

Report on Text and Sequences

Kiran Kour

Objective: Apply RNNs to text and sequence data demonstrating how to improve the network's performance, especially when dealing with limited data and determining which approaches are more suitable for prediction improvement.

Models	Training	Validation	Test	Loss Test	Accuracy Test
Model 2	100	10000	5000	0.6929	0.5058
Model 3	1000	10000	5000	0.5401	0.7314
Model 4	900	10000	5000	0.6265	0.6908
Model 5	4000	10000	5000	0.4667	0.7744
Model 6	30000	10000	5000	0.3277	0.896
Pre Model 1	100	10000	5000	0.6711	0.6376
Pre Model 2	20000	10000	5000	0.3024	0.87

Findings:

- Model 1 was just a base model using the entire dataset to know how a model with embedding layers works and its accuracy on the test set. I have not considered it in the evaluation of the model performance.
- The initial models- Model 2 and Model 4 was built using a very small set of training samples, and as a result, they could have performed better. Model 4 was built using embedding layers along with Conv1D and dense layers and a dropout value of 0.5.
- On the other hand, in Model 3 and Model 5, when built by increasing the training samples along with Conv1D, dense layers, and the dropout value of 0.8, the accuracy of the models increased, and the loss value decreased. Model 3 and Model 4 were built using Conv1D, embedding, and dense layers, but the training sample sizes differed.
- Model 6 has the highest accuracy and the most negligible loss value. It was built using the maximum training sample size of 30000, resulting in maximum accuracy.
- GloVe network was used to build the Pre model 1 and Pre model 2, which are pre-trained models. One major lesson while building Pre model 1 with just 100 training

samples learned is underfitting. Because this model was trained using 100 training samples, it could not perform better. A clear increase in the accuracy rate can be seen when the training samples were increased from 100 to 20000 in the Pre model 2. This concludes that providing more data to the model increases its accuracy because it learns more and more about the data.

