**Experiment Report: Comparison of Models with Embedding Layer vs. Pre-trained Word Embedding**

**1. Introduction**

This experiment aimed to compare the performance of two models in sentiment analysis on the IMDB dataset. The models were trained using different approaches: one with an embedding layer initialized with random weights and the other with pre-trained GloVe word embeddings.

**2. Experiment Setup**

* **Dataset**: IMDB dataset consisting of movie reviews with binary sentiment labels.
* **Data Preprocessing**:
  + Reviews truncated to 150 words.
  + Training samples limited to 100.
  + Validation on 10,000 samples.
  + Considered only the top 10,000 words.
* **Models**:
  + **Model 1**: Embedding layer + Bidirectional LSTM + Dense layer.
  + **Model 2**: Pre-trained GloVe word embeddings + Bidirectional LSTM + Dense layer.
* **Training Parameters**:
  + Epochs: 10
  + Batch Size: 32
  + Optimizer: RMSprop
  + Loss Function: Binary Crossentropy

**3. Results**

**Model with Embedding Layer**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Epoch** | **Training Loss** | **Training Accuracy** | **Validation Loss** | **Validation Accuracy** |
| 1 | 0.6947 | 0.4500 | 0.6942 | 0.5027 |
| 2 | 0.6855 | 0.5800 | 0.6940 | 0.5027 |
| 3 | 0.6828 | 0.5800 | 0.6947 | 0.5027 |
| 4 | 0.6781 | 0.5800 | 0.6949 | 0.5027 |
| 5 | 0.6732 | 0.5800 | 0.6997 | 0.5027 |
| 6 | 0.6640 | 0.5800 | 0.6977 | 0.5027 |
| 7 | 0.6581 | 0.5800 | 0.7032 | 0.5027 |
| 8 | 0.6479 | 0.5800 | 0.6968 | 0.5027 |
| 9 | 0.6412 | 0.5800 | 0.6933 | 0.5039 |
| 10 | 0.6314 | 0.6600 | 0.6961 | 0.5030 |

**Model with Pre-trained Word Embedding**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Epoch** | **Training Loss** | **Training Accuracy** | **Validation Loss** | **Validation Accuracy** |
| 1 | 0.6993 | 0.5300 | 0.7219 | 0.5034 |
| 2 | 0.6730 | 0.5800 | 0.7087 | 0.5024 |
| 3 | 0.6652 | 0.6000 | 0.7309 | 0.5028 |
| 4 | 0.6620 | 0.5800 | 0.7459 | 0.5027 |
| 5 | 0.6622 | 0.5800 | 0.7110 | 0.5026 |
| 6 | 0.6472 | 0.6000 | 0.7065 | 0.5037 |
| 7 | 0.6417 | 0.6100 | 0.7045 | 0.5017 |
| 8 | 0.6387 | 0.6100 | 0.6982 | 0.4983 |
| 9 | 0.6328 | 0.7100 | 0.7056 | 0.5031 |
| 10 | 0.6244 | 0.6400 | 0.7100 | 0.5013 |

**4. Discussion**

* Both models demonstrated poor performance with validation accuracies around 50%, indicating limited generalization capability.
* The fluctuation in validation accuracy suggests that the models may be overfitting to the training data, given the small dataset size.
* Despite utilizing pre-trained word embeddings, the second model did not yield significant improvements over the first model with randomly initialized embeddings.

**5. Conclusion**

* The experiment highlights the challenges of training deep learning models for sentiment analysis with limited data.
* Addressing the limitations of data scarcity is crucial for improving model performance in natural language processing tasks.
* Future work should focus on acquiring more labeled data, exploring data augmentation techniques, and experimenting with different model architectures to enhance performance.