Automatic Labeling of Chest Radiography Images

Fateme Sadat Haghpanah

11 April 2019

1 Problem Statement

Radiography using X ray is one of the most common imaging techniques. It can be used to image lots of different parts of the body, including hands, arms, foot, and chest. Chest radiography is the most common imaging examination globally, critical for screening, diagnosis, and management of many life threatening diseases [1]. Researchers from Stanford university provided a large dataset of chest radiography images, named CheXpert [1]. In this project, I am going to use CheXpert in order to train a model that can automatically label the chest radiography images.

2 Dataset

The CheXpert dataset consists of 224316 chest radiographs of 65240 patients. The authors of [1] extracted the labels for each image using the radiology reports. Also, they provided a set of 200 radiography images that annotated manually by 3 certified radiologists, for evaluation purposes. They also provide a challenge on a hidden test set of 500 images that you can submit your code there and see your results.
In the dataset, each image might have more label. Actually, there is a set of 13 observations that each image can have some of them. If no observation found in the image, image labeled as "No Finding" which result in having 14 different labels overall. Also, each label has a degree of certainty assigned to it. In the table 1 you can see the statistics of the labels.

3 Models

Since my dataset is consisted of images, I am going to use CNNs to solve the problem. I am going to first use popular CNN architectures like ResNet (different sizes), Xception, and etc. Then, I will try to figure out if it would be possible to find a better network.

Pathology	Positive (%)	Uncertain (%)	Negative (%)
No Finding	16627 (8.86)	0 (0.0)	171014 (91.14)
Enlarged Cardiom.	9020 (4.81)	10148 (5.41)	168473 (89.78)
Cardiomegaly	23002 (12.26)	6597 (3.52)	158042 (84.23)
Lung Lesion	6856 (3.65)	1071 (0.57)	179714 (95.78)
Lung Opacity	92669 (49.39)	4341 (2.31)	90631 (48.3)
Edema	48905 (26.06)	11571 (6.17)	127165 (67.77)
Consolidation	12730 (6.78)	23976 (12.78)	150935 (80.44)
Pneumonia	4576 (2.44)	15658 (8.34)	167407 (89.22)
Atelectasis	29333 (15.63)	29377 (15.66)	128931 (68.71)
Pneumothorax	17313 (9.23)	2663 (1.42)	167665 (89.35)
Pleural Effusion	75696 (40.34)	9419 (5.02)	102526 (54.64)
Pleural Other	2441 (1.3)	1771 (0.94)	183429 (97.76)
Fracture	7270 (3.87)	484 (0.26)	179887 (95.87)
Support Devices	105831 (56.4)	898 (0.48)	80912 (43.12)

Figure 1: Statistics of labels in the dataset

4 Metrics

The most obvious metric that I am going to check would be accuracy. I will also calculate and consider precision, recall, F1, and AUC measures. Also, I am going to plot the confusion matrix to find the places that my model cannot work well.

5 Responsibilities

I have no group-mate, so this section is not needed! :D

References

[1] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, et al. Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In *Thirty-Third AAAI Conference on Artificial Intelligence*, 2019.