Movie Genre Classification from Posters using Convolutional Neural Networks

Akansha Agarwal, Darakshan Anwar and Rishabh Garg
Department Of Computer Science And Engineering
Texas A&M University

***Abstract*—Movie posters have a major influence in depicting the content of the movie. A good movie poster can attract a lot of viewers if it illustrates the content well. It is crucial for the success of a movie that particular genre(s) admirers get motivated to watch the movie by just looking at its poster. In this study, we perform movie genre classification from movie posters using Convolutional Neural Network.The idea is to feed pre-processed movie posters to the network.This pre-processing is done using two techniques,by taking center area of each poster image and by taking area containing the detected objects in the poster.**

In this paper, poster classification is modeled as a multi-label classification task, where a single movie may belong to more than one class (genre).Three metrics for multi-label prediction evaluation are described given the classification results obtained from Convolutional Neural Network.

***Index Terms*—Movie poster, Convolutional Neural networks, Multi-label classification , Object detection**

I. INTRODUCTION

People often judge a book by its cover and same is the case for movie posters. People generally grasp certain information about the movie by looking at a movie poster and often judge whether to watch a movie based on their initial impression of a movie poster. Thus, movie producers often try to make movie posters which resembles to nature of the movie so that the poster can grab attention and attract audience. Hence, movie posters can be an important element in labeling the genres associated with it.

Moreover, movie information sites such as IMBD, Rotten Tomato maintains a huge database of movies and movie reviews and often classify and recommend movies based on genres. However, as genre labelling for movies is based on user suggestions, this process is prone to error. To make this more accurate, it can be automated using efficient machine learning models that is far more accurate.

Also it can be assumed that if a person can quickly grasp the genre from a movie poster regardless of level of details,then the poster possesses some characteristics which could be utilized in machine learning algorithms to predict its genre.

Rest of the paper is organized as follows: Related work is reviewed in section II. Section III, describes the proposed methodology of our work.Implementation and experimental results are given in section IV. Section V and VI include acknowledgment and conclusion for the paper.

II. RELATED WORK

Solutions proposed for movie genre classification range from linear models to the advanced deep learning approaches. Some work attempts to perform genre classification based on poster image or video data which would often demand deep convolutional neural network based solutions. [1] [2]

Several works discusses use of movie posters for genre classification. One of the work on genre detection using movie posters demonstrates the extraction of semantic features like Theme, Emotion, Nationality, Composition, Layout, Number of males and females, Background focus, dominant colour and Background color from the movie poster using Neural Network and combining these features for predicting the genre. [3] Another research emphasizes the idea that colors present in a poster are essential in determining its genre ,for example dark colors may corresponds to the 'horror' genre. So, features like colour and edge can be extracted from the movie poster and thus can be used for classifying the movie into different genres [4].Another related work discusses the use of object detection to extract object information combined with visual representation feature extraction from the poster using deep neural network for classifying the movie posters. [5]

In this work, we explore how to improve the performance of movie genre classifier with efficient data-processing and feature extraction on movie poster images. We aim to process the important areas of the poster image which contains most useful information, like center of the poster or areas that consist objects. And feed these relevant processed data to the convolutional neural network. This kind of data pre-processing extracts important features from the image and helps neural network to perform better.

III. METHODS

We introduce the construction of deep learning classifier for movie genre classification using original movie poster images and our proposed approaches for improvement of the classifier with efficient feature extraction from the poster image.

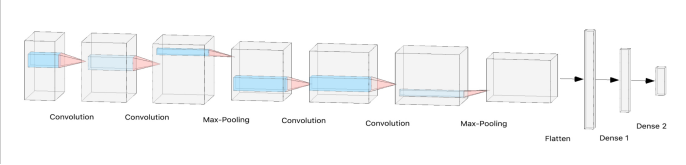

Fig. 1. Convolutional Neural Network Architecture

A. Approach 1: Convolutional Neural Network based Genre Classfication using Original Poster Images

We construct a deep learning genre classifier as follows. Our model takes original poster image as input and applies several layers of convolution and max pooling, followed by a fully connected layer.

Overall architecture of our model is outlined in figure 1. Model takes an input of size [150,101,3] and applies two 3-D convolution with 32 and 64 filters respectively and both with stride 1. Then it applies a batch normalisation and a ReLU non-linearity after each convolutional layer. After second convolutional layer, Max pooling with stride 2 is applied followed by a dropout with possibility 0.5 to the pooled representation. Again , two 3-D Convolution each with 64 filters and stride 1 are applied, followed by batch normalisation and ReLU activation function. Afterwards, there is a flatten layer followed by two dense layers of size 128 and 10 respectively. Size of last dense layer is determined by number of genre classes in the dataset. The activation for the last layer is sigmoid in order to deal with multi-label classification problem.For training we optimize the Cross Entropy loss function using the SGD optimizer with a learning rate of 0.001.

B. Approach 2: Genre classfication using centre of poster images

Most of the movie posters have relevant information in the center of the image. Instead of giving whole image to the model,we intend to improve the performance of genre classifier by extracting the important portion from the movie poster and feeding only the extracted portion to the network. Here, we aim to extract only the center portion of the image.

In order to extract the center portion of the image, we crop the original image by leaving one-eighth margin from all the sides. This extracted image is then feed into the same convolutional network outlined in figure 2, Approach 2.

C. Approach 3: Genre classfication using object detection on poster images

In order to improve the performance of the classifier further, we intend to extract object information from the poster images and feed only the portion of poster which contains the detected objects into the network. Main idea behind this approach is that most of the information related to movie genre are captured in the objects present in the poster and if instead of giving whole poster as input,the portion

Fig. 2. Convolutional Neural Network Architecture

of poster containing the objects is processed and extracted to feed directly into the model, our model learns in a more efficient way.

In order to extract the object information,object detection is performed on each movie poster and then area formed from union of bounding boxes for each detected object in the poster is cropped and given as input to the convolutional neural network outlined in figure 2, Approach 3.

Figure 2 displays the above mentioned approaches.In the first approach, whole poster image is used as input. In the second approach, center portion of the poster is cropped and used as input. In the last approach, using object detection all men in the poster are detected and so the area covering all the detected men is cropped and used as input for the model.

D. Prediction Evaluation for Multi-label classification

The output of the network [2] is an M-dimensional probability vector, where each dimension indicates how likely a given poster belongs to a movie genre. Each movie can belong to multiple genres and number of genre for each movie also varies which makes this a multi-label classification problem.

We explore three methods to solve this multi-label classification problem and corresponding metrics to evaluate the performance of the model.

1) Metric 1: This method uses a concept of thresholding i.e If the value of the ith dimension of output probability vector is larger than a particular threshold, we say the given poster belongs to the ith movie genre. A threshold of 0.5 is taken here. After thresholding, different performances metrics like Accuracy, Precision, Recall and F1 are calculated in a conventional way.

2) Metric 2: Similar to Metric 1 above, this method also uses a concept of thresholding over output probability vector. But here in order to calculate performance metrics,the partial correctness(weighted) in multi-label classification is taken

into consideration.

3) Metric 3: In this method, thresholding is not done.Instead, top k genres from the output vector with highest probability are taken and evaluated, where k would vary and would be equal to the total number of genres a particular movie has.

IV. EXPERIMENTS AND RESULT

A. Dataset

Used the 'IMDB dataset' that contains movie's name, IMDB link(which is used for web-scraping) and its poster link(for some movies). Posters are retrieved by downloading the posters from IMDB website using the links provided in the dataset. Since a movie may belong to multiple genres, this is a **multi-label** image classification problem. This implies that a movie can be classified into more than 1 movie genres. To transform the target Movie Genres column, which initially contains the names of the movie genres separated by the $'|'$ delimiter, we created the target genres representation for true labels using one-hot encoding. This is done in such a way that target label(genre) vector consists '1' in a position if that particular genre is present otherwise it contains '0'.

B. Preprocessing the dataset

The IMDB dataset consists of 27 movie genres. Due to the limited availability of resources, we constrained our implementation into 10 movie genres. For doing this we first retrieved the total number of movies belonging to each genre, then amongst the top 15, using [8] as reference, we chose the most popular genres and finally proceeded with 10 movie genres.

After this, we proportioned the data using sampling. Since the dataset was large and resources were limited, instead of over sampling, we proceeded with under sampling. We initially retrieved the counts of movies belonging to each genre, and then chose 1000 movies for each of the genres. Now because each movie can have multiple genres, the final count for each genre with resulting collection of posters after under sampling might not be equal. Finally we had approximately 9000 samples of which 10 percent are used for testing and 90 percent are used for training purposes.

Further, we moved ahead by downloading all the images. Now, as we used 3 methodologies, we created separate datasets for Approach 2 and Approach 3 using the extracted posters. For Approach 2, we cropped the downloaded images by leaving one-eighth margin from all the sides, in order to take the center portion of the image. For Approach 3, we first applied object detection to the image to retrieve the bounding boxes for all the objects in the image, and then we took the union of all the resulting bounding box dimensions to extract an image containing all the objects. If no object is detected for an image, then the center portion of that image is taken for that particular poster.

This was followed by resizing the images to size 150x101 and normalising them.

C. Hyperparameter Tuning

We chose different batch size and epoch combinations for tuning our model. We used different values of epochs equal to 25, 50, 75 and 100, and values of batch sizes equal to 10 and 20 to determine the best parameters for each approach. As we increase the values of the epochs the training improvises, but after a certain threshold the model tends to overfit. When overfitting occurs, training accuracy begins to increase but testing accuracy decreases. In neural networks, small batch sizes are usually preferred for large datasets, but since the data set we used is not very large we could achieve comparable results with batch size 10 as well as 20.

D. Evaluation Matrices

Since there is no definite and pre-defined method to get prediction scores for multi label classification problems, we used 3 different approaches to determine the performance of our models.

In the first method, we calculated the accuracy, precision, recall and f1 score using the confusion matrix formula. The predictions from convolutional neural network is a list of 10 probabilities, where each probability corresponds to a particular genre. We used a threshold of 0.5 to calculate true positive, true negative, false positive and false negative for each of these predictions.

In the second method, we used partial correctness to achieve the accuracy, precision, recall and f1 score. This method is also called Hamming distance. For determining the accuracy, the ratio of intersection of the predicted and actual labels with the union of the predicted and actual labels is calculated. Similarly for precision, the ratio of intersection of the predicted and actual labels with the number of positive predictions is calculated. For recall, the ratio of intersection of the predicted and actual labels with the number of positive actual labels is calculated. Lastly, F1 score is calculated by taking the harmonic mean of precision and recall. Threshold of 0.5 in this method to convert probabilities retrieved using sigmoid activation function to 0 or 1.

In the third method, we calculated the percentage correctness of the predicted labels as a accuracy measure. In this, if an image has k genres, we extract the top k predicted genres and count the number of correct predictions. The total number of correct predictions calculated is divided by the total number of actual genres to get the percentage of correctly predicted labels.

E. Results

We have 3 approaches for predicting the genres of a movie. And for each of the approaches we used 3 different metrics

to calculate the result.

Using the normal images as input to the Convolutional Neural Network, we achieved an accuracy score of 0.75 using method 1, weighted accuracy of 0.18 using method 2 and percentage correctness of 0.26 using method 3.

Using the center image area as input to the Convolutional Neural Network, we achieved an accuracy score of 0.66 using method 1, weighted accuracy of 0.23 using method 2 and percentage correctness of 0.28 using method 3.

Using the union of object bounding box areas from the images as input to the Convolutional Neural Network, we achieved an accuracy score of 0.74 using method 1, weighted accuracy of 0.22 using method 2 and percentage correctness of 0.34 using method 3.

We achieved percentage age correctness of 0.26 in approach 1, 0.28 in approach 2 and 0.34 in approach 3.

Since, the percentage correctness appears to be the most logical metric since it doesn't involve any thresholding and uses the top k probabilities for corresponding k genres of a movie. We see that we achieved improvement in our results when we progressed from approach 1 to approach 2. For approach 1, approximately 26 percent of the genres were correctly predicted while for we observe a 2 percent increase in correct prediction of genres using approach 2, with 28 percent correct genre predictions. In approach 3 we achieve 34 percent correctly predicted genres, which is 8 percent increase from approach 1 and 6 percent increase from approach 2.

The percent increase in correct prediction of genres can be seen as a significant improvement, since the total number of genres are in thousands.

F. Tables And Figures

a) Result Tables: Below are the result tables, for calculating correctness of our predicted result using 3 different metrics. Table 1 represents accuracy measures obtained using metric 1. Table 2 represents accuracy measures obtained using metric 2. Table 3 represents percentage correctness of predicted genres using metric 3.

TABLE I
METRIC 1

Method	Batch	Epochs	Accuracy	Precision	Recall	F1
1	10	100	0.75	0.40	0.17	0.24
1	10	50	0.68	0.22	0.17	0.19
2	10	100	0.66	0.22	0.20	0.21
2	10	75	0.62	0.20	0.23	0.22
2	10	50	0.64	0.16	0.14	0.15
3	10	100	0.74	0.36	0.21	0.26
3	20	100	0.73	0.33	0.22	0.26
3	10	75	0.59	0.24	0.37	0.29

TABLE II
METRIC 2

Method	Batch	Epochs	Accuracy	Precision	Recall	F1
1	10	100	0.18	0.38	0.18	0.25
1	10	50	0.21	0.23	0.72	0.35
2	10	100	0.23	0.23	0.86	0.37
2	10	75	0.22	0.22	0.96	0.36
2	10	50	0.14	0.18	0.24	0.20
3	10	100	0.22	0.22	0.99	0.36
3	20	100	0.22	0.22	0.99	0.36
3	10	75	0.21	0.25	0.50	0.33

TABLE III
METRIC 3

Method	Batch	Epochs	Percentage correct
1	10	100	0.24
1	10	50	0.26
2	10	100	0.25
2	10	75	0.28
2	10	50	0.18
3	10	100	0.34
3	20	100	0.33
3	10	75	0.20

b) Prediction Examples: Figure 3 represents the actual and predicted genres by our model for some of the posters.

Poster	Actual Genres	Predicted Genres – Approach 3
	Action, Thriller	Action, Drama
	Comedy, Drama, Romance	Comedy, Drama, Adventure
	Sci-Fi	Horror

Fig. 3. Actual vs Predicted Genres

V. ACKNOWLEDGEMENT

We would like to thank Dr. Bobak Mortazavi for providing us with the opportunity to do research on such an interesting topic as part of our project. Also, we appreciate the guidance provided by Arash Pakbin in successful completion of the project.

VI. CONCLUSIONS AND FUTURE WORK

As most of the important information regarding movie poster would be present in the center portion rather than the borders, it can be used to improve our model. To further improve our prediction, object detection is used to extract relevant features from the posters to obtain efficient results.

To further improve our model, semantic feature extraction can be combined with our methodology to obtain better genre classification.

VII. LIMITATIONS AND IMPROVEMENTS

Limitations: Since movie posters can be complex and sometimes vague, recognition of a movie's genre by its poster gets difficult.

Improvement: Model performance can further be enhanced by using various known efficient architectures such as VG16, Google Net for neural network.

REFERENCES

[1] Z. Rasheed and M. Shah,"Movie genre classification by exploiting audio-visual features of previews"

[2] Sanjay K. Jain ; R.S. Jadon, "Movies genres classifier using neural network"

[3] Sorratat Sirattanajakarin and Panita Thusaranon, "Movie Genre in Multi-label Classification Using Semantic Extraction from Only Movie Poster".

[4] Marina Ivasic-Kos, Miran Pobar and Luka Mikec, "Movie Posters Classification into Genres Based on Low-level Features".

[5] Wei-Ta Chu and Hung-Jui Guo, "Movie Genre Classification based on Poster Images with Deep Neural Networks".

[6] https://github.com/davideiacobs/-Movie-Genres-Classification-from-their-Poster-Image-using-CNNs.

[7] Sean Maxfield, "ANALYZING THE MOVIE-VIEWING AUDIENCE "

[8] https://www.statista.com/statistics/188658/movie-genres-in-north-america-by-box-office-revenue-since-1995/