NLP for Healthcare Triage

Project Report

Prasad Valavade

Abstract

This project is about building Healthcare Triage classifier based on clinical notes maintained by the hospitals. It compares mainly 2 approaches

1. The common BERT (DistilBERT) based classification technique
2. Using domain specific (clinicalBERT) based classification technique

Introduction

'NLP for Healthcare Triage' aims to enhance the current Healthcare Triage system.
Language is a complex social process especially when it comes to the Healthcare Industry. As interpreting and analysing statements will be a deciding factor of further course of action for a patient's Life Saving decisions.

Healthcare Triage is a system of clinical risk management employed in Emergency Departments worldwide to manage patient flow safely when clinical need exceeds capacity. Systems are intended to ensure care is defined according to the patient's need and in a timely manner.

Background

Medical staff like nurses play a crucial role in the Triage system. This staff understands the patient's situation by having a conversation with him/her and then determining the course of next actions.

Current Challenges

1. The medical staff taking decisions do not have sufficient information for all types of emergency cases.
2. The availability of medical staff taking Triage decisions.
3. The knowledge/learnings from historical cases, which is key for new cases, are difficult to maintain.
4. All these challenges pose the risk to the patient's safety.

Methods

Data

- Getting clinical notes data was a big challenge because of medical data privacy related policies and guidelines.
- Fortunately, MTSamples: Transcribed Medical Transcription Sample Reports and Examples provides a big collection of Transcribed Medical Transcription Sample Reports and Examples.
- A dataset of Consult - History and Phy. t from this site was used for this project. It is having total 516 records
- The data was not labelled from Healthcare Triage perspective
- To label this data, I took help from one of the senior Doctors who is working at the department of Emergency Medicine, Milton Keynes NHS University Hospital, UK.
- Each record was classified into the following categories based on the clinical notes

Category Code	Category Name	Category Description
RED	Immediate resuscitation	Patient in need of immediate treatment for preservation of life
Orange	Very urgent	Seriously ill or injured patient whose lives are not in immediate danger
Yellow	Urgent	Patients with Serious problem but apparently stable condition
Green	Standard	Standard cases without immediate danger or distress
Blue	Non Urgent	Patients whose conditions are not true accidents or emergencies

Solution Design

There were 2 approaches tried to design and build HealthcareTriage system

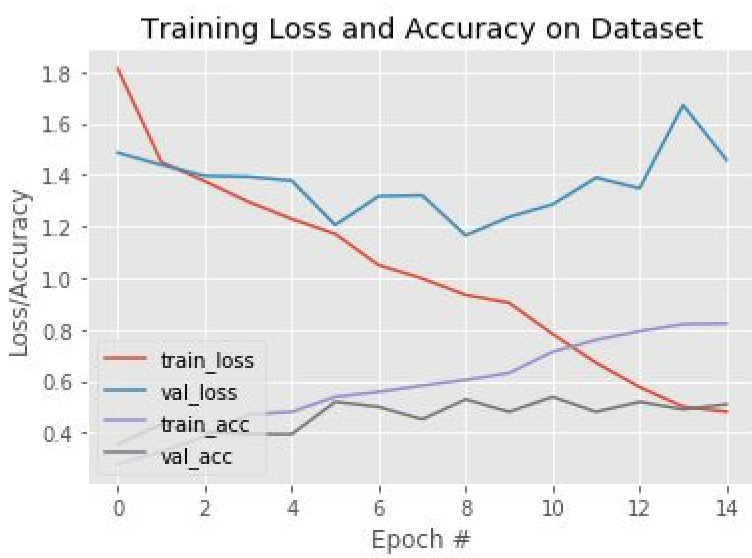
1. In the first approach, the DistilBERT model was used as a base model and Triage classifier was trained using transfer learning mechanism

2. In the second approach, the ClinicalBERT model was used as the base model instead of DistilBERT.
3. The overall solution design was as follows
    a. The data was cleaned to remove common but non clinical words and other letters/numbers/characters
    b. The tokenizer was used to token clinical notes. It was observed that tokenizer was representing clinical notes in a more efficient way when all sentences were sent to it in one go. The maximum number of tokens was set to 512. It was observed that every record was having words between 800 to 10000. However, in most of the cases, all words beyond 500 were not related to the current condition of the patient.
    c. The data was then divided into training and test/validation datasets using 80:20 ratio.
    d. For approach 1, 15-20 epochs were used. The learning rate was changed between 0.00001 to 0.001.
    e. For approach 2, 10 epochs were used. The learning rate was set to 0.000023

Results and Discussion

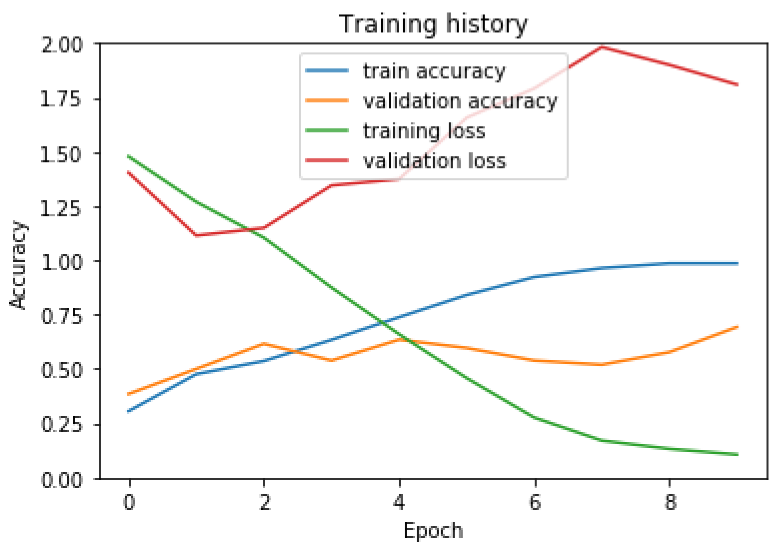
Approach 1 Results (using DistilBERT):

The following figure shows training/validation loss and accuracy graphs for the model built using DistilBERT. The best training accuracy achieved was 82% while best validation accuracy achieved was 53.85%. When it comes to loss, it is evident that training loss decreased significantly over all epochs. However, validation loss did not decrease much.



Approach 2 Results (using ClinicalBERT):

The following figure shows training/validation loss and accuracy graphs for the model built using ClinicalBERT. In this approach, the best training accuracy achieved was 98% while best validation accuracy achieved was 69% The training loss decreased exponentially. However validation loss increased after decreasing in the first few epochs and started decreasing again in the last few epochs.



The following table gives a classification report for the test data (based on final trained model). When it comes to precision 'RED category has done really well and ;Green' category is having the best recall. For the 'Yellow' category, both precision and recall scores are very poor. The overall accuracy for the test data is 62 %

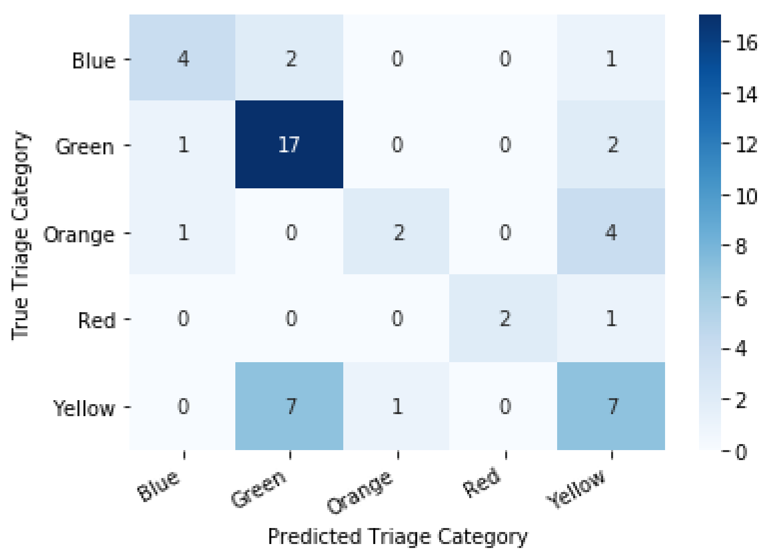
Classification Report for ClinicalBERT based Model:

	precision	**recall**	**f1-score**	**support**
Blue	0.67	0.57	0.62	7
Green	0.65	0.85	0.74	20
Orange	0.67	0.29	0.40	7
Red	1.00	0.67	0.80	3
Yellow	0.47	0.47	0.47	15
accuracy			**0.62**	52
macro avg	0.69	0.57	0.60	52

weighted avg	0.62	0.62	0.60	52

Confusion Matrix for ClinicalBERT based model:

From the confusion matrix it is evident that most of the Orange records were categorized as Yellow. Hence it is having very poor recall and it translated into poor precision for the Yellow category . Also most of the Yellow records were categorized as Green. It translated into poor recall for Yellow. However Yellow and Green are adjacent categories and the same is the case with Yellow and Orange.



Conclusion

1. It is quite evident that the ClincalBERT based model has performed significantly better than DistilBERt based model. It reiterates the fact that BERT based models do really well if they are pre-trained for a given specific domain
2. While training accuracy is very good for approach 2, the test and validation accuracies are fairly ok. It may be attributed to the less number training as well as test/validation records. Building a new BERT based pre-trained model only for triage application will give much better results.
3. When precision/recall scores are poor, the model has classified the corresponding record into adjacent categories. This may be acceptable to certain extent as the model has not confused between extreme categories like Red and Blue. This is really important from the practical application of this classifier for Healthcare Triage..

References

1. Kevin Mackway-Jones, Janet Marsden,Jill Windle(2013). Emergency Triage Manchester Triage Group.
2. Sentiment Analysis with BERT and Transformers by Hugging Face using PyTorch and Python
3. Akash Desarda, Working with Hugging Face Transformers and TF 2.0 | by Akash Desarda