University of Hertfordshire UH

Applied Data Science 2

Multimodal IMDB Analysis with Keras

Author:

Cristina Baron Suarez 23069038

April 21, 2025

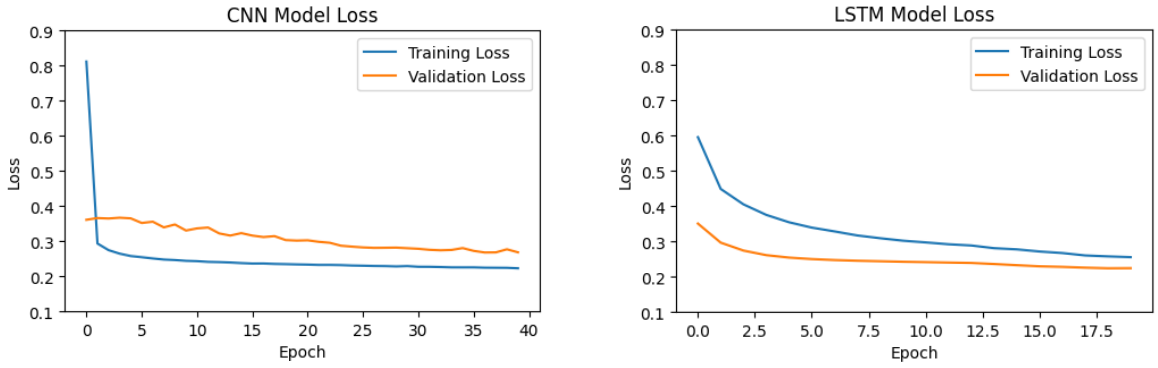
1 Introduction

In this assignment, we want to classify film posters and overviews by genre. For that, we will implement and train a CNN model and an LSTM model, and we will determine which is better at classifying films.

2 CNN Model VS LSTM Model

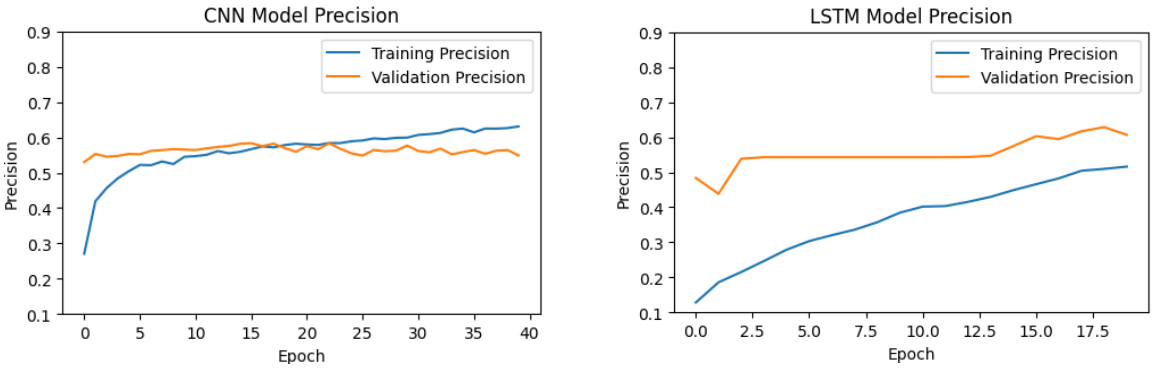
In order to evaluate the models, several plots have been created using the losses and metrics of the trained models. In this section, we discuss these results and evaluate the performance of the models.

2.1 Loss



The CNN model learned to classify film posters by genre faster than the LSTM model learned to classify film overviews by genre. However, the decrease in training loss in the CNN model without a corresponding improvement in validation performance suggests that it might be suffering from overfitting. In contrast, since LSTMs learn sequential dependencies, their training is more gradual, especially if the film overviews are long. Although this model learns more slowly, it may be generalizing better, as indicated by the validation loss, which continues to decrease over several epochs after the initial stages.

2.2 Precision



The CNN model is very good at detecting features and may reach high precision when classes have clear local patterns. However, if classification relies on the film overview, the LSTM needs more epochs, as it captures a wider context.

CNN consistently shows higher training precision than the LSTM, which might lead to the conclusion that this last model is undertrained or that posters give clearer genre clues than overviews. However, CNN's validation precision stops improving early, while training precision continues to increase, suggesting overfitting.

Meanwhile, the LSTM validation precision keeps improving overtime, indicating better generalization and effective use of the overview. So, rather than being undertrained, the LSTM likely just needs more epochs, while the CNN may be overfitting.

2.3 Recall

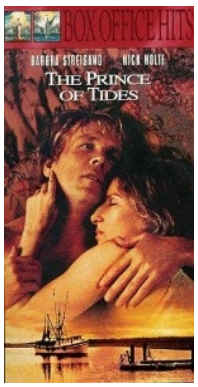
In the CNN model, the validation recall decreases over time, despite the training recall continuing to increase. This divergence suggests, once again, overfitting.

In constrast, the LSTM model shows an upward trend in validation recall after some epochs, even if it starts lower. This indicates that although the model learns more slowly, it eventually generalizes better. This gradual improvement supports the idea that film overviews may provide deeper contextual clues that help to retrieve more difficult examples.

Furthermore, while CNN initially shows stronger performance in training recall, LSTM demonstrates better long-term generalization, which is more desirable.

3 Examples

Three film examples were used to demonstrate the classification performance of the CNN model on their posters and of the LSTM model on their overview. For that, we are going to compare, for each example film, the top three most probable genres predicted by the CNN and by the LSTM models, with its truth genres.



Overview: A troubled man talks to his suicidal sister's psychiatrist about their family history and falls in love with her in the process.

CNN Top-3 Predicted Genres:

- Drama: 0.572
- Comedy: 0.449
- Romance: 0.398

LSTM Top-3 Predicted Genres:

- Drama: 0.728
- Comedy: 0.521
- Romance: 0.265

Actual Genre(s): Drama

Overview: The Griswold family's plans for a big family Christmas predictably turn into a big disaster.

CNN Top-3 Predicted Genres:

- Drama: 0.604
- Comedy: 0.452
- Romance: 0.421

LSTM Top-3 Predicted Genres:

- Drama: 0.616
- Comedy: 0.460
- Romance: 0.224

Actual Genre(s): Comedy

Overview: Former cop Brian O'Conner is called upon to bust a dangerous criminal and he recruits the help of a former childhood friend and street racer who has a chance to redeem himself.

CNN Top-3 Predicted Genres:

- Comedy: 0.534
- Drama: 0.531
- Romance: 0.371

LSTM Top-3 Predicted Genres:

- Action: 0.421
- Comedy: 0.336
- Drama: 0.300

Actual Genre(s): Action, Crime, Thriller

The LSTM's ability to understand sequences often helps it better identify the correct main genre (compared to the CNN), like "Action" in *2 Fast 2 Furious*. However, there are ambiguous cases, such as *Christmas Vacation*, where both models incorrectly guess "drama". This could be because words like "family," "big," "disaster," "plans" could appear in the synopsis of many dramatic or family-oriented films classified as "drama".

4 Conclusion

By analyzing the curves, we can tell that CNNs typically excel in tasks where local, spatial features are critical and may converge faster; while LSTMs are better at capturing long-term dependencies, which can lead to improved recall, and sometimes better precision, after an adequate number of training epochs.

The CNN model fits the training data quickly and effectively, but its performance in the validation data does not improve accordingly, exhibiting signs of overfitting. However, the LSTM model learns more gradually, but its performance in validation data improves over time, suggesting a more stable generalization.

When analyzing examples, we notice that both models tend to overpredict the genre "drama." This may be due to an overrepresentation of that genre in the training set or because certain words or visual elements commonly linked to dramatic content appear frequently.

Moreover, instead of focusing on just a few test cases, it would be interesting to evaluate the performance of the model per genre. Calculating the recall for each genre after training could reveal whether the "drama" bias is a consistent issue and help identify which genres are being misclassified more often.

Also, to achieve better results, a potential solution could be designing a hybrid model that combines CNNs and LSTMs to take advantage of both architectures.