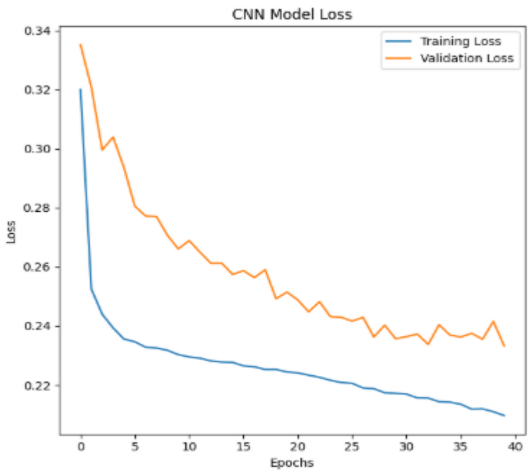
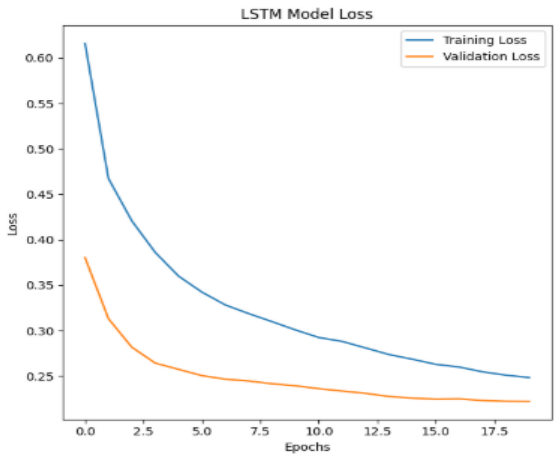
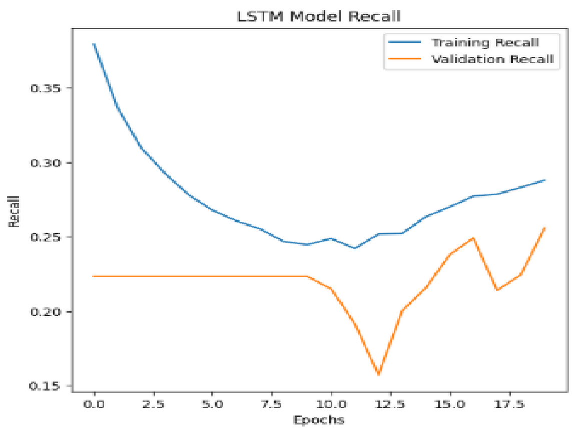
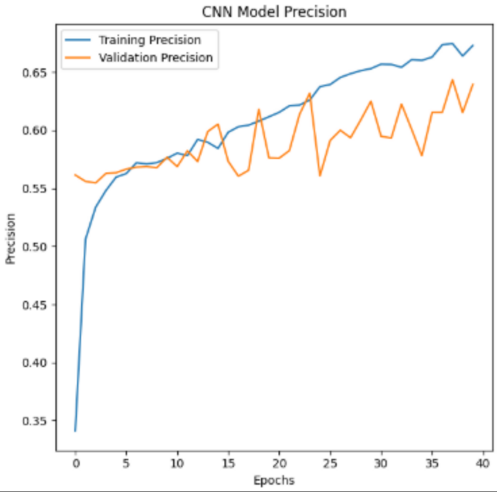
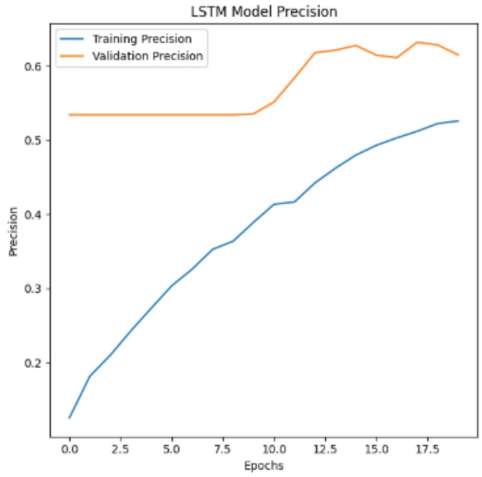
Multi-Model Analysis with Keras

Introduction

We have evaluated two machine learning models. They categorise films using CNN for posters and LSTM for overviews. The models were trained on an IMDb dataset, and their performance was assessed using loss, precision, and recall metrics on both training and validation sets. A critical analysis was conducted, illustrating the strengths and weaknesses of each model with practical examples. This analysis highlights how each model performs with different types of data and where they might fall short, providing insights into their effectiveness for movie genre classification tasks.

Performance Review of Models

CNN Model	LSTM Model
In the graph, we can observe that there is steady decrease in the Training Loss, indicating that the model is learning effectively on the training set. However, the Validation Loss decreases initially but remains constant after 20 epochs with minor fluctuations.This indicates that the model may be experiencing overfitting after 20 epochs. The performance on the model continues to improves on the training data, but on the validation data stagnates, implying it is not generalizing well to unseen data.  	The graphs shows consistent decline in both Training Loss and Validation Loss throughout the training process, which indicates that the LSTM model is learning effectively. Additionally, the gap between training and validation loss remains relatively small, which suggests that the LSTM model is experiencing less overfitting compared to the CNN model 
The below graphs shows a gradual improvement in Training Recall, which stabilizes around 30 epochs.  Indicates that the model is effectively learning to identify the correct categories within the training set.	The below graph shows Training Recall line initially decreases and then begins to increase after about 5  epochs. Indicating a learning curve where the model adjusts before starting to improve. However, the Validation Recall remains mostly flat and low

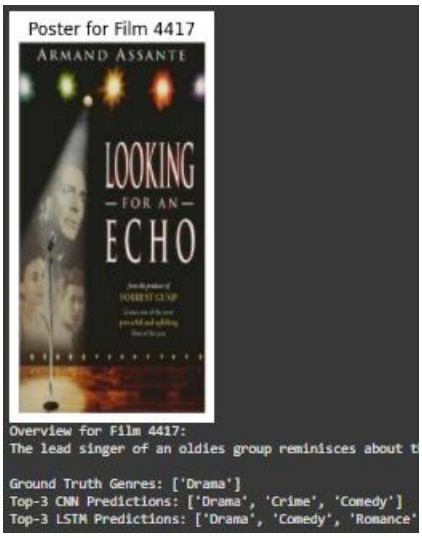
<table>
<tr>
<td>However, the Validation Recall fluctuates significantly, showing inconsistency after the initial improvement. The model performs well on training data but fails to maintain consistent recall on unseen validation data.</td>
<td>throughout training, suggesting that the model struggles to correctly recall categories on new data.The model may not be effectively capturing the patterns in the data needed to improve recall performance.</td>
</tr>
<tr>
<td>In the Precision Graphs, the CNN model demonstrates a steady improvement in Training Precision, reaching around 0.65 by the end of 40 epochs. The Validation Precision shows an initial increase, followed by fluctuations after 15 epochs. The fluctuations in validation precision indicate the presence of some overfitting, where the model's performance on new data is less consistent than on the training data.

</td>
<td>In the Precision Graphs, theTraining Precision in LSTM increases consistently over time, though it remains below the validation precision throughout the training process. The Validation Precision stays relatively high and stable, indicating that the LSTM model maintains good precision on unseen data.

</td>
</tr>
</table>

The CNN model shows clear signs of overfitting, as its training metrics (loss, accuracy, and recall) continue to improve while the validation metrics either plateau or fluctuate. This overfitting results in inconsistent recall, indicating the model struggles to identify all relevant instances in the validation set, even though its accuracy suggests it can provide accurate predictions. In contrast, the LSTM model shows less overfitting due to consistent loss reduction for both training and validation sets. The LSTM produces fewer but more accurate predictions, achieving steady accuracy. However, it struggles with recall, indicating that while it performs well in terms of precision, it may still miss relevant instances in the data.

Prediction Analysis of Models

Let's provide the models with some movie posters or overviews as input and analyze their predictions.

For Film 4308, the ground truth genres are Crime, Drama, and



For Film 4417, the sole ground truth genre is Drama. Both the

For Film 4290, the ground truth genre is Comedy. The CNN model

Thriller. The CNN model predicted Drama, Action, and Thriller, correctly capturing two out of the three genres but missing Crime. The LSTM model predicted Drama, Action, and Comedy, capturing only Drama while missing Crime and Thriller. The CNN performed better by identifying more relevant genres, while the LSTM struggled to match the ground truth and included the unrelated genre Comedy.

CNN and LSTM models correctly identified Drama. However, the CNN also predicted Crime and Comedy, while the LSTM predicted Comedy and Romance. Although both models identified the correct genre, they added extraneous predictions. This indicates that while both models can recognize Drama accurately.

correctly predicted Comedy but also included Adventure and Drama. The LSTM model similarly predicted Comedy but added Drama and Romance. Both models successfully identified Comedy, but their additional predictions suggest they struggle to limit their outputs to only relevant genres.

When comparing the performance of the CNN and LSTM models, the CNN model outperforms the LSTM model in terms of total movie genre prediction accuracy. CNN, which analyses visual inputs from movie posters, consistently identifies more relevant genres, especially those with significant visual characteristics such as Action, Thriller, and Drama. While the CNN shows evidence of overfitting (the model performs well on training data but less consistently on validation data), its capacity to attain better recall makes it more successful for genre prediction tasks based on visual cues. Despite occasional misclassifications, CNN forecasts are often closer to the ground reality than the LSTM.

On the other hand, the LSTM model, which analyses textual overviews, exhibits less overfitting and has rather steady accuracy. However, it suffers greatly with recollection, often failing to record all important genres. The LSTM predicts fewer genres, preferring accuracy over breadth, which means it may overlook important characteristics of the film's genre categorisation. This constraint reduces the LSTM's effectiveness when complete genre identification is needed. As a result, although both models have advantages, the CNN model is the superior option for movie genre prediction owing to its stronger recall and greater capacity to use visual information in posters.

Conclusion:

The CNN model is often better at detecting genres from visual inputs, accurately predicting more relevant genres while displaying evidence of overfitting and over-predicting unimportant categories. The LSTM model does relatively well with textual inputs, often capturing at least one proper genre but suffering with memory and adding extraneous genres. Both models might benefit from strategies that promote generalisation and accuracy, such as regularisation, balanced datasets, and ensemble methods, to improve overall performance.

GitHub Link for code:

https://github.com/pratapponnam/ADS2-MultiModel

Author: Pratap Bhargav Ponnam

ID:23027654