ASSIGNMENT-4

TEXT AND SEQUENCE

We are training the RNN model on the IMDB data set to find the reviews and finally make a binary prediction to identify the positive and negative reviews. Which is a method of the neural network model designed to handle the sequential data that process in one direction, RNN has a recurrent connection that helps to process data sequentially.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Review words | Max length | Training size | Validation size | Embedding accuracy |
| 150 | 10000 | 100 | 10000 | 75.23 |
| 150 | 10000 | 1000 | 10000 | 85.81 |
| 150 | 10000 | 10000 | 10000 | 87.32 |
| 150 | 20000 | 15000 | 10000 | 81.24 |
| 200 | 20000 | 20000 | 10000 | 86.16 |

Different samples:

In the above samples, the model trained on a small amount of data, and the network architecture used to run this model was very simple just with an embedding layer. Accuracy was not having many variations on the data set. But the best accuracy for the model at 150 top words was 87.32

The sample at 200 samples reaches the maximum accuracy and upon changing it was similar with slight variation. changing hyper parameters will affect the performance of the model. Which are learning rate, dense layers, and drop-out rate.

**Accuracy of the pre-trained model**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Review words | Max length | Training size | Validation size | Pre-trained accuracy |
| 150 | 10000 | 100 | 10000 | 80.1 |
| 150 | 10000 | 1000 | 10000 | 81.27 |
| 150 | 10000 | 20000 | 10000 | 80.1 |

The vector dimensions in the pre-trained network model and layers. The model's performance seems to be poor on unseen data because the model training is unable to identify new data, affecting the greater performance on the test set.

**Conclusion**

The final model was across all models the model with embedding layers’ performance was high as the sample size increases the accuracy was increased to accuracy was reached to 88.16 and for the pre-trained model was 80.1. As we can say by the dense layers and learning late and convolution layers affecting the performance. From all sample sets in embedding layer with top reviews of 150 samples the performance of model was reached to 8.32