의료 데이터의 자기지도학습 적용을 위한 pretext task 분석

공희산[1] · 박재훈 · 김광수

성균관대학교 소프트웨어학과

Pretext task analysis for self-supervised learning application of medical data

Heesan Kong · Jaehun Park · Kwangsu Kim

SungKyunKwan University College of Software

E-mail : khtks@g.skku.edu / zlrnwlzladl@gmail.com / kim.kwangsu@skku.edu

요 약

의료 데이터 분야는 레코드 수는 많지만 응답값이 없기 때문에 인공지능을 적극적으로 활용하지 못하고 있다. 이러한 문제점을 해결하기 위해 자기지도학습(Self-Supervised learning)을 의료 분야에 적용하는 연구가 등장하고 있다. 자기지도학습은 model이 레이블링이 없는 데이터의 semantic 표현을 이해할 수 있도록 pretext task와 supervision을 학습한다. 그러나, 자기지도학습의 성능은 pretext task로 학습한 표현에 의존하므로 데이터의 특성에 적합한 pretext task를 정의할 필요가 있다. 따라서 본 논문에서는 의학 데이터 중 활용도가 높은 x-ray 이미지에 적용할 수 있는 pretext task를 실험적으로 탐색하고 그 결과를 분석한다.

ABSTRACT

Medical domain has a massive number of data records without the response value. Self-supervised learning is a suitable method for medical data since it learns pretext-task and supervision, which the model can understand the semantic representation of data without response values. However, since self-supervised learning performance depends on the expression learned by the pretext-task, it is necessary to define an appropriate Pretext-task with data feature consideration. In this paper, to actively exploit the unlabeled medical data into artificial intelligence research, experimentally find pretext-tasks that suitable for the medical data and analyze the result. We use the x-ray image dataset which is effectively utilizable for the medical domain.

키워드

Self-Supervised learning, Pretext-task, Supervision, presentation learning, Medical Deep-learning

Ⅰ. 서 론

현재 인공지능 분야의 지배적인 방법론인 지도학습은 다양한 분야에서 높은 성능을 보이지만, 의료 분야에서는 데이터가 많음에도 불구하고 레이블링이 되어있지 않은 레코드가 많아 인공지능 활용이 어렵다. 특히, 의학 데이터의 레이블링을 위해서는 의학 분야에 전문성을 가진 인력의 시간과 노력이 필요하므로 비용이 많이 발생한다.

이러한 한계점을 극복하기 위해 레이블이 없는 데이터를 사용해 학습하는 비지도학습 방법론 중 하나인 자기지도학습(Self-Supervised learning)을 의학 분야에 적용하는 연구가 등장하고 있다. 자기지도학습이란, 간단한 문제(pretext task)와 그 문제에 맞는 정답(supervision)을 새롭게 정의해 레이블이 없는 데이터의 표현을 학습함으로써 인공지

[1] speaker

능을 이용해 실제로 풀고 싶은 문제(downstream task)의 성능을 높이는 방법이다.

자기지도학습의 성능은 pretext task를 통해 학습한 데이터의 표현에 의존적이므로, 데이터의 특징에 맞는 pretext task를 정의할 필요가 있다. 그러나, 의료 데이터에 적합한 pretext task를 정의하는 연구는 아직 진행되지 않아 자기지도학습을 적용하기 어려운 상황이다. 따라서 의료 데이터에 적합한 pretext task를 찾는다면, 이를 인공지능에 적극 활용할 수 있다.

의료 데이터 중 흉부 X-ray는 정밀검사 전 환자의 병의 유무를 판단하는 가장 기초적인 검사에 사용되고, 오랜 기간 누적되어 활용가치가 높으므로 본 연구에서는 'NIH Chest X-ray' 데이터[1]를 이용해 흉부 X-ray 데이터에 자기지도학습을 적용하기 위해 알맞은 pretext task를 실험적으로 탐색했다.

II. 자기지도학습

2.1 Self Supervised Learning

데이터와 정답(label)을 함께 이용해 문제를 푸는 방법을 학습하는 지도학습은 다양한 task에서 뛰어난 성능을 보이지만 모든 데이터에 레이블링이 필요하다는 근본적인 문제점이 있다. 반면, 자기지도학습은 레이블이 없는 데이터로부터 유용한 표현을 학습하는 데 목적이 있다. 학습한 데이터의 표현은 다른 task에 손쉽게 적용이 가능하다. 즉, 자기지도학습은 직접 문제를 풀기보다 데이터의 표현을 학습하고, 학습한 표현을 다른 task에 전이시켜 성능을 높이는 방법이다.

자기지도학습에서 데이터의 표현을 학습하기 위해 구성한 문제를 pretext task라고 지칭하고, 실제로 풀고 싶은 문제를 downstream task라고 한다.

2.2 Pretext task

모델이 데이터의 표현을 학습할 수 있도록 새롭게 정의한 문제인 pretext task는 자기지도학습의 목적에 맞게 일반적인 goal을 갖도록 구성한다. 이에 더해, pretext task에 맞는 답 또한 함께 정해 사용함으로써 레이블이 없는 다량의 데이터도 학습할 수 있다.

2.3 Downstream task

이미지 분류(Image classification), 객체 탐지(Object detection)와 같이 레이블이 없는 데이터로 학습한 표현을 전이시켜 실제로 풀고 싶은 target task를 downstream task라고 한다.

전이시킨 downstream task는 자기지도학습으로 학습한 표현의 유용성을 평가하는 지표 중 하나로 사용한다.

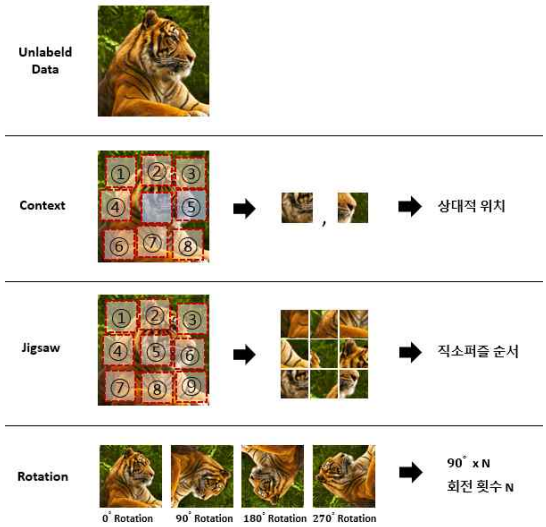

그림 1. Context, Jigsaw, Rotation pretext task 모식도

III. 실험 및 결과

3.1 실험 설계

3.1.1 Dataset 'NIH chest X-ray' 데이터를 기반으로 실험에 다양성을 주기 위해 데이터를 256x256과 128x128 크기로 변환해 자기지도학습을 적용했다. 학습/검증/평가를 위한 데이터는 'NIH chest X-ray'에서 제공하는 목록에 맞게 나누어 사용했다.

3.1.2 Pretext task learning 본 연구에서 그림 1.과 같이 Context, Rotation, Jigsaw 3가지 pretext task를 이용해 자기지도학습을 의학 분야에 적용할 때 알맞은 pretext task를 탐색했다.

Context(relative patch location) [2]는 이미지를 9등분 후 중앙과 나머지 임의의 부분에서 샘플링한 두 patch[1]의 상대적인 위치를 맞추는 pretext task이다. 이미지 내부의 context를 활용해 두 patch 사이의 유사성을 학습한다.

Jigsaw [3]는 그림 1.에서 볼 수 있듯, 9등분 한 이미지의 patch들을 섞어 만든 직소퍼즐의 순서를 맞추는 pretext task이다. 9개의 patch로 만들 수 있는 순서의 순열은 수가 많고 비슷한 경우가 많아 학습에 제약이 있으므로, 길이가 같은 두 문자열의 거리를 측정하는 척도인 'Hamming Distance'를 기준으로 100개의 순열을 뽑아 사용하였다. 모델은 직소 퍼즐을 푸는 과정을 통해 의미상 관련 있는 내용을 찾도록 학습한다.

Rotation [4]은 90°단위로 회전시킨 이미지의 회전 횟수를 맞추도록 pretext task를 설계하였다. 이

1) 원본 이미지를 특정 크기만큼 샘플링한 이미지

미지의 회전 정도를 맞추기 위해 모델은 이미지 내 객체의 위치나, 방향 등의 묘사를 인식하도록 학습한다.

데이터의 표현을 학습하기 위한 pretext task는 ImageNet pretraining 된 VGG16을 모델로 사용했고, 초기 learning rate 0.001과 learning rate decay 0.9, momentum 0.9를 사용한 SGD를 사용해 100 epoch 동안 학습했다.

3.1.3 Linear evaluation 자기지도학습의 성능 평가 및 비교를 위해 'Chest X-ray' 이미지의 이상 유무를 이진 분류하도록 downstream task를 구성하였다. Linear evaluation protocol을 따라 자기지도학습을 통해 학습된 모델의 가중치를 고정한 후, 2개의 dense layer를 추가해 10 epoch 동안 이진 분류 task에 맞게 추가로 학습하였다.

본 연구에서는 pretext task로 사용한 Context, Jigsaw, Rotation 모델과, ImageNet pretraining 모델, Random initialization 모델 총 5가지를 비교했다.

3.2 실험 결과 및 분석

표 1은 데이터를 128x128과 256x256 크기로 변환해 자기지도학습을 위한 3가지 pretext task와 학습한 표현을 평가하기 위한 linear evaluation 결과이다. ImageNet Pretrain은 지도학습으로 학습된 모델을 의미한다.

표 1. X-ray 데이터 실험 결과

128x128 / 256x256		Linear evaluation	Baseline 대비 성능
Baseline model	ImageNet Pretrain	68.84	-
		71.98	-
Pretext model	Context	64.03	93.01
		66.31	92.12
	Jigsaw	65.91	95.74
		65.37	90.82
	Rotation	**67.09**	**97.46**
		67.84	**94.25**

실험 결과 레이블이 없는 데이터를 이용한 자기지도학습 모델이 지도학습 모델과 비견될만한 결과를 보이며, 모든 데이터에 레이블링이 필요하다는 지도학습의 한계를 극복했다. 이러한 결과를 보았을 때, 레이블이 없는 의료 데이터에 인공지능 기술을 적용할 수 있다.

Chest X-ray 데이터는 이미지 내 객체의 방향성과 특징이 명확하고, 데이터 자체의 변화가 크지 않기 때문에 국소적인 부분을 보고 학습하는 Context, Jigsaw pretext task 보다 전체적인 부분을 보고 학습하는 Rotation pretext task가 더 유용한 표현을 학습했다고 할 수 있다.

데이터의 크기도 결과에 영향을 미쳤는데, 표현의 학습에는 차이가 없지만, linear evaluation 결과 데이터의 크기가 큰 경우 downstream task로 표현을 전이한 결과가 더 좋다는 것을 확인했다.

Ⅳ. 결 론

본 연구에서는 'NIH Chest X-ray' 데이터를 이용해 흉부 X-ray 데이터에 알맞은 pretext task를 탐색했다. 실험 결과, 이미지의 지엽적인 부분만을 보고 학습한 pretext task 보다 전체적인 부분을 이용해 학습한 pretext task가 적합하다는 것을 알 수 있었다.

향후 연구에서는 파라미터를 최적화하거나 contrastive learning을 적용해 자기지도학습의 성능을 높일 수 있을 것으로 기대한다. 또한, 흉부 X-ray 데이터 이외의 다양한 의료 데이터에도 확장 가능하다.

Acknowledgment

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No.2020-0-00990, 5G-IoT 환경에서 이기종ㆍ비정형ㆍ대용량 데이터의 고신뢰ㆍ저지연 처리를 위한 플랫폼 개발 및 실증)

References

[1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri and RM. Summers, "ChestX-ray8: Hospital-scale Chest X-ray Database and Benchmarks on Weakly-Supervised Classification and Localization of Common Thorax Diseases" in *Proceedings of the IEEE conference on computer vision and pattern recognition,* Hawaii Convertion Center, pp. 2097-2106, 2017.

[2] C. Doersch, A. Gupta, and A. A Efros, "Unsupervised Visual Representation Learning by Context Prediction" in *International Conference on Computer Vision,* Santiago, pp. 1422-1430, 2015.

[3] M. Noroozi and P. Favaro, "Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles" in *European Conference on Computer Vision,* Amsterdam, pp. 1422-1430, 2016

[4] S. Gidaris, P. Singh and N. Komodakis, "Unsupervised Representation Learning By Predicting Image Rotations" in *International Conference on Learning Representations*, Vancouver, 2018