Assignment 3

Recurrent Neural Networks

Deep learning models called recurrent neural networks (RNNs) are excellent for sequential data processing tasks like natural language processing (NLP) tasks. RNNs excel at tasks involving sequences of words, phrases, or paragraphs because of their exceptional capacity to grasp sequential dependencies in data. In this study, we conducted sentiment analysis on the IMDb movie review dataset using RNNs with embedded layers.

Train sample 100, Validation 10000, Test 12500:

* Initial Setup:
  1. IMDB review dataset has been imported for this assignment.
  2. The initial setup for the model was taking 100 training samples with each review of length 150 words max and a total of 10000 words are taken as input for the model.
  3. Also, this model is validated against 10000 validation samples of both positive and negative reviews.
  4. The loss function “binary cross-entropy” was used as it was a classification model with optimizer “Adam”.
* Models Trained:
  1. There are four models trained, validated, and tested using the initial setup with performance metric as accuracy.
  2. One hot encoded sequence model has achieved a test accuracy of 0.62 and test loss of 0.6497.
  3. Embedded model without masking gave a test loss of 0.5423 and test accuracy of

0.6952.

* 1. Embedded model with masking gave a test loss of 0.6726 and test accuracy of

0.5823.

* 1. A pre-trained model Global Vectors for word representation (GloVe) gave a test loss of 0.7139 and test accuracy of 0.586.

Training Models Across Different Sample Sizes:

The models have been trained using different training sample sizes from 100 to 10000 and their test loss and accuracy are recorded in below table,

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| sample size | one hot encoded sequence | | Embedded | | Embedded masked | |  | Pretrained |
|  | Test  Loss | Test  Accuracy | Test  Loss | Test  Accuracy | Test  Loss | Test  Accuracy | Test  Loss | Test Accuracy |
| 100 | 0.6497 | 0.62 | 0.5423 | 0.6952 | 0.6726 | 0.5823 | 0.7139 | 0.586 |
| 500 | 0.6114 | 0.687 | 0.6822 | 0.555 | 0.6534 | 0.617 | 0.6806 | 0.57 |
| 1250 | 0.6164 | 0.7495 | 0.6478 | 0.641 | 0.6457 | 0.639 | 0.5272 | 0.7744 |
| 2500 | 0.5897 | 0.715 | 0.6120 | 0.679 | 0.6169 | 0.679 | 0.4843 | 0.77 |

The investigation's findings demonstrated that, when it came to sentiment analysis, RNNs with embedded layers outperformed alternative word embedding strategies like one-hot encoded sequence. Test accuracy and test loss were consistently higher with the embedded layer-based models than with other methods.

Additionally, it is evident that as sample size increased, the RNN-based models' performance improved. The test loss of the RNN-based models decreased from approximately 64.9% to less than 48% as the sample size rose from 100 to 2500 samples. This implies that greater sample numbers provide the model access to more training data, which improves performance.

Additionally, several embedded layer types—such as conventional embedded and masked embedded layers—are compared. When compared to masked embedded layers, the normal embedded layer-based models performed marginally better in terms of test accuracy. Even though the masking technique enables the model to ignore padding tokens and concentrate only on the word embeddings, improving performance and producing more meaningful representations, it can be observed that masking has no effect on the imdb dataset used in this model implementation.

Using pre-trained word embeddings—specifically, GloVe embeddings—produced more effective and efficient models than training embedded layers from scratch, according to the IMDb movie review dataset. After training on 2500 samples, the pre-trained model outperformed both masked and standard embedded layers in terms of test accuracy, achieving a test loss of 48.4%.

**Conclusion:**

It may be inferred from the experiment's analysis and results that higher sample numbers typically translate into better performance. The model has more data to work with and is probably going to generalise better to new data as the sample size grows.

The ideal sample size, however, may change based on the particular job, dataset, and model architecture employed. Sample sizes of 1250 and 2500 were found to produce good results in terms of test accuracy and test loss in the Imdb dataset that was presented; lower numbers indicated higher performance. With respect to various embedding approaches, such as pre-trained Glove embeddings and masked and conventional embedded layers, these sample sizes consistently demonstrated increased performance.

Furthermore, compared to alternative embedding techniques, the pre-trained GloVe embeddings consistently performed better across a range of sample sizes, demonstrating higher accuracy and reduced loss. This indicates that the advantages of pre-trained GloVe embeddings are more pronounced when there is a shortage of training data since they give the model a solid initialization and make use of the information gathered from a large volume of data during pre-training.

In conclusion, it can be seen that the pre-trained GloVe embeddings performed better than the other embedding approaches based on the findings from the examination of masked and conventional embedded layers. Across a range of sample sizes, the pre-trained GloVe embeddings consistently outperformed the one-hot encoded sequence and embedded layers in terms of test loss and test accuracy. Although they did not consistently beat the pre-trained GloVe embeddings or the regular embedded layers, the masked embedded layers also displayed encouraging results.

Therefore, compared to other embedding techniques, pre-trained GloVe embeddings are more effective for sentiment analysis tasks because they capture a large amount of semantic and syntactic information from large corpora, require less training data, provide a standardised representation, and are simple to use and implement.