**CW1 - TensorFlow - Multimodal IMDB Analysis with Keras**

## **CNN and LSTM Model Performance Analysis**

**Introduction**

In this report, two models are analyzed – (1) CNN and (2) LSTM – to predict movies genre from their posters and overviews, respectively. A selection of films in the dataset were their posters, in JPEG format and overviews from Internet Movie Database (IMDB). Finally, the report will discuss the performance of the two models by using some of the examples to demonstrate how they do better and how they do worse.

### **Data Processing and Model Definition**

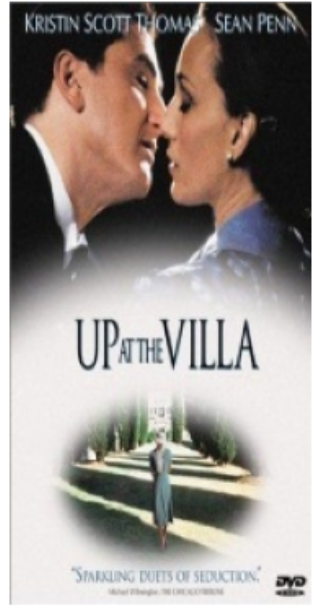
First, we processed the data in the IMDB dataset to prepare the data for both the model. The 64x64 images were resized to float32 datatype. The vocab size for the overviews was 10,000, built with the tf.keras.layers.TextVectorization layer. The dataset was fed on the CNN model built by Keras Functional API with multiple convolutional and max pooling layers followed by fully connected layers. However, to build the LSTM model, the embedding layer, two bidirectional LSTM layers, and two dense layers are used. Both models were compiled with precision and recall as the metrics and binary cross entropy loss, using the Adam optimizer.

### **Model Training and Evaluation**

Several callbacks were used to enhance energy model performance during training. For both models, the weights were saved for the best performing epoch based on validation loss using the ModelCheckpoint callback. To fine tune the models, we also used a LearningRateScheduler callback in order to decrease the learning rate during training. The learning rate of the CNN model decreases on the 10th epoch and trained for 40 epochs. As LSTM is a complex model, we trained it with 20 epochs using the same learning rate scheduling.

### **Model Performance Analysis**

**CNN Model:** Poster classification based on movie genres using the CNN model yielded high values and moderate results of classifying. It was able to often identify the dominant genre, but not multi label classification, and also not accurately predict secondary genres.



**Example 1:** The CNN model recognizes "Drama" (with confidence level 71.20%) and "Romance" (35.05%) as the most likely and the very second, respectively. Supposedly, the model incorrectly identifies the film as 'Up At the Villa' (tt0153464). As evidenced by ground truth genres of 'Drama' and 'Romance', these predictions agree. Yet, the model also guessed "Comedy" third most likely genre with 29.65%, not a correct ground truth genre for this film.



**Example 2:** "Alexander and the Terrible, Horrible, No Good, Very Bad Day" (tt1698641) - This film had "TV Movie" model prediction of genres for "Comedy". Nevertheless, the ground truth genre “Family” was not predicted but was replaced by “Drama” and “Adventure.”

The limitations of the CNN model can be drawn out from these examples. Most movie posters concentrate on their visual aspects whilst leaving out on the subtler sense of the storyline carried by secondary genres.

**LSTM Model:** However, overall LSTM showed better performance when predicting movie genres from text summaries than CNN model.

**Example 1: "Up At The Villa" (tt0153464)** - The LSTM model had very high confidence of predicting both 'Drama' (81.35%) and 'Romance' (50.30%), which line up with the ground truth genres.

**Example 2:** We can also look at "Alexander and the Terrible, Horrible, No Good, Very Bad Day" (tt1698641), where, though the LSTM model also predicted Comedy as the most probable genre, it missed out on Family category and simply chose Drama and Crime instead**.**

From these examples emerges the intuition that the LSTM model can handle the textual overview, and in some cases tends to capture storylines nuances which the CNN model, which only looks at the poster, overlooks. Nonetheless, although it can still run into trouble in deciding all the applicable genres, when confronted with multi-genre films, for example.

### **Discussion**

There is a performance difference between the CNN and LSTM models, because the nature of the input data and how each model processes it differ. With visual features obtained from the movie poster dataset, the CNN model is more susceptible to the lead theme or the most dominant element visually. As such, this can result in misclassifications, particularly when we have issues with multi label classification where a film is a part of more than one genre.

On the contrary, the LSTM model is able to gain from the richer information presented in the textual overview. By seeing the sequence of words and seeing the context the LSTM is able to catch the little hints about the different genres the film comprises. However, the performance of the model is still constrained by quality and detail in the overview. This wouldn’t give enough information for genre classification that is accurate, a brief or generic overview might not.

**Conclusion**

The results of both models yield valuable insights to the prediction capability of multimodal analysis for movie genre prediction. Yet the type of model and input data plays a significant part in performance. While the CNN can accurately identify dominant genres using poster cues, it fails at multi-label classification; and also, at capturing secondary genres. The LSTM model excels at genre identification because it is able to leverage richer information in textual summaries than can the CNN, and it can better handle short story nuances in a storyline that result from genre.