Movie Sentiment Analysis using Deep Learning for NLP

• What problem did you select and why did you select it?

In this project, we will be performing sentiment analysis, which is a type of text classification using Keras deep learning library.

Word embeddings are a technique for representing text where different words with similar meaning have a similar real-valued vector representation. They are a key breakthrough that has led to great performance of neural network models on a suite of challenging natural language processing problems.

• What database/dataset will you use? Is it large enough to train a deep network?

The IMDB (International Movie Database) trained dataset has 25000 movie reviews for natural language processing and two columns - review and sentiment. The review column contains text reviews and the sentiment column has the corresponding sentiments with 12500 positive and 12500 negative movie reviews. The test dataset has 25000 reviews. There are additional 50,000 IMDB reviews provided without any rating labels.

Thus, this data set may be used to predict positive and negative reviews using either classification or deep learning algorithms.

• What deep network will you use? Will it be a standard form of the network, or will you have to customize it?

We will use convolutional neural networks (CNN) to classify the text. CNN tries to find specific features in an image in the first layer. In the next layers, the initially detected features are joined together to form bigger features and in this way, the whole image is detected.

- **What framework will you use to implement the network? Why?**

We would develop a deep learning model for neutral networks to automatically classify movie reviews as positive or negative with Keras.

A great way to use deep learning to classify images is to build a convolutional neural network (CNN), and the Keras in python makes it pretty simple to build a CNN.

- **What reference materials will you use to obtain sufficient background on applying the chosen network to the specific problem that you selected?**

https://blog.keras.io/using-pre-trained-word-embeddings-in-a-keras-model.html

https://keras.io/api/layers/core_layers/embedding/

- **How will you judge the performance of the network? What metrics will you use?**

We will use the ‘accuracy’ metric from the compiled model of Keras. We will also use the standardized Cohen Kappa test to check our model’s performance.