Comparing Feed-Forward, Bi-LSTM, and Regularized Networks for IMDB Sentiment Analysis

Arpit Kumar

July 28, 2025

Abstract

This paper presents a comparative study of three neural network architectures for sentiment analysis: a baseline Feed-Forward Neural Network (FFNN), a sequence-aware Bidirectional Long Short-Term Memory (Bi-LSTM) network, and an FFNN enhanced with Dropout regularization. The models are evaluated on the IMDB movie review dataset. Experimental results show that the Bi-LSTM achieves the highest test accuracy (86.76%), narrowly outperforming the regularized FFNN (86.68%) and the original FFNN (86.54%). The findings highlight the inherent advantage of sequence-modeling architectures for language tasks, while also demonstrating that regularization provides a marginal but clear benefit to the simpler FFNN model.

1 Introduction

Automated sentiment analysis from text is a core task in modern Natural Language Processing. The challenge lies in creating models that understand language beyond a simple "bag-of-words" approach. This study compares three distinct deep learning architectures to investigate the trade-offs between model complexity, sequential processing, and regularization.

We evaluate a simple FFNN baseline, a sequence-aware Bi-LSTM, and a regularized FFNN using Dropout. The goal is to quantify the performance differences and analyze the architectural reasons behind them on the IMDB sentiment classification task.

2 Methodology

2.1 Dataset and Preprocessing

The study utilizes the IMDB dataset [1], containing 50,000 movie reviews. Preprocessing was standardized for all models: a vocabulary limited to the top 10,000 words, integer tokenization, and padding sequences to a length of 200.

2.2 Model Architectures

2.2.1 Feed-Forward Network (FFNN)

Our baseline FFNN uses an `Embedding` layer, a `GlobalAveragePooling1D` layer to create a fixed-size vector from the sequence, a `Dense` hidden layer, and a final `sigmoid` output layer. This architecture intentionally discards word order information.

2.2.2 Bidirectional LSTM (Bi-LSTM)

This model is designed for sequential data. It replaces the pooling and dense layers with a `Bidirectional(LSTM)` layer [2], allowing it to process text in both forward and reverse chronological order, thereby capturing contextual dependencies [3].

2.2.3 Regularized FFNN

This model extends the baseline FFNN by adding a `Dropout` layer [4]. Dropout is a regularization technique that randomly sets a fraction of input units to 0 at each update during training. This helps prevent complex co-adaptations on training data, thus improving the model's ability to generalize to unseen data. A dropout rate of 0.5 was used.

3 Results

All models were trained for 3 epochs using the Adam optimizer. The final test accuracies are presented in Table 1 and visualized in Figure 1.

Table 1: Performance Comparison on the IMDB Test Set.

Model Architecture	**Test Acc.**
Original FFNN	0.8654
Regularized FFNN	0.8668
Bidirectional LSTM (Bi-LSTM)	0.8676

The Bi-LSTM model achieved the highest accuracy, followed closely by the Regularized FFNN. The Original FFNN had the lowest accuracy of the three, though the performance difference between all models is marginal.

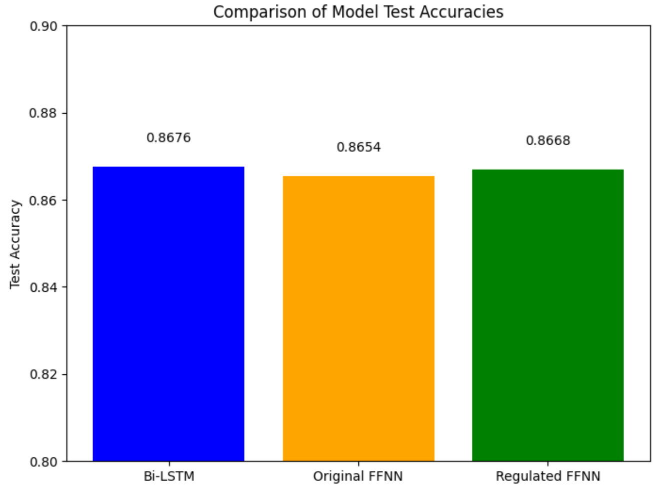

Figure 1: Comparison of Final Test Accuracies for the three models.

4 Discussion

The results confirm that for text classification, architectures that model sequence information, like the Bi-LSTM, hold an advantage. Even a small one indicates a better capacity to understand context.

Interestingly, the introduction of Dropout regularization provided a tangible improvement to the FFNN's performance, pushing it closer to that of the Bi-LSTM. This suggests the baseline model was experiencing minor overfitting, which the Dropout layer helped mitigate. However, the fact that regularization alone did not close the gap entirely reinforces the importance of architectural design choices tailored to the data's nature (i.e., sequential text).

Analysis of misclassified examples from the regulated model showed it still struggled with certain nuances, occasionally predicting "Positive" for reviews that were clearly negative. This indicates that even with high overall accuracy, all models have specific linguistic blind spots.

5 Future Scope

While this study provides a clear comparison, several avenues exist for future work:

- **Hyperparameter Tuning:** The performance of the regularized FFNN could potentially be improved by systematically tuning the dropout rate, number of neurons, and number of training epochs.
- **Deeper Architectures:** Exploring deeper FFNNs (with more dense layers) could be beneficial, as initial tests showed promise in improving validation accuracy during training.
- **Error Analysis:** A detailed qualitative analysis of misclassified reviews could reveal patterns in the types of language or sentiment that the models find challenging, informing future model improvements.
- **Advanced Models:** Comparing these models to state-of-the-art Transformer-based architectures like BERT [5] would provide a more complete picture of their relative performance in the modern NLP landscape.

6 Conclusion

This study demonstrated that while a simple FFNN provides a strong baseline, sequence-aware models like the Bi-LSTM are architecturally better suited for sentiment analysis. Furthermore, we showed that regularization techniques like Dropout can improve the generalization of simpler models, though not enough to overcome the inherent advantages of sequence modeling. The close performance of all three models suggests that for this specific dataset, much of the sentiment can be gleaned from a "bag-of-words" approach, but a small, consistent edge is gained by understanding context and word order.

References

[1] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts, "Learning word vectors for sentiment analysis," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, 2011, pp. 142–150.

[2] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997.

[3] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[4] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[5] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008.