Predicting Video Memorability Using Features for MediaEval 2018

Anubhav Gupta
School of Computing
Dublin City University
Dublin, Ireland
anubhav.gupta5@mail.dcu.ie

***Abstract*—With the recent survey on Internet, it is proved that video content has been increasing with video sharing, which in turn is increasing the competition among different creators. Research and analysis on videos, therefore, has become important and many tries have been done before. Memorability is one aspect of videos that can help in competing the metric of different videos but is not been discovered in the past. In this paper, I will be computing the memorability of the clips provided in the MediaEval Predicting Media Memorability Task, 2018. I have used various features and applied various algorithms to increase the accurately predict the short- and long-term memorability.**

***Index Terms*—Captions, Linear Regression, Decision Trees, C3D, Random Forest, Gradient Boosting, Bayesian Ridge**

I. Introduction

This paper is related to the MediaEval 2018- predicting memorability of 6000 short videos using different descriptive features. Combination of different features like Captions and C3D (Convolutional 3D) which were pre-processed features were used against different algorithms like linear Regressor, Decision trees, Random Forest, Bayesian Ridge and Gradient Boosting Regressor deep learning algorithm to produce the best scores in achieving memorability. MediaEval Predicting Media Memorability Task was presented to generate automatic prediction of the memorability of videos as it has various use cases like advertising, education, search, learning, etc. The data consists of 6000 videos with different features to train the model on and further 2000 videos are there to test the algorithm on. In this paper will be the final result for my approach and the paper will further explain the process of achieving short- and long-term memorability.

II. Literature Review

All the work done in the past in this prediction was done focused on C3D feature with different other features like captions and HMP. I tried to follow the same path and used Eoin Brophy's given code for estimating Spearman rank correlation coefficient and to extract and preprocess the features like captions and C3D. Then applied number of different algorithms to get better scores. I learnt from the contest results at MediaEval'18 that teams who used 2 or more than 2 features resulted in better scores also by researching different papers in this area I learnt how to approach in the process of prediction with different features combined.

III. Background Modelling

A. Feature Exploration

The key component in this project was to find out which of the features to be used in the model. I researched a lot on the previous work in the field done and also the teams that took part in the competition. Every team who used two or more than two features got a better spearman coefficient and therefore I tried to choose atleast 2 features. Reason in choosing captions and C3D features was because they were directly extracted from the videos and it was really efficient to compute both of the features. At first I tried using regression models on them individually and results were relatively low so, this made me think of combining the features and try some of the models over them to find the better solution.

B. Model Exploration

I have used different regression algorithms like Linear Regression, Decision Tree Regression, Random Forest Regression, Bayesian Ridge and Gradient boosting Regression on captions and C3D to get better results in the spearman coefficient. I used simple models so as to maintain overfitting as the features were of high dimensionality and collinearity.

- Linear Regression
- Decision Trees
- Random Forest Regression
- Bayesian Ridge
- Gradient Boosting Regression

For each model I used captions sequences and C3D merged them together for every model, explored the spearman coefficient based on the Eoin Brophy's code and compared the models to predict the memorability of the the videos on the basis of the spearman correlation coefficient which can bee seen in the table in the results section.

IV. Explanation

My project has been worked on the Google Colaboratory to run my python code and to mount my drive to access the data provided for the computation. Many of the pre-computed features like C3D, HMP, captions were available. I used captions and C3D because they can be easily merged and computed, also i researched a lot of erlier work and the teams at MediaEval also got better results with 2 features combined. I pulled captions and C3D and the captions were preprocessed using tokenization to count the words and splitting the caption. The captions and C3d Were merged and using pandas, then I used the 5 models to check the spearman rank. As in the result section shows, Random Forest and Gradient Boosting Regressor were showing the best results. Both the model were trained on 500 decison trees (n estimators). However the learning rate for gradient boosting regressor was 0.02. The model was splitted into 90 percent train data and 10 percent train data which was splitted in a 5400-600 videos.The results of which is shown in Table 1. Then 2000 videos were tested which was the main goal.

V. Results

As followed in the python notebook attached with the report, various of the regression algorithms were used and applied on captions and C3D feture to obtain the spearman coefficient. As in the table below , the comparison has been shown with the results obtained with the data.

TABLE I
SPEARMAN COEFFICIENT

Model	*Short Term*	*Long Term*
Linear Regression	0.187	0.047
Decision Tree	0.007	-0.007
Random Forest	0.264	0.132
Bayesian Ridge	0.222	0.088
Gradient Booster	0.289	0.122

As per the table the results for short-term were obtained by Gradient Boosting Regression while for the long-term were obtained by the Random Forest. So I decided to move with the Gradient Boosting Regression and one is noticed here is all the algorithms score better in their short-term memorability predictions rather than the long-term.

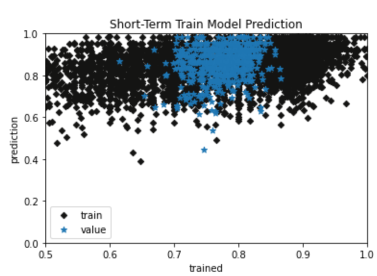

Fig. 1. Short Term Memorability Graph

Figure 1 is the scattered graph that is based on the no. of predictions vs trained values for the Short-Term memorability of the videos. Figure 2 depicts the long -term memorability

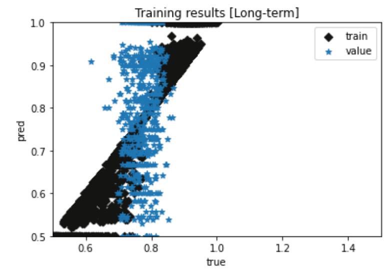

Fig. 2. Long Term Memorability Graph

of the predicted vs trained values. The short term memorability graph depicts that the predictibility of the short term memorability was higher that the long term memorability.

VI. Conclusion

In conclusion, this project was the basis to learn different algorithms and methods to implement a machine learning project in practical and to dive in the practicality of importing, preprocessing, training and testing the machine learning algorithm to obtain the results. The results that I achieved will not be the best as many improvements are possible like adding HMP feature may improve the score and having no prior knowledge of machine learning in practical, I am happy with the results achieved. I am now pretty confident in working with Machine learning algorithms and as my future work will try to apply new algorithms and use HMP feature and try to increase the score for long term memorability aswell.

References

[1] Brophy, E. (2018). Google Colaboratory.
`https://colab.research.google.com/drive/1X7l5MGrDZa2IdMCOxwg`

[2] Machine learning Algorithms
`https://scikit-learn.org/sklearn.ensemble.RandomForestRegres`
`https://scikit-learn.org/sklearn.ensemble.GradientBoostingRe`

[3] SuperDataScience
`https://www.superdatascience.com/pages/machine-learning`

[4] Google Machine learning
`https://developers.google.com/machine-learning/crash-course/`

[5] Linear Models for Video Memorability Prediction Using Visual and Semantic Features- R.Gupta K motwani
`http://ceur-ws.org/Vol-2283/MediaEval`$_1 8_p aper_3 1.pdf$