

CNN을 활용한 의료 영상 분석 모델 개발

20195092_이경준

Kyeong-Jun Lee

이 경 준

Abstract

지금까지도 코로나 19는 끝나지 않았다. 코로나 19 발병 후 3년이 지났지만 지금도 우리는 아직 코로나 19속에 살고 있다. 그리하여 코로나 19로 인한 영향을 최소화하기 위해 인공지능을 학습시켜 흉부 x-ray 데이터 분석을 인공지능 학습의 목표로 삼았다. 345장 중 259장의 train dataset과 86장의 test dataset을 활용하여 인공지능을 학습시켰고 학습시킨 모델로 88퍼센트의 정확도로 일반인과 확진자의 흉부 x-ray를 구분할 수 있게 되었고 이러한 결과들을 통해 의료진의 코로나 19 진단에 유용한 도구로 활용될 수 있기를 기원하며 나아가 더 정확하고 확실한 모델 개발을 통해 의료진과 정부에 이바지하여 코로나 19로부터 대응할 수 있도록 하는 가능성을 보았습니다.

1. 서론

코로나 종식이 가까워지고 있는 지금, 우리가 코로나로 인해 얼마나 힘들고 고생했는지를 상기시키기 위해 코로나 시대의 흉부 X-ray를 주제로 선택하게 되었습니다. 또한, 흉부 X-ray 모델을 인공지능으로 학습시켜 코로나 바이러스로 인한 고통을 최소화하고자 하는 바람과 의료진들의 진단에 도움이 되고자 하는 바람으로 주제를 선정하였습니다.

2. 본론

2.1 데이터 세트

데이터 세트는 총 259장의 훈련 데이터로 구성되었습니다. 이 중에는 149장의 확진자의 흉부 x-ray 이미지와 110장의 일반인의 흉부 x-ray 이미지가 포함되어 있습니다. 또한, 86장의 테스트 데이터 세트도 구축되었으며, 이는 46장의 확진자의 흉부 x-ray 이미지와 40장의 일반인의 흉부 x-ray 이미지로 구성되어 있습니다. 따라서, 총 345장의 흉부 x-ray 데이터가 모델 학습과 테스트에 사용하였습니다.

2.2 수행 환경

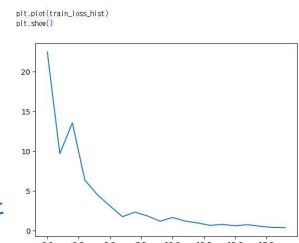
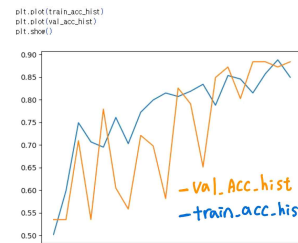
딥러닝과 인공지능 연구를 위한 머신 러닝 프레임워크 모델인 파이토치 언어를 사용했고 인공지능을 학습시키기에 적합한 환경인 코랩에서 인공지능 학습을 수행하였습니다.

2.3 알고리즘

CNN 모델을 사용하여 학습시켰고 정확도를 증가시키기 위해 배치사이즈를 60으로 조정하였습니다. 그리고 손실함수의 값을 최소화하기 위해 반복을 요구하는 에폭의 값을 20으로 변경하였습니다. 그리고 기존 패딩의 값을 1에서 2로 변경하여 데이터의 크기를 조정하였고 더욱 좋은 결과값을 낼 수 있

었습니다.

3. 결과

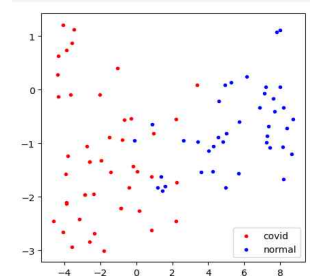
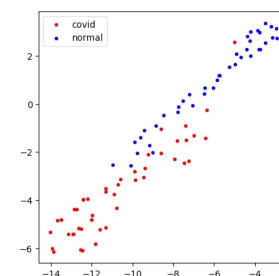


```
train_acc_hist, train_loss_hist = train_model(model,
Epoch 0/19
/usr/local/lib/python3.10/dist-packages/torch/nm/fun
warnings.warn(warn_msg)
Loss: 80.3124 Acc: 0.5019
Epoch 1/19
Loss: 1.5419 Acc: 0.8069
Epoch 2/19
Loss: 29.1084 Acc: 0.5985
Epoch 3/19
Loss: 13.6605 Acc: 0.7066
Epoch 4/19
Loss: 7.4636 Acc: 0.6950
Epoch 5/19
Loss: 5.2519 Acc: 0.7606
Epoch 6/19
Loss: 3.6278 Acc: 0.7027
Epoch 7/19
Loss: 3.0181 Acc: 0.7722
Epoch 8/19
Loss: 2.5192 Acc: 0.7992
Epoch 9/19
Loss: 1.7755 Acc: 0.8147
Epoch 10/19
Loss: 1.5419 Acc: 0.8069
Epoch 11/19
Loss: 1.5283 Acc: 0.8185
Epoch 12/19
Loss: 0.6210 Acc: 0.8340
Epoch 13/19
Loss: 0.6479 Acc: 0.7876
Epoch 14/19
Loss: 0.7716 Acc: 0.8533
Epoch 15/19
Loss: 0.5164 Acc: 0.8456
Epoch 16/19
Loss: 0.5441 Acc: 0.8147
Epoch 17/19
Loss: 0.3851 Acc: 0.8571
Epoch 18/19
Loss: 0.5114 Acc: 0.8880
Epoch 19/19
Loss: 0.4293 Acc: 0.8494
Training complete in 3h 17m
Best Acc: 0.888031
val_acc_hist = eval_model(model, test_loader, device)
saved_model('covid/00.pth', 'covid/01.pth', 'covid/02.pth')
Loading model covid/00.pth
Acc: 0.5549
Loading model covid/01.pth
Acc: 0.5549
Loading model covid/02.pth
Acc: 0.7093
Loading model covid/03.pth
Acc: 0.5549
Loading model covid/04.pth
Acc: 0.7791
Loading model covid/05.pth
Acc: 0.8047
Loading model covid/06.pth
Acc: 0.8581
Loading model covid/07.pth
Acc: 0.7209
Loading model covid/08.pth
Acc: 0.8577
Loading model covid/09.pth
Acc: 0.8023
Loading model covid/10.pth
Acc: 0.8255
Loading model covid/11.pth
Acc: 0.7937
Loading model covid/12.pth
Acc: 0.6612
Loading model covid/13.pth
Acc: 0.8495
Loading model covid/14.pth
Acc: 0.8721
Loading model covid/15.pth
Acc: 0.8837
Loading model covid/16.pth
Acc: 0.8837
Loading model covid/17.pth
Acc: 0.8721
Loading model covid/18.pth
Acc: 0.8721
Loading model covid/19.pth
Acc: 0.8837
Validation complete in 1h 35m
Best Acc: 0.883721
```

train 데이터세트를 학습시킨 결과 최고 0.888031

test 데이터세트를 활용한 eval모델의 결과 최고 0.883721

train_loss_hist(손실함수)의 결과 최저 0.4293



4. 고 찰

주어진 CNN 모델의 train데이터세트의 batchsize는 32, num_epoch는 5, 그리고 padding의 값은 1로 지정되어 있었습니다. 그리하여 손실함수의 값은 10~15의 값이 나왔었고 훈련 데이터세트와 학습데이터세트의 정확도가 60~70퍼센트로 낮게 측정되었습니다. 그리하여 정확도를 높이기 위해 batchsize를 32에서 60으로 변경하고 padding을 1에서 2로 변경하였습니다. 그리고 손실함수의 값을 줄이기 num_epoch의 크기를 5에서 20로 변경하여 더욱 괜찮은 결과가 나왔고 원하는 결과를 도출해낼 수 있었다고 생각합니다. batchsize는 한 번의 학습상황에서 모델에 입력되는 데이터 샘플의 수를 의미하기에 적절한 batchsize를 지정하여 학습을 시킨다면 정확도가 올라가는 사실을 확인할 수 있었습니다. 그리고 num_epoch은 데이터 세트를 전체적으로 몇 번 반복하여 학습하는 것을 의미하기에 적절한 num_epoch을 지정한다면 훈련의 양이 늘어나 성능이 향상되고 과적합을 방지할 수 있고 마지막으로 적절한 패딩의 조정 역시 이미지의 크기를 유지하거나 출력 크기를 조정하는 등의 목적으로 활용되기에 다음과 같은 결과가 나왔다고 생각합니다.

코로나19 발생 후 경과한 3년 동안 확진자 수와 관련된 뉴스를 보면서, 우리 삶에 코로나가 어떤 영향을 미치는지 심층적으로 고민하게 되었습니다. 이 기간에 많은 사람들이 코로나로 인해 건강상의 문제를 겪었으며, 정부 역시 질병관리본부를 질병관리청으로 개편하는 등 열정과 노력을 기울였지만, 코로나 대응은 여전히 어려운 상황이었습니다. 이러한 현실을 바탕으로, 코로나 시대를 함께 지나온 의료진과 우리 모두에게 도움이 될 수 있는 방안을 모색하던 중, 평범한 사람들과 확진자들의 흉부 x-ray 데이터세트를 접하게 되었습니다. 이러한 데이터세트를 인공지능을 통해 학습시킨다면, 코로나 19로 인한 고통을 최소화하고 의료진과 정부에게도 지원이 가능할 것으로 판단했습니다. 이에 따라, 일반인과 확진자의 흉부 x-ray를 분석하여 비교하는 정확도를 88%로 향상시켰고, 앞으로 이 모델을 발전시켜 더욱 정확한 결과를 도출하길 희망합니다.

5. 결 론

이번 인공지능 학습을 통해 코로나 19로 인한 확진자와 일반인의 흉부 x-ray 데이터를 활용한 모델을 개발하였고 총 345장의 흉부 x-ray 데이터 세트를 사용하여 모델을 학습 및 평가하였습니다. 이 중에는 259장의 훈련 데이터 세트와 86장의 test 데이터 세트가 포함되어 있습니다. train 데이터 세트는 확진자의 흉부 x-ray 149장과 일반인의 흉부 x-ray 110장으로 구성되어 있으며, test 데이터 세트에는 확진자의 흉부 x-ray 46장과 일반인의 흉부 x-ray 40장이 포함확진자의 흉부 x-ray의 정확도를 88퍼센트까지 향상시키는 결과를 도출해낼 수 있었습니다. 이 연구를 개선하면 의료진의 업무의 도움을 주고 나아가 정부 차원에 도움을 줄 수 있지 않을까 하는 기대감과 다른 의료분야에서의 활용도도 넓힐 수 있지 않을까 하는 기대감이 생겼습니다.

참고문헌

[1] 어떤 문서, 논문, 블로그 등을 보고 배경, 목적, 결과 등을 작성했는지 쓰시면 됩니다.

참고문헌

캐글(Covid-19 Dataset)

과학기술정책연구원(인공지능 기술 전망과 혁신정책 방향)

대한경영정보학회「경영과 정보 연구」

- 합성곱 신경망을 활용한 지능형 아토피피부염 중증도 진단 모델 개발

배제대학교 대학원(컴퓨터 공학과)

- 딥러닝을 이용한 안면 여드름 분류 모델 연구

서울대학교 컴퓨터공학부 바이오지능 연구실

- 딥러닝 : 인공지능을 이끄는 첨단 기술