Predicting Media Memorability

Nikhil Mittal
Data Analytics
19210509
nikhil.mittal2@mail.dcu.ie

***Abstract*-In this paper we will be predicting the media memorability of video i.e. a predicting task for our machine learning task. The task is about to get the score as to what is the probability of remembering a video by using machine learning models. We apply various models like Neural network, Random forest, etc. to get the best model and get the short- term and long-term memorability score.**

Keywords: -Neural Network, Random Forest

I. INTRODUCTION

Media memorability is something that tells how much we video someone can remember after seeing it for one time. So predicting media memorability is getting a score of each media based on using machine learning techniques.[4]

Firstly, we load the data from the drive where data is being stored and import necessary libraries. Then we load what features are to be tested for the memorability so that they are loaded. Spearman score function is declared and loaded into the function.

Then we apply various Machine learning models on the various features and get the spearman score. The best model with those features is then applied to the test set and eventually getting the short-term and long term-term memorability score

II. RESEARCH

There are some works done on media memorability done and the better part was to look at them as to understand the task for which it was being designed and get a better understanding of the problem.

There is a paper that is using visual and semantic features on another type of dataset and predicting image memorability using CNN. The output is very good and is outperforming the originally predicted model genericity of the model is also taken out. [1]

Another paper uses deep visual features and recurrent networks to predict video memorability. Features extracted from CNN from several frames are put into the LSTM network to model the structure and predict memorability. [2]

Lastly, there are several features like aesthetics, color, etc were used and models used were SVR and CNN with LSTM. CNN with LSTM was the most useful case and spearman score was better than SVR.[3]

III. APPROACH

Algorithms used are Random forest and Neural network and also created an ensemble of these models using the weighted average method. All models were subjected to the features that were concatenated into one and there is no separate prediction and score that is taken out for a particular feature.

IV. FEATURE EXTRACTION

There are three features used in this, and all of these features have been used as a result of previous research work on this.The choice of features is based on research. These features contain information about a constant frame, or the motion of an object, rather than a single frame. In order to avoid overfitting, all frames should be considered or taken into account in some cases, the appearance of objects is too quick and too short. In that case, it's difficult to predict memorability even though the video was frightening. Here are the combinations of features that are being used..

Captions: The textual feature describing the video. The subscriptions are created manually. The captions are the features of objects in videos.
C3D: C3D is obtained through the training of a deep 3D convolutionary network. It contains different concepts, including objects, actions, and scenes.
HMP: histogram of motion patterns. They encode the global
ALL three were combined when using it on the models for both the dev-set and the Test-Set

V. MODEL DESCRIPTION

Two algorithms are being used: Random Forest and Neural Network. There is also an ensemble method used using a weighted average using the predictions of the random forest and neural network to get a better result. The problem that it is a classification type problem so these models were used for the prediction.

The dataset splitting was done at a ratio of 80:20 for both models. The maximum depth of the decision tree is set at 10 for the random forest. Keras is used for the implementation of neural networks. There are three layers, each of which is set to use the Scaled Exponential Linear Unit (Selu) as an activation. Other activation functions, such as soft plus or soft sign, significantly reduce the accuracy of the validation and the score of the spearman. The size of the output shape of the layers is 30, 25, 2. If lower accuracy is set, start lowering this parameter as if you were going down the layers to assume

memorability exactly.. For the same purpose, the dropout is smaller between the first and the second layers. Adam Optimizer is used in the neural network.

VI. ANALYSIS OF THE RESULTS

On applying all the features on thr model random forest the spearman score is as below:

```
GET_SCORE(rfpredictions,Y_test)
```

```
The Spearmans correlation coefficient is: 0.300
The Spearmans correlation coefficient is: 0.141
```

Figure:1

Neural network was best in the terms of score and the graph accuracy while using the dev set for the same was also good which made it obvious to use the neural network on in test set.

Below results for dev-set:

```
GET_SCORE(nnpredictions, Y_test)
```

```
The Spearmans correlation coefficient is: 0.319
The Spearmans correlation coefficient is: 0.187
```

Figure:2

While using the ensemble using the weighted average of the two models .i.e random forest and neural network the scores become less as can be seen.

```
predictions_test = npy.divide(predictions_test, 2)
print(predictions_test)
```

```
[[0.42805293 0.38809542]
 [0.43112751 0.3906117 ]
 [0.44331048 0.42327931]
 ...
 [0.42834106 0.39190222]
 [0.41419602 0.35138929]
 [0.4308663  0.38073851]]
```

Figure: 3

While using the same model on test-set the score gets better.

```
predictionnn_test = model_test.predict(X_test)
print(predictionnn_test)
```

```
[[0.85105133 0.7683709 ]
 [0.7918561  0.75290555]
 [0.8995825  0.83575606]
 ...
 [0.8783476  0.7972223 ]
 [0.91141295 0.8647998 ]
 [0.883858   0.7973548 ]]
```

Figure:4

VII. CONCLUSION

It can be concluded that short-term memorability is higher than long-term memorability. Using the combination of features, the score of the models is lower as data is large. Neural network is best for memorability and can be used in different aspects to increase the memorability score.

VIII. REFERENCES

1] H. Squalli-Houssaini, N. Q. K. Duong, M. Gwenaelle and C. Demarty, "Deep Learning for Predicting Image Memorability," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, 2018, pp. 2371-2375.

[2] Tran-Van, D., Tran, L., & Tran, M. (2018). Predicting Media Memorability Using Deep Features and Recurrent Network. MediaEval.

[3] Joshi, T., Sivaprasad, S., Bhat, S. and Pedanekar, N., 2018. Multimodal Approach to Predicting Media Memorability. In MediaEval.

[4] MediaEval Benchmarking Initiative for Multimedia Evaluation

(http://www.multimediaeval.org/mediaeval2018/memorability/)

[5] Spearman's Rank-Order Correlation

(https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php)