Multi Label Image Classification Using Deep Neural Network

Adithya Nair, Akshay Harikumar, Sreejith Kumara Pai
Department of Computer Science
Amrita Vishwa Vidyapeetham,
Amritapuri, India

***Abstract*—Movie posters encompasses the whole message and feeling of the film. Posters are more than just a promotional material, that captures a viewer's attention. A good poster is able to convey important qualities of a film such as theme and genre to make the movie seem as appealing to as wide of a viewership as possible. We aim to train and compare different Multi-label classification models based on their accuracy rate of learning the features from the posters and successfully predicting the genres of the movie. We were able to find out that our current model which follows a Deep Neural Network (DNN) Architecture was not to able to predict the genres with high accuracy.**

I. Introduction

A Movie Poster is the first impression of a movie. Posters are the face of a movie when it comes to promoting and advertising. Good posters are able to communicate important aspects of a film such as cast, theme, and elements of plot. Thus, designers have incentive to include salient features in their posters to make their movie attract more viewers. The ability to grasp the concept of the movie in such a way that, a viewer can identify the theme behind it, is crucial to a poster. The film-makers are able to get a feedback on how a viewer is able to visualize their film. Therefore a model which can extract the features of a movie poster and identify the genres can become handy for both the film-makers as well as the designers.

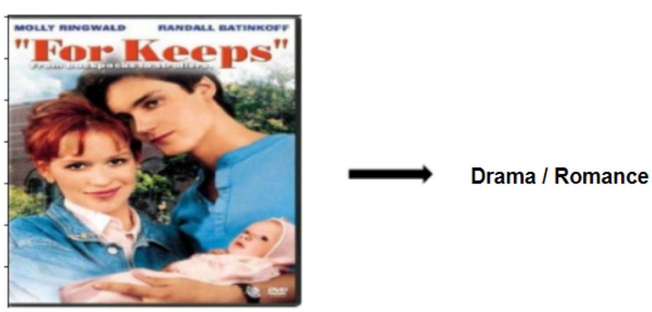

Fig. 1. A visualization of our problem

II. Methodology

Our model is trained using a feed-forward neural network. It takes the input, feeds it through several layers one after the other, and then finally gives the output. With the additional help of back-propagation and several hyper-parameter tuning techniques, we were able to minimize the loss function and improve our model.

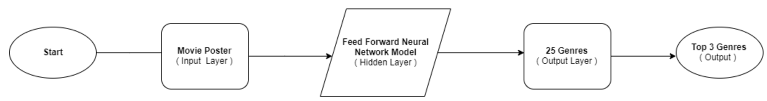

Fig. 2. Block Diagram of Feed Forward Neural Network Model

A. DataSet

We have collected a dataset from IMDB which contains Hollywood movies released between the years 1980 to 2015. Dataset includes 7867 images of movie posters with 25 genres where each posters belonging to more than one genre.

Each sample in the dataset consists of the image ID, corresponding categories of genres in text and the labels, where each label in the dataset consists of 25 genres which are one hot encoded according to their respective genres.

B. Data Preprocessing

Our image dataset consists of 3 channel RGB color space images of varying sizes, which had to be resized to a dimension of 350 x 350 pixels for better and efficient training. The resized images were then transformed to tensors. The dataset was then later split into training and test samples of sizes 75% and test 25% respectively, and loaded into training and testing dataloaders in batches of size 33.

Fig. 3. Image batches of dimension 350 x 350

C. Hyper Parameter Tunings

Our training process makes use of a gradient descent optimization method, which is able to find optimal values for the weights and biases such that the loss function is minimum. The model parameters are updated accordingly at every iteration using Stochastic Gradient Descent optimization with a learning rate of 0.05. A typical Feed Forward Model consists of a input layer, a hidden layer and an output layer. Each layer consists of neurons which are triggered based on the activation function applied. Our network architecture consists of a single linear input layer, three hidden layers and a final output layer. Both the input and hidden layers neurons are activated using ReLU activation, whereas the output layer is activated using a Sigmoid function. We use the Sigmoid activation function on the final layer as it converts each value of the final node into a probability score between 0 and 1 for easier classification.

D. Loss Function

When coming to loss function, we have used Binary Cross Entropy Loss (BCE Loss) which can be used along with sigmoid activation. Binary cross entropy compares each of the predicted probabilities to actual class output which can be either 0 or 1. It then calculates the score that penalizes the probabilities based on the distance from the expected value.

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \cdot log(p(y_i)) + (1 - y_i) \cdot log(1 - p(y_i))$$

Fig. 4. Binary Cross Entropy Loss

III. Solution Approach

Since the labels were one hot encoded, in order to test our dataset correctly, we had to transform the model outputs which were in their corresponding probability scores to one hot encoded values (1's and 0's). Therefore we decided to go with the top three probability scores in each sample, convert those values to 1 and the remaining to 0's and and predict the corresponding three genres.

IV. Experimental Results and Performance Evaluation

A. Results

Our model was able to predict the genres of the test dataset with an accuracy of only 21-23 %.

```
TRAIN LOSS:  0.2542370557785034
Epoch [0], val_loss: 0.2555, val_acc: 21.9061
TRAIN LOSS:  0.2609313726425171
Epoch [1], val_loss: 0.2506, val_acc: 21.9506
TRAIN LOSS:  0.2470485270023346
Epoch [2], val_loss: 0.2505, val_acc: 21.9108
TRAIN LOSS:  0.23429888486862183
Epoch [3], val_loss: 0.2481, val_acc: 21.9342
TRAIN LOSS:  0.24423915147781372
Epoch [4], val_loss: 0.2467, val_acc: 21.9310
```

Fig. 5. Accuracy Result Of Test Dataset

We were able to observe that a Deep Neural Network model is not efficient for our dataset and a further comparison with other models is necessary.

```
PREDICTED:

Comedy
Drama
Romance
```

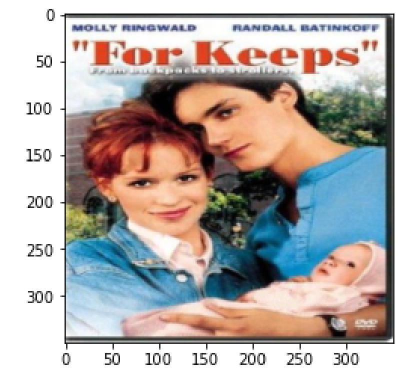

Fig. 6. Predicting Genres of New Poster