrained=True) #loading model...
38     optimizer = Adam(model.parameters(), lr=hp.initial_lr, weight_decay=hp.weight_decay)
39     lr_scheduler = CosineAnnealingWarmRestarts(optimizer, T_0=100, T_mult=1, eta_min=0.00001)
40
41     train = TrainandEval(20, 32, train_loader, valid_loader, model, optimizer, lr_scheduler)
42
43     modelresult = train.best_model
44     test_loader = DataLoader(testset, shuffle=False, num_workers=4)
45     inference = Inference(test_loader, modelresult, hp.device)
46     submission = Submission(inference)

```

시퀀셜한 객체 생성에 따라 프로젝트를 진행할 수 있도록 하였고, 실험의 용이성을 위하여 대다수의 하이퍼파라미터를 메인 파일에서 진행할 수 있도록 했습니다. 별도의 훈련 모듈을 만들어 진행한 것은, 지금의 메인 파일이 온전히 프로젝트 실행만을 위한 파일이 되기를 원했기 때문입니다. 이후 베이스라인 코드가 나왔을 때 협업을 위하여 움직였지만, 랩업 리포트를 쓰는 지금 그냥 넘어가기엔 아쉬운 작업이었습니다.

이후 팀 코드에서 작업하게 되었을 때까지만 해도 저는 잘 알지 못하였습니다.

실제 대회 또는 현업에서 또한 그렇다는 것을 이후 멘토님께 들었지만, 머신러닝 서비스 개발에 인공 신경망 모델 자체는 생각보다 작은 부분을 차지한다는 것을 말합니다.

다만 그것을 알았을 때, 저는 이미 일곱개의 모델과 네 개의 옵티마이저, 그리고 세개의 러닝레이트 스케줄러에 따른 결과값 비교를 마친 후였습니다. 이거 혹시 모델은 생각보다 그렇게 중요하지 않은게 아닌가 하는 생각을 하게 되었을 때 즈음, 멘토님께서 시원하게 생각보다 모델은 중하지 않다는 것을 알려주셨습니다.

말아서 알아본 작업이 큰 리턴을 얻지 못한 이후, 팀 코드에 기여하기 위해 편의성 코드를 작성하였습니다.

```
root@31b69b637581:~/teamrepo/team# ./29.sh
Wish you get 0.9 f1 score this time...
Please enter project name, if you don't ,it will be exp
for report
Do you want to train model one by one? Input anything
yes
Do you want to split wandb project one by one too? Input anything
no
Please enter model number from below or Input model name what you want.
=====
=====Model Menu=====
1. EfficientNetB3
2. Res2Next50
3. ResNext50
4. DenseNet121
5. Inception-ResnetV2
6. Inception-ResnetV1,FaceNet
7. I want another model
=====
4
You chose DenseNet121 model for mask
```

먼저 실험 실행 시 Argument가 길고, 많아, 자주 쓰는 옵션에서 조금씩 변형하여 실험하는 것이 번거로울 것 같아 쉘 스크립트 파일을 작성해서 남겼습니다.

이후 데이터셋 자체를 교차 검증을 위해 훈련 셋과 검증 셋으로 나누어 훈련한 모델을 제출 전에 한번 더 전체 데이터셋으로 훈련할 수 있도록 나누지 않고 전체 데이터셋을 이전 훈련에서 가져온 체크포인트부터 훈련할 수 있도록 코드를 작성하였고, 구동하였습니다. 다만 이 또한 결과에 큰 영향을 주지 못하여 최종 코드에 포함되어 있지 않습니다.

대회가 끝난 이후 다른 팀의 솔루션과 저희 팀의 코드를 비교하고, 스페셜 피어 세션을 통해 어떤 부분을 놓쳤는지 알아보았고, 안타까움을 금할 수 없었습니다.

상당히 초창기에 팀에서 논의되었던 라벨 스무딩이 아닌 나이 라벨링의 기준값 자체를 바꿔서 라벨링을 하자는 이야기가 있었는데, 해당 방법만으로 79%의 성능을 낸 조가 있었습니다. 논의했던 것을 모두 행하지 못하고, 옳지 못한 것에 주목하여 생각한 것을 모두 하기 전에 시간을 다 사용한 것을 가장 큰 아쉬움으로 가지고 있습니다.