“WORD EMBEDDINGS USING EMBEDDED LAYERS AND PRE-TRAINED MODELS”

This report aims to investigate the effects of different hyperparameters on the performance of a sentiment analysis model using the IMDB dataset. The methodology involved training the model with 10,000 training samples and limiting the movie reviews to the top 10,000 common words, with a maximum review length of 150 words. The input shape of the embedding layer was set to (1000,64). By modifying one hyperparameter at a time, several experiments were conducted to analyze the impact on the model's performance.

**The values are set to:**

Cutoff reviews set to 150 words  
 training samples = 100  
 Validate samples = 10,000   
 words= 10,000

**A) EMBEDDING LAYERS trained from scratches**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cutoff Review | No. of words | Training samples | Validation accuracy | Test accuracy |
| 150 | 10000 | 100 | 68.87 | 70.7 |
| 150 | 10000 | 500 | 80.7 | 84.4 |
| 150 | 10000 | 1000 | 83.3 | 84.7 |
| 200 | 10000 | 2000 | 82.21 | 84.5 |
| 150 | 10000 | 10000 | 83.43 | 84.0 |

Cutoff review of 150 with 1000 samples shows more accuracy in scratch model as 84.7

**B) Embedding layer with masking enabled**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Cut off Review | No. of words | Training samples | Validation accuracy | Test accuracy |
| 150 | 10000 | 100 | 71.65 | 70.2 |
| 150 | 10000 | 500 | 80.62 | 82.74 |
| 150 | 10000 | 1000 | 83.2 | 83.6 |
| 200 | 10000 | 2000 | 82.21 | 84.9 |
| 150 | 10000 | 10000 | 82.9 | 82.6 |

Using an embedding layer with masking helps the network learn meaningful representations of input sequences with varying lengths.

Here, we can observe the accuracy has improved.

**C) Model that used pre-trained embedding layer**

|  |  |  |  |
| --- | --- | --- | --- |
| Max length | Training samples | Validation accuracy | Test accuracy |
| 150 | 100 | 61.46 | 67.1 |
| 150 | 500 | 83.36 | 82.7 |
| 150 | 1000 | 84.06 | 82.6 |
| 200 | 2000 | 83.86 | 83.7 |
| 150 | 10000 | 82.73 | 82.8 |

Generally pretrained model gives better performance but here, the accuracy has decreased but we can observe that accuracy is improved because of changing hyperparameters also like cut off and training samples .

**CONCLUSION:**

* For the **embedding layer model**, the validation accuracy is higher than the test accuracy for all cutoff reviews and training sample sizes. This suggests that the model might be overfitting to the training data.
* For the **pre-trained model,** the validation accuracy is higher than the test accuracy for some cutoff reviews and training sample sizes, but lower for others. This suggests that the performance of the model is more variable than the embedding layer model.
* The findings indicated that the straightforward **embedding layer** model performed **better** than the **pre-trained model**, which goes against the common assumption that pre-trained embeddings enhance model performance. Generally, it’s important to consider that the pre-trained model here is not optimized for the task specified and didn’t **fine-tune** the embeddings during training. Essentially, Fine-tuning the embeddings might lead to better performance.
* Finally, we should be attentive in drawing conclusions from these results as they are based on a small number of training samples and a **limited set of hyperparameters.** It's possible that different hyperparameters or more training data could lead to different conclusions.