

개인 회고록

목차

1. 나는 무엇을 했는가?
 2. 회고
 3. 배운것들
 4. 아쉬운 점
 5. 다음에는...
 6. 총평
-

1. 나는 무엇을 했는가?

1주차

- baselinecode 작성
- 초기데이터에 대한 eda
- 데이터 오류 검수
- sota model 찾기
- 간단한 모델 실험하기

2주차

- confusion matrix 작성 및 eda
 - public score, data에 대한 eda
 - age filter model 실험
 - ansemlcode 작성
-

2. 회고

1주차는 굉장히 혼란스러웠습니다. 딥러닝의 학습 흐름을 제대로 이해하지도 못했고, torch도 익숙하지 않고, 목표는 score이면서 동시에 학습이었습니다.

제가 생각할 때 대회를 위해서는 모든일을 효율적으로 처리해야 한다고 생각합니다. 팀원들끼리 철저한 분업을 하고 했던 일을 두 번하는 것을 최대한 지양해야합니다.

그리고 학습을 위해서는 처음하는 이들에겐, 이것 저것 다해보는 것이 맞다고 생각합니다. baseline도 만들어보고, eda도 해보고, model도 건드려보고 등등...

이러한 부분들이 충돌해서, 대회임에도 효율적으로 임하지 못하는 부분들이 있었습니다. 학습을 위해서 사실 저부터도 그랬습니다. baseline을 다뤄보지 않고서는 공부 제대로 되지 않을 것 같아 요한님과 같은 날에 dataloader와 dataset을 만들었습니다.

또한 모델을 넣기위해서는 torch를 공부해야했고, template을 공부하는 등 대회의 스코어에는 영향을 주지 못할 부분들에 시간을 쏟는 일이 많았었습니다.

아쉽긴 하지만 어쩔 수 없었다라는 생각과 다음 번엔 더 잘할 수 있겠다라는 생각이 듭니다.

팀에 기여한 부분

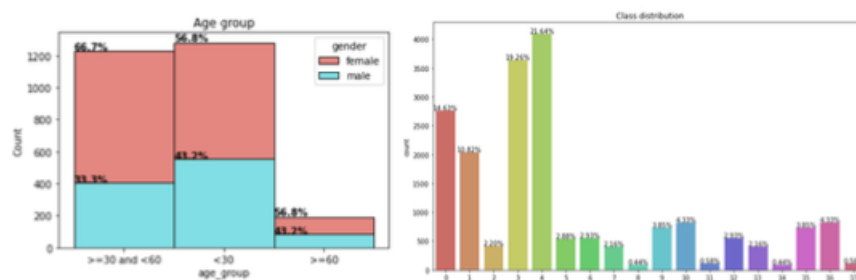
- 초기 데이터 검수
- confusion matrix를 통한 전체 모델의 약점, 강점 파악 및 해결 방안 제시
- eval dataset 분포와 submission 분포의 차이를 비율로 제시
- 이로 인한 f1 스코어 개선 방향 제시

3. 배운것들

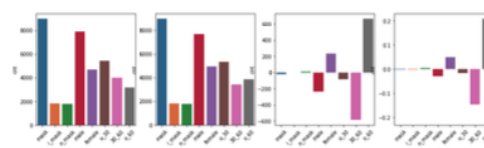
validation

score 제출은 1일 10회, 7명의 팀원들은 평균적으로 1.xx번 정도의 제출을 할 수 있습니다. 이 점은 모델의 성능이 어느정도 확보가 된 것들만을 제출해야하는 것을 의미합니다. 그리고 이를 위해서는 validation score가 필요하고, 좋은 validation set이 중요하다는 사실을 배웠습니다.

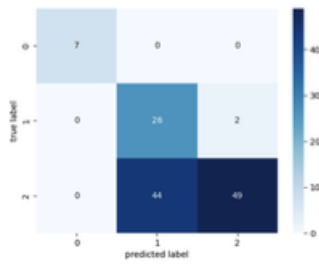
visualization



나이대별 성별 분포 및 class별 분포



퍼블릭 테스트 데이터에 대한 모델 시각화



confusion matrix를 통한 validation set에 대한 시각화

다양한 시각화가 굉장히 중요하다는 것을 느꼈습니다. 저희 모델 같은 경우 3가지 모델을 사용해서 각 class를 예측하는 방식을 사용했는데, visualization을 하기 전까지는 정확한 개별 모델에 대한 성능을 감잡기 어려웠으나, visualization 이후에는 마스크모델은 거의 완벽하게 해결을 하고있다는 사실을 알 수 있었고, gender에서는 나이 많은 여성, 남성을 헛갈려하는 것을 알 수 있었습니다. 또한 age에서는 60대의 경계선이 불분명해서 생기는 문제들이 바로 드러났습니다.

실험 결과 기록

실험의 결과에 대한 공유를 하려면 기록이 중요했습니다. 좋은 모델들은 추후에 재현할 수 있어야했고, 그렇게 하기 위해서는 정확한 하이퍼파라미터에 대한 기록, 실험의 설계에 대한 내용들이 있어야했습니다.

4. 아쉬운 점

- 첫 대회, 첫 딥러닝이기에 대회에 좀 더 초점을 맞출 수 없는 점들이 아쉽게 느껴졌습니다. 정말 좋은 팀원을 만났기에 제가 좀 더 익숙했다면 더 나은 시너지를 냈을텐데, 그러지 못한 것이 아쉽습니다.
- 3개 model로 class별 예측 후 합치는 것과 18 label을 model 한 개로 학습하는 것, 이 두개의 방법론에 대해서 우열을 찾아보고 싶은데 그러지 못했습니다. 또한 우열이 생긴다면 왜 생기는지 알고 싶다는 생각이 듭니다.

5. 다음에는...

- 기록을 잘 해야겠습니다.
- 데이터를 좀 더 잘 관리해야겠다는 생각이 듭니다. (중간에 추가된 데이터들의 경로들이 팀원들끼리 공유가 부족했던 것 같습니다. 한명이 데이터를 전부 관리하는 것이 낫지 않나 하는 생각을 했습니다.)
- 코드를 좀 더 유연하게 작성해야겠습니다.
- 협업을 좀 더 신경을 써봐야겠습니다.

6. 총평

- 즐거운 대회였습니다.

- 최고의 팀원, 최고의 멘토, 최고의 순간들이었습니다.
-