Predicting Media Memorability Using Machine Learning Models

Aditya Gupta
School of Computing
Dublin City University
Dublin, Ireland
aditya.gupta4@mail.dcu.ie

***Abstract*—Memorability is defined as the state of being easy to remember or worth remembering [1]. Nowadays with an increasing number of users in social media platforms, there are more than billions of videos that are viewed by us and when we see these videos it gets pivoted and focussed in our mind. Through these videos, our visual power is more grasping than the usual readings or listenings. Memorability thus rotates or oscillates in our mind to capture the attention of our brain clocks, as certain features have a sharp impact on our memorability. But now the question arises that how memorable these videos are to us? Or Are these videos worth remembering? Or Is there any feature which can predict the memorability. Therefore in this project, I have implemented various features with different algorithms in predicting the short-term and long-term memorability scores.**

***Index Terms*—Linear Regression, Decision Tree, Random Forest, Captions, HMP, InceptionV3, Sequential Neural Network**

I. Introduction

In this paper, we investigated the Media Memorability scores and further predicted the short-term and long-term scores based on the various features and using different algorithms. As part of the MediaEval Media Memorability challenge 2018, we were given short videos including different descriptive features to predict memorability such as InceptionV3, C3D, HMP, ColorHistogram, and so on. Initially, I trained the models separately with all the given features but then later chose Captions, HMP, and Inception to predict the memorability as previous work has also shown us that using captions and HMP yields good results as compared to other features like ColorHistogram or C3D. Therein, I trained various models using different algorithms to predict the memorability score. The models are then evaluated using Spearman Correlational Coefficient.

Based on the analysis it can be said that:

- Captions yield better results than any other features available and whereas Inception being the worst.
- Short-term memorability scores were more accurate than the long-term memorability scores.
- If we increase the size of the test-set in the prediction the performance will go down and will become more sturdy.

This paper will further explain the analysis and process for achieving short-term and long-term memorability scores.

II. Literature Review

In the past years, a vast amount of research work has been done in predicting video memorability. The winner of MediaEval 2018 competition R. Gupta et al. 2018 [2] trained the model by combining semantic features and visual features and then predicted the memorability scores. They showed in their paper that using the C3D video feature and HMP has outperformed all the other features such as AestheticFeature, ColorHistogram, and LBP. In addition, they have used captions in building their model and the final conclusion they gave was that words related to nature had negative coefficients and words related to humans had positive coefficients. Therefore, I explore more on captions features and certain bags of words.

III. Feature Extraction

In the dataset, we had various descriptive features available to predict the memorability of short videos like captions, C3D, HMP, Aesthetic features, etc. I used Captions, HMP, and InceptionV3 to predict short-term and long-term memorability. All three were individually used to predict the memorability.

The feature extraction of captions was performed using Natural Language Toolkit Library and some defined methods. Then cleaned the dataset by removing punctuation marks and stopwords from captions such as at, any, before, them, etc. Used bag of words for the captions and should be a 2D array.

The HMP features were transformed into data frames and then merging ground truth with it. These were read into frames and sent as an independent variable and further joined on video names but yielded us unsatisfactory results.

Inception features, I implemented a deep learning model to predict the short-term and long-term memorability score. Therein I merged the dataset with the same video names and finally converted the features into NumPy arrays as our model will only accept NumPy arrays but this feature also gave us poor results.

IV. Model Exploration

I simply preferred linear regression algorithms for the prediction of memorability scores and ran different algorithms on different features like:

- For Captions, I used Random Forest Regression algorithm and used Support Vector Regression algorithm.
- For HMP, I used Random Forest and Decision Tree Regression algorithm.
- For Inception, I used Sequential Neural network algorithm.

V. Explanation

In this project, I used Colab Notebook[3] because it offers us an environment that requires no setup to use and also it runs completely on the cloud. I mounted my Google Drive in the notebook to acquire the data provided for the computation. In the dataset, various extracted features were present like C3D, HMP, Caption, etc. After this, I imported all the required libraries and installed pyprind which is a module that enables you to visualize the progress of various tasks in Python. Hence, I used Captions, Inceptions, and HMP as a feature to predict memorability scores.

When using the Inception feature, I defined a function to read the file then merged the dataset with the same video names and with ground_truth. For pre-processing, I converted features into NumPy arrays as our model will merely accept NumPy arrays. After performing this I split the dataset into a train set and test set then finally trained the model using Sequential Neural Network for 20 epochs and compiled it through 'adam' optimizer and metrics accuracy. I used EarlyStopping with early_stopping_monitor to prevent overfitting.

Underneath is the visualization when we trained the model for 20 epochs which during the training phase fits the data.

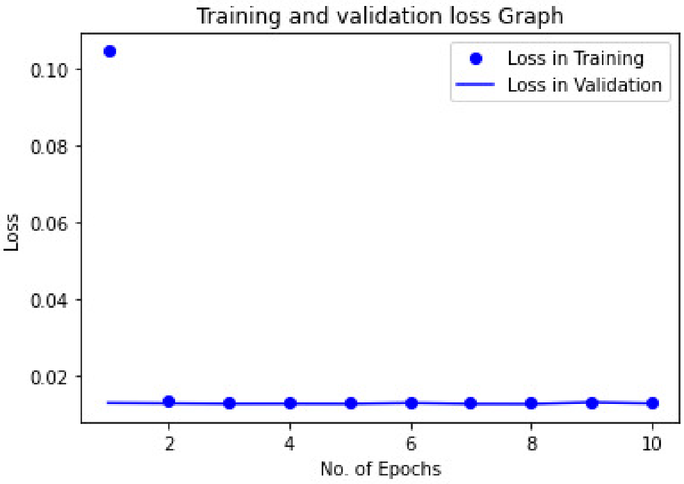


Fig. 1. Training and validation loss Graph

Lastly, we calculated the Spearman's Correlation Coefficient and it yields very poor results of both short-term and long-term memorability.

When using the Captions feature, I defined a function to load the captions and to preprocess the dataset. And next providing the path to the dataset by loading ground_truth and captions from the drive folder. The essential thing was to clean the dataset in which I removed punctuation marks and stopwords like at, is, are, etc. from the dataset and replaced the punctuation marks with the white spaces, and also I converted all the words(if any) to lowercase. Subsequently, I used CountVectorizer which provides a way to tokenize a collection of text documents and builds a vocabulary of known words. Ultimately split the dataset into the training set and test set and then after splitting the dataset I implemented two algorithms Random Forest regression and Support Vector Regression to train the model and it was found out that Spearman's Correlation Coefficient score was better when we used Random Forest algorithm then Support vector Regression algorithm.

Underneath is the scatter plot of predicted to actual results when I train the model using the Random Forest regression algorithm.

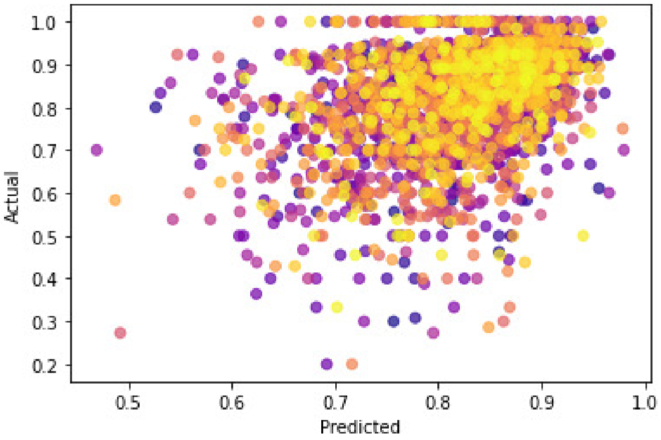


Fig. 2. Feature Captions Using Random Forest Model

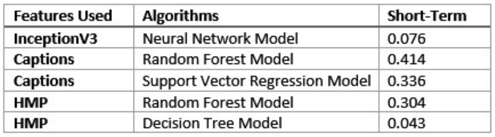
When using the HMP feature, I defined the function to load the HMP feature and to preprocess the dataset and then transformed the feature into a dataframe. We then load groundand drop the unnecessary columns such as 'nb_short-term_annotations' and 'nb_long-term_annotations' and at last merge the ground_truth and HMP features into a dataframe. Ultimately, split the dataset into a training set and test set. In this feature I implemented two different algorithms to train the model first is Random Forest Regression and the other is the Decision Tree Regression algorithm. In Random Forest I used n_estimators = 100 which produces us the good results in our analysis but in case of using the Decision Tree Regression algorithm, it very yields poor results.

After training all the models, it was found out that the captions feature when used with Random Forest algorithm gives us the most outstanding results so at last, I predicted the 2000 short videos with the same model which gave us the best spearman's coefficient score and lastly saved the results to a CSV file.

VI. Results

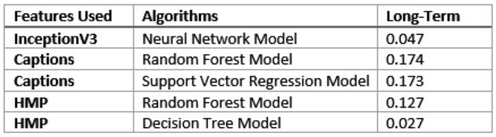
Results were calculated using Spearman's Correlation Coefficient which is a nonparametric measure of rank correlation. During the evaluation of models, we found out that captions, when used with the Random Forest Regression algorithm, outperformed all the other features. And it can on top be noted that HMP when used with the Random Forest Regression model also yielded good results but not better than Captions. From the table, it is also understandable that InceptionV3 when used with a sequential neural network algorithm and HMP when used with the Decision Tree algorithm yields very poor results.

TABLE I
SHORT-TERM MEMORABILITY



Features Used	Algorithms	Short-Term
InceptionV3	Neural Network Model	0.076
Captions	Random Forest Model	0.414
Captions	Support Vector Regression Model	0.336
HMP	Random Forest Model	0.304
HMP	Decision Tree Model	0.043

TABLE II
LONG-TERM MEMORABILITY



Features Used	Algorithms	Long-Term
InceptionV3	Neural Network Model	0.047
Captions	Random Forest Model	0.174
Captions	Support Vector Regression Model	0.173
HMP	Random Forest Model	0.127
HMP	Decision Tree Model	0.027

VII. ANALYSIS AND DISCUSSION

The Below Fig 3. and Fig 4. depicts the short-term and long-term memorability score when various features were used with different algorithms.

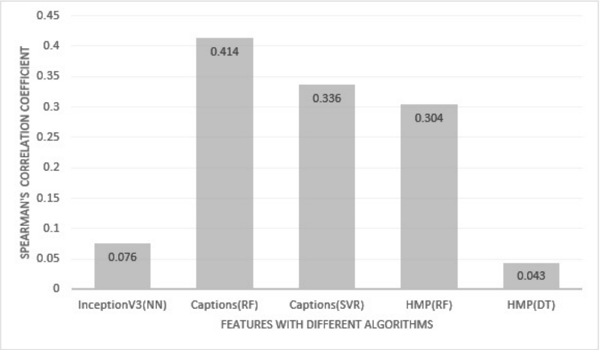


Fig. 3. Short-term Memorability scores.

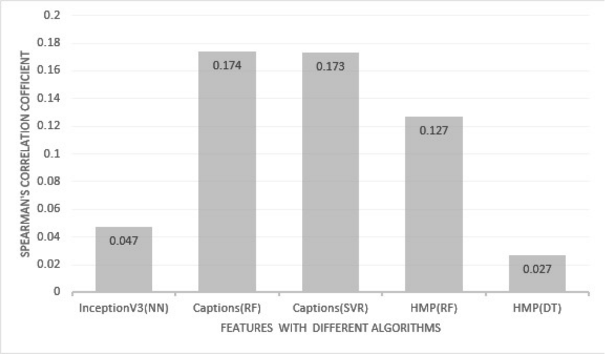


Fig. 4. Long-term Memorability scores.

These visualizations describe us that captions, when used with the Random Forest algorithm, have outperformed all the other features. The vital thing which is to be noted is when we used captions with the Support Vector Regression algorithm it also yielded us the impressive results. Moreover, we can on top see that the InceptionV3 when used with a sequential neural network algorithm and HMP when used with the Decision Tree algorithm yielded us the poor results.

CONCLUSION AND FUTURE WORKS

The overall analysis shows us that captions provided us better short-term and long-term memorability scores when used as a feature with the Random Forest Regression Algorithm. Also, it can be said that short-term predictions are more precise than long-term predictions. Captions feature when used with Support Vector Regression Model also yields us good results but it was not better than the Random Forest Regression algorithm.

When we see the graph, InceptionV3 with Sequential Neural Network model, and HMP when used with the Decision Tree Regression model seems to be a poor choice with very low short-term and long-term memorability scores whereas HMP feature, when used with Random forest Regression algorithm, gives us significantly better results. However, to other researchers, combining ResNet and captions fetched better results so further improvements we would like to have are Pre-extracted ResNet features that can help us to measure the impact on memorability scores.

For future work, I would like to work on Logistic Regression and ResNet as it is a pre-trained model that was built to perform well on image features.

REFERENCES

[1]”Merriam Webster,” [Online]. Available: https://www.merriamwebster.com/dictionary/memorability. [Accessed 2019].
[2]R. Gupta, ”Linear Models for Video Memorability Prediction using Visual and Semantic Features,” MediaEval, 2018.
[3]Google Colab https://colab.research.google.com/