

2023년 K-ium 의료인공지능 경진대회

> 익명화된 뇌혈관조영술 영상을 기반으로 뇌동맥류 여부, 위치를 진단하는 소프트웨어 개발

한성대학교 컴퓨터공학부

팀명 : Almagest

김동준(팀장), 박경민, 박상준, 최유진

[요약]

PyTorch를 활용하여, 사전학습 모델인 "EfficientNet_b0"을 주최측에서 제공한 데이터셋을 이용해 전이학습한 뇌동맥류 진단 모델을 개발하였습니다. 데이터 전처리를 거쳐, 각각의 뇌혈관 위치와 이미지 데이터를 활용하여 학습을 진행하였습니다. target_column을 설정하여 각 위치마다 별도의 모델을 학습시키고, 검증 성능이 가장 우수한 모델을 최종 모델로 선정하는 방식을 채택하였습니다.

[방법]

➤ 데이터 전처리

크기가 다른 각 이미지를 640 x 640 사이즈로 통일시켰으며, 모델 입력을 위해 Pytorch tensor로 변환시켰습니다. 혈관의 위치정보 및 해당 위치의 명도가 중요한 흑백 이미지 데이터셋임을 고려하여, 별도의 이미지 증강은 적용하지 않았습니다. 원-핫 인코딩 방식을 통해 위치 정보를 벡터화 했으며, 클래스 가중치 조정 기법을 통해 '1' 클래스의 데이터 불균형을 완화하였습니다.

➤ 학습 및 검증

EfficientNet_b0 사전학습 모델을 활용하여 전이학습을 진행해 이미지에서 특징을 추출하였습니다. 또한, 혈관의 위치 정보를 원-핫 인코딩하여 특징 벡터와 함께 모델에 입력하였습니다.

train.csv의 각 혈관 위치열을 target_column으로 설정하여 개별적인 학습을 진행했습니다. 배치 정규화, 드롭아웃, 학습률 스케줄링 기법을 적용하여 모델의 일반화 성능을 향상시켰습니다. 드롭아웃 비율은 0.2로 설정하였으며, 초기 learning rate는 0.001, 가중치 감소 파라미터는 1e-5로 설정하여 과적합을 방지하였습니다. 최적화 기법으로 Adam을 적용했습니다. Batch는 32, epoch는 25로 각 학습을 진행했습니다.

엘리스탑 기법을 이용하여 성능향상이 되지 않는 경우에, 학습을 멈추고 최적의 성능을 보인 에포크의 모델을 저장하였습니다. 위의 과정을 통해, train.csv의 각 위치열 별로 환자 이미지 데이터(8장)에 대한 각각의 학습모델을 생성하고 해당 모델들의 검증 성능을 비교하여 가장 최적의 모델을 채택했습니다.

[결과]

■ C-statistics (AUROC)

```
41 def main():
42     auc, acc = eval_model()
43     print('AUROC of the provided model')
44     print (auc)
45     print('Accuracy for locations')
46     print (acc)
47
```

PROBLEMS TERMINAL OUTPUT DEBUG CONSOLE

```
PS C:\Users\Android\Desktop\AIImagine_Cal_AUC> & c:/Users/
(AI) PS C:\Users\Android\Desktop\AIImagine_Cal_AUC> & c:/U
exe c:/Users/Android/Desktop/AIImagine_Cal_AUC/evaluate.py
AUROC of the provided model
0.9710144927536232
Accuracy for locations
0.9638095238095238
```

주최측이 제공한 evaluate.py을 이용하여 정답 값인 groundtruth.csv과, 모델이 도출해낸 output.csv 파일을 비교하여 C-statistics(AUROC)를 출력했습니다.

그 결과,

AUROC of the provided model : 0.9710144927536232

Accuracy for locations : 0.9638095238095238

모델 AUROC : 0.971, 위치에 대한 정확도 : 0.9638 값을 도출해 냈습니다.

※ 추론코드에 대한 실행방법은

README 모형 검증방법(run_inference.py 실행 가이드라인).pdf를 참고하세요.