CW1 - TensorFlow - Multimodal IMDB Analysis with Keras

Student Name: Sridhar Guggilla

Student ID: 23021710

Critical Analysis Report for Multimodal IMDB Genre Classification

This report showcases a summary of critical analysis regarding the outcome obtained after implementation and training of a Convolutional Neural Network (CNN) and a Long Short-Term Memory (LSTM) network in multi-label genre classification on the IMDB dataset experiment. The critical analysis will look into the model architectures, training process, performance trends based on training logs, and ultimately critically evaluate the efficacy of both models.

Model Architectures

Architecture of the CNN Model

The CNN model implemented using the Keras Functional API possesses a hierarchical architecture to extract features from images and classify them. It consists of three convolutional blocks where each is comprised of two convolutional layers with ReLU as the activation function, followed by a dropout layer for avoiding overfitting with a rate of 0.2, and finally, a MaxPooling layer downsampled the feature maps to reduce computational complexity while preserving the most salient features.

LSTM Model Architecture

This is done in the form of a sequential model and it takes input text as a sequence of numeric tokens from a *TextVectorization* layer. After that, the *embedding layer* maps these tokens into size 256 dense vectors while handling semantic relationships between words. A mask_zero=True is set in this embedding layer to handle variable-length input sequences efficiently.

Model Training and Evaluation

Training Process

Both CNN and LSTM are trained by using the Adam optimizer due to its adaptive learning rate capabilities, which make it suitable for complex and high-dimensional data as in image and text classification. The learning rate of 1e-4 has been selected as a trade-off between speed and stability. Here, the binary cross-entropy loss is used since this is a multi-label classification task because each movie could be tagged with more than one genre. It is a multi-class and multi-label problem. For both models, precision and recall are used.

Evaluation and Discussion

CNN Model Performance Analysis

The training of the CNN model was completed using 40 epochs, which has been recorded from the training logs. The figure below shows an analysis plot which shows the training logs for both training and validation for CNN model across all 40 epochs.
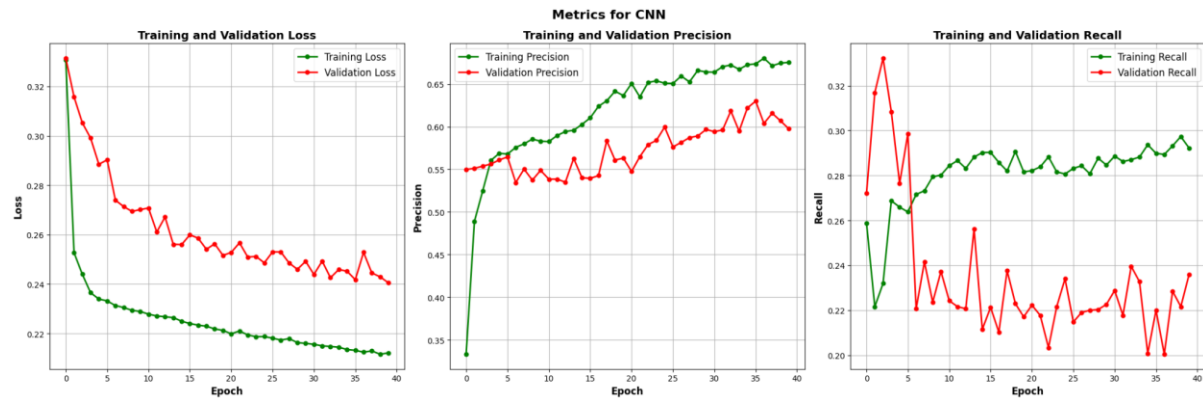
LSTM Model Performance Analysis

This trained the LSTM model for 20 epochs, with the dynamics of performance captured in the training logs. The figure below shows how the performance of LSTM model across each of the epochs as a trend plot.
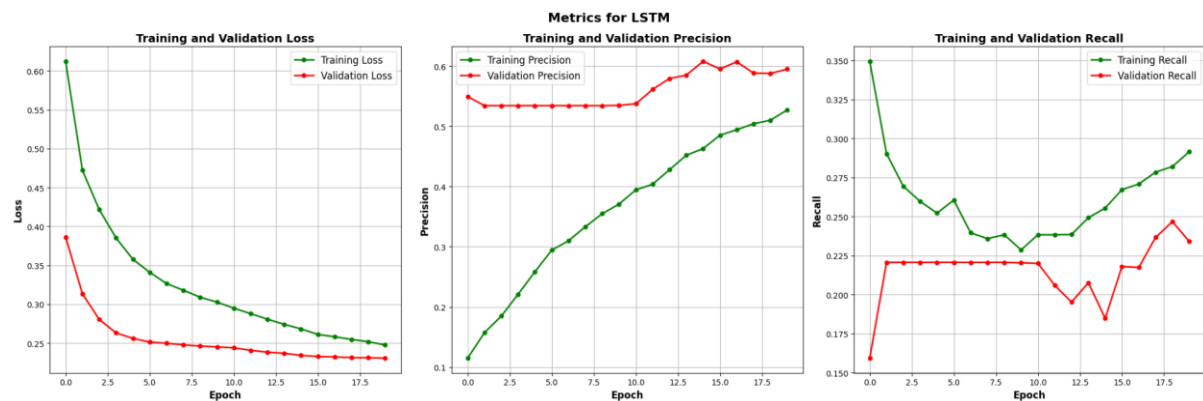
Performance Comparison

In addition to above metrics, further performance testing for the models was done by comparing predictions each has made against ground truth genre labels for a sample size of 10 films. The screenshot below shows the top 3 labels predictions from the two models against the actual label.

To accurately compare the below, an accuracy plot was done by counting the number of accurately predicted genres by each model on the ten selected items and after that an accuracy plot was done. The figure below shows a performance pie chart which tries to compare how LSTM and CNN performed on 10 randomly selected posters.

```
################################################################################
Showing Predicitions Vs actual for each model:
################################################################################
Film: /content/drive/MyDrive/Multimodal_IMDB_dataset/Images/tt0117331.jpg
Ground truth genres: ['Action', 'Adventure', 'Comedy']
Top 3 genres predicted by CNN (poster): ['Drama' 'Comedy' 'Romance']
Top 3 genres predicted by LSTM (overview): ['Action' 'Drama' 'Comedy']
-------------------------------------------------
Film: /content/drive/MyDrive/Multimodal_IMDB_dataset/Images/tt0089304.jpg
Ground truth genres: ['Comedy']
Top 3 genres predicted by CNN (poster): ['Comedy' 'Drama' 'Romance']
Top 3 genres predicted by LSTM (overview): ['Drama' 'Comedy' 'Romance']
-------------------------------------------------
Film: /content/drive/MyDrive/Multimodal_IMDB_dataset/Images/tt0095186.jpg
Ground truth genres: ['Comedy', 'Drama', 'Romance']
Top 3 genres predicted by CNN (poster): ['Drama' 'Romance' 'Comedy']
Top 3 genres predicted by LSTM (overview): ['Drama' 'Comedy' 'Romance']
-------------------------------------------------
Film: /content/drive/MyDrive/Multimodal_IMDB_dataset/Images/tt0417614.jpg
Ground truth genres: ['Drama']
Top 3 genres predicted by CNN (poster): ['Comedy' 'Drama' 'Romance']
Top 3 genres predicted by LSTM (overview): ['Drama' 'Comedy' 'Romance']
-------------------------------------------------
Film: /content/drive/MyDrive/Multimodal_IMDB_dataset/Images/tt0108410.jpg
Ground truth genres: ['Comedy', 'Drama']
Top 3 genres predicted by CNN (poster): ['Drama' 'Comedy' 'Romance']
Top 3 genres predicted by LSTM (overview): ['Drama' 'Comedy' 'Romance']
-------------------------------------------------
Film: /content/drive/MyDrive/Multimodal_IMDB_dataset/Images/tt2172071.jpg
Ground truth genres: ['Comedy', 'Romance']
Top 3 genres predicted by CNN (poster): ['Drama' 'Comedy' 'Romance']
Top 3 genres predicted by LSTM (overview): ['Drama' 'Comedy' 'Romance']
-------------------------------------------------
Film: /content/drive/MyDrive/Multimodal_IMDB_dataset/Images/tt0117774.jpg
Ground truth genres: ['Action', 'Crime', 'Drama']
Top 3 genres predicted by CNN (poster): ['Drama' 'Action' 'Thriller']
Top 3 genres predicted by LSTM (overview): ['Drama' 'Comedy' 'Romance']
-------------------------------------------------
Film: /content/drive/MyDrive/Multimodal_IMDB_dataset/Images/tt0112950.jpg
Ground truth genres: ['Comedy', 'Drama', 'Music']
Top 3 genres predicted by CNN (poster): ['Comedy' 'Drama' 'Adventure']
Top 3 genres predicted by LSTM (overview): ['Drama' 'Comedy' 'Romance']
-------------------------------------------------
```

Future Work

To be able to address the issues identified above, several ways to further improve the performance of multimodal movie genre classification models may be considered as the following;

- Some additional data that are more recent needs to be extended by collecting more recent movies with more variation to really teach the model to learn and generalize better.

- Also, to ease issues with imbalance genres, techniques for label balancing such as oversampling the minority classes or with weighted loss functions maybe used as it may further improve performance on these under-represented genres.

- The deeper architecture of CNN and LSTM combined with an attention mechanism could further help improve the features learned to classify better.

- The experiment can be oriented toward alternative methods of multi-label classification, including hierarchical classification or label embedding techniques.

To conclude the experiment, the potential of multimodal analysis was shown by the CNN and LSTM models developed in this experiment for the classification of movie genres. However, a number of factors limited the performances of these two models. Such limitations need further research and development for the full realization of the potential of such models in this challenging task.