Machine Learning Engineer

Capstone Project Proposal

Xun Ding

June 28th, 2018

Overview

This project explores existing pre-trained *CNN* networks and uses transfer learning to distinguish normal chest X-ray images from images that indicate pneumonia. Furthermore it classifies those with pneumonia as either viral or bacterial. Dataset for this project can be downloaded from kaggle.

Domain Background

Neural networks is at the heart of many state of art of machine learning applications. Aside from self-driving cars that are on the cusp of entering mainstream, Google recently showcased a new technology *Duplex* built on top of *RNN* (recurrent neural networks) and *ASR* (automatic speech recognition) technology that can conducting natural conversations to perform everyday tasks in a closed business domain[1]; In computer vision, *CNN* (convoluted neural networks) is applied especially in image classification tasks. *CNN* are used to teach computer agent to play video games, build comprehensive and accurate street maps by reading street signs.

Ever advancing deep learning technology makes revolution possible in many other areas, such as healthcare. Since 2008, adoption of EHR (electronic health records) has increased ninefold[2] , health care data has not only been increasing in volumes, it has also been growing in diversity (behavioral data, social environment, genomics, baseline health exams, vital signs, lab exams). Healthcare data is making enormous progress in standardization by using International Classifications of Diseases (ICD).

While the healthcare sector is being transformed by the availability of data and capacity of recording data, the enormous volume and complexity of data becomes impossible for any human to analyze. Machine deep learning provides a clear path for data classification, clustering and regression. More and more real life deep learning applications have

[1] Google Duplex: An AI System for Accomplishing Real-World Tasks Over the Phone
[2] Adoption of Electronic Health Record Systems among U.S. NonFederal Acute Care Hospitals: 2008-2014

demonstrated machine learning is able to diagnose disease at a accuracy that is on par or exceed medical experts. For example, deep learning based tissue analysis is able to predict outcome of colorectal cancer[3], and CNN image classification can provide dermatologist-level classification of skin cancer[4]. Deep learning is becoming more and more of a key in future healthcare applications.

Personal motivation

I currently work at Maestro Health, a company that provides an integrated platform of health benefit administration, including care coordination solutions for employees. MaestroHealth aims to provide a comprehensive and long-term population health management solution for better care at lower cost. I think a deep learning application in chest X-ray image classification would be interesting and would illuminate the power of neural networks and can be used in further applications such as patient hospital readmission identification and X-ray musculoskeletal image readings.

Problem Statement

Pneumonia is an infection that causes inflammation in one or both of the lungs and may be caused by a virus, bacteria, fungi or other germs. Viral pneumonia oftens needs no medication, however bacterial pneumonia needs antibiotics and sometimes hospitalization. Doctors often conduct physical exams and order chest X-rays of the lungs and other types of tests for further evaluation[5]. Lung abnormalities on chest X-rays will either present as areas of increased or decreased density[6].

The goal of this project is to create a CNN model to correctly identify images of pneumonia from normal images; the model will also classify the cause of pneumonia as either virus or bacteria. Furthermore, the project will create a website to make real-time predictions on any given chest X-ray images.

Database and inputs

The X-ray dataset consists of a total of 5856 chest X-ray images (JPEG) and 2 categories (Pneumonia/normal). The kaggle dataset is pre-organized into 3 folders (train, test, val)[7].

Available labeled images are distributed as the following:

---

[3] Deep learning based tissue analysis predicts outcome in colorectal cancer
[4] Dermatologist-level classification of skin cancer with deep neural networks
[5] Pneumonia
[6] Chest X-ray - lung disease
[7] Chest X-Ray Images (Pneumonia)

	Normal	Pneumonia		**Total**
		Virus	Bacteria	
Train	1349	1345	2538	**5232**
Test	234	148	242	**624**
Total	**1583**	**1493**	**2780**	**5856**

I would say the dataset is, though skewed towards the abnormal end, well balanced in the normal and abnormal (viral vs bacterial) spectrum.

Sample labeled images are included in the project folder: *chest-xray-sample-images*. The full dataset can be downloaded from [kaggle](). The images are in grayscale (black for dense structures, white for metal and contrast media, muscle, fat, and fluid will appear as shades of gray) and have varied dimensions.

Benchmark Model

Baseline Model
I have built a baseline model that uses extracted features from Keras' VGG-16 model and a *GlobalAveragePooling2D* pooling and fully-connected top layer. I am able to get the following result (obtained from test set prediction):

	Accuracy	Sensitivity	Specificity
Normal vs Pneumonia	76.1%	81.3%	52.2%
Bacterial vs Others	82.2%	78.4%	65.5%

Code for the baseline model and metrics calculations are included with the project repository.

Secondary Model
The team at University of California has built a model that trained on both Optical Coherence Tomography (OCT) and Chest X-Ray Images and published their results in this article [Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning](). The results for chest X-ray is the following:

	Accuracy	Sensitivity	Specificity
Normal vs Pneumonia	92.8%	93.2%	90.1%

(Note: the team also has accuracy, sensitivity and specificity score for **Binary** comparison of bacterial vs viral pneumonia. However, it is not clear to me how they derive the scores. Other than training the images with pneumonia separately, I am not sure how to reach a binary comparison of only bacterial and viral pneumonia).

Solution Statement

I plan to use transfer learning and experiment with the following pre-trained models: Resnet-50, Inception and Xception model. I will choose the best performing architecture for my final model.

For this project, my goal is to reach higher than 85% accuracy, sensitivity and specificity.

Evaluation Metrics

I will use the following metrics to evaluate my CNN architecture:

Normal vs Abnormal (Healthy vs Pneumonia)	Accuracy	How often the model correctly label the x-ray result normal /healthy vs abnormal (pneumonia) ?
	Sensitivity	Of all the images with pneumonia, how many are correctly identified?
	Specificity	Of all the normal images, how many are correctly identified?
Bacterial vs Others (Urgent care vs Non-urgent care)	Accuracy	How often the model correctly label the x-ray result as bacterial vs viral?
	Sensitivity	Of all the images with viral pneumonia, how many are correctly identified?
	Specificity	Of all the images with bacterial pneumonia, how many are correctly identified?

Confusion Matrix: how well the model is able to classify each of the three categories, normal, viral and bacterial?

Project Design[8]

At this stage, everything is exploratory.

[8] Preliminary code is submitted together with the proposal. Please see vgg_16.py

As the skin cancer classification study has shown that, even for images that are drastically different as those of skin cancer and dog cat images, there are shared commonalities among them, especially during the initial learning phase; it is based on this insight, the team of Sebastian Thrun is able to use a “GoogleNet Inception v3 CNN architecture that was pre-trained on approximately 1.28 million images” and achieve great result[9].

For this project, I plan to try with pre-trained Resnet-50 and Inception and Xception model. For each of the three models, I will take the following steps:

Pre-process the data:

1. Reorganize the images in the given train, validation and test directories, so they are in three categories: normal, pneumonia_virus, pneumonia_bateria, in doing so I can easily use Keras prebuilt the *ImageDataGenerator* and *flow_from_directory* functionality of Keras.
2. Re-scale the images by dividing every pixel in every image by 255

Feature extraction and model training:

1. Use the pre-trained model from Keras, extract and save bottle neck features with the top layer off.
2. Train the network using the saved bottleneck features to classify normal vs pneumonia, viral vs bacterial. Use checkpoint to save the best model weights. Plot the history of accuracy and losses by epochs.
3. Fine tune the various parameters (learning rate), activation functions (relu, vs. elu vs. sigmoid) and layers (dropout layers, batch normalization, max pooling, etc), batch size, image dimensions.
4. Retrain and continue to fine tune the network until satisfied
5. Load the model weights and make predictions on the test set. Note the performance by checking the accuracy, sensitivity and specificity

Further steps to take, if needed:

1. Randomize and remix the validation, test, and train data set;
2. Create copies of training samples;
3. Create more training data through image augmentation;
4. Collect more data from the internet;

[9] Dermatologist-level classification of skin cancer with deep neural networks