Project Report: Sentiment Analysis with RNN and LSTM Models

Juan David Bahamon
Samuel Hernández
David Peñaranda

1. Introduction and Objectives of the Project:

The goal of this project is to develop a sentiment analysis model using supervised learning with two different types of Recurrent Neural Networks (RNN): a basic RNN and a Long Short-Term Memory (LSTM) network. Sentiment analysis involves determining the emotional tone behind a text, whether positive or negative. The project aims to explore the effectiveness of these models in classifying sentiments and identify possible improvements in sentiment analysis accuracy.

2. Dataset Details and Preprocessing Steps:

The dataset used for this project is the "Sentiment Labelled Sentences Dataset" obtained from the University of California, Irvine Machine Learning Repository. This dataset consists of sentences extracted from three different websites: amazon.com, imdb.com, and yelp.com. Each sentence is labeled as 1 (positive) or 0 (negative). The dataset was preprocessed with tokenization, lowercasing, and the removal of stop words, using the NLTK library.

3. Description of RNN and LSTM Sentiment Analysis Models:

3.1 Basic RNN:

The basic RNN model is a simple recurrent neural network architecture. It consists of a simple chain of recurrent layers. The training procedure involves forward and backward steps over time to capture sequential dependencies. The model is trained to understand the sequential patterns in sentiment data.

3.2 LSTM:

The LSTM model, a type of recurrent neural network, incorporates long-term memory capabilities. It mitigates the problem of gradient vanishing and is particularly effective in capturing long-term dependencies in sequential data. The architecture includes memory cells, input gates, forget gates, and output gates. Training involves learning the weights that optimize the model for sentiment analysis.

These more advanced models are better at understanding and describing the complexity of how words relate to each other in sentiment texts. This means they are able to more accurately classify whether a text is positive or negative because they can notice finer patterns and capture the nuances of emotions in the text. In the case of the LSTM model, which has a kind of "long-term memory," this helps a lot in understanding the context and makes the model even better at predicting and understanding emotions in sentiment analysis.

4. Performance Evaluation:

The models were evaluated using standard metrics, including accuracy, precision, recall, F1 score, and Cohen's kappa. These metrics provide a comprehensive understanding of the models' performance in sentiment classification. The DummyClassifier was used as a baseline for comparison:

Accuracy: 0.5091
Precision: 0.5091
Recall: 1.0000
F1 Score: 0.6747

RNN:
Accuracy: 0.5915
Precision: 0.6009
Recall: 0.5880
F1 Score: 0.5944
Cohen's Kappa: 0.1830

RNN with best hyperparameters:
Accuracy: 0.7333
Precision: 0.7475
Recall: 0.7190
F1 Score: 0.7330
Cohen's Kappa: 0.4668

LSTM:
Accuracy: 0.7963
Precision: 0.7863
Recall: 0.8238
F1 Score: 0.8046
Cohen's Kappa: 0.5922

LSTM with best hyperparameters:
Accuracy: 0.7975
Precision: 0.8019
Recall: 0.8
F1 Score: 0.8009
Cohen's Kappa: 0.5950

5. Comparative Analysis:

Progressive Improvement:
All models outperform the DummyClassifier, demonstrating significant improvements in all evaluated metrics.

Impact of Hyperparameters on RNN and LSTM:
Hyperparameter optimization substantially improves the performance of both RNN and LSTM, evidenced by an increase in all metrics, with a particularly notable improvement in the F1 Score and Accuracy

Superiority of LSTM:

Even with the best RNN hyperparameters, LSTM retains its superiority, exhibiting the highest accuracy, precision, and F1 Score, as well as the highest Cohen's Kappa.

DummyClassifier (Baseline):
Precision: 0.5091

RNN (Original):
Precision: 0.6009
Significant improvement in precision compared to DummyClassifier, indicating that the RNN can more accurately discern between positive and negative sentiments.

RNN (Best Hyperparameters):
Precision: 0.7475
Further optimization of the RNN results in a substantial increase in precision, highlighting the importance of adjusting hyperparameters to improve performance.

LSTM (Original):
Precision: 0.7863
LSTM outperforms both DummyClassifier and RNN in precision, demonstrating advanced capability for more accurate sentiment classification.

LSTM (Best Hyperparameters):
Precision: 0.8019
Hyperparameter optimization in LSTM confirms the trend of improvement, achieving the highest precision among all evaluated models.

6. Conclusion and Future Work:
In conclusion, the project successfully implemented and evaluated sentiment analysis models using basic RNN and LSTM architectures. Both models showed promising results, outperforming the reference DummyClassifier. Future work could include experimenting with more sophisticated architectures, exploring ensemble methods, and incorporating domain-specific embeddings to enhance sentiment analysis.