Predicting Video Memorability using HMP features

Krithika Sharon Komalapati
Department of Computing
Dublin City University
krithika.komalapati2@mail.dcu.ie

ABSTRACT

Memorability has been made an important aspect in the world of AI and Machine learning. Due to massive amount of data in the digital platforms, memorability is being used for retrieving lots of useful data. We work on the MediaEval 2018 Task where the based on the given features the short-term and long-term memorability is predicted. We use a pre-extracted feature HMP (Hierarchical Matching Pursuit) features to analyze memorability for videos using Regressor models. The K-Nearest Neighbor Regressor turned out to show the best results when compared to others.

KEYWORDS

Video Memorability, HMP features, Machine Learning, Regressors, Spearmen's Correlation Coefficient.

I. INTRODUCTION

In this paper, we use the HMP features to predict the video memorability of soundless videos. We are going to run the given ground truth values along the videos to train different regressor models and determine which one gives better results. These features are passed in a Random Forest Regressor, Decision Tree Regressor, K-Nearest Neighbor Regressor, AdaBoost Regressor, Gradient Boost Regressor and Neural Network models for prediction.

Using the spearmen's score for evaluating we selected the best model for the test set. Based on the scores we concluded that:

1. Short-term predictions were more accurate than long-term predictions.
2. KNN regressor gave better results when compared to all other models.

The rest of the paper is organized into the following sections: Section II discusses the literature review of this paper with respect to what has already been done. Section III presents deeply talks about the approach used for the video memorability test along with the regressors used for training and prediction. Section IV displays the results obtained and how they differed. Section V discusses the conclusion and future work.

II. LITERATURE REVIEW

Video memorability was initially investigated from the medical sector more where they used brain imaging to study the memorability with different people [1]. But with the boom of machine learning and deep learning, by learning the data we can predict and analyze a lot of them. For this, features of a video had to be studied and considered. [2] works with finding the short-term and long-term memorability values based on neural feedback. They use ensemble models to train data and get required results which were compare with the spearmen's coefficient.

Out of many different features, captions and HMP gave better results. We know that HMP gives best results with respect to video features [3] and is also best in achieving a high value for short-term memorability. From these papers we can say that using HMP features to obtain our results is the best option. Many features along with multiple machine learning models were used to get state-of-art results [4], but this paper show different results obtained when passed through regressor models.

III. APPROACH

A memorability dataset which consists of 10,000 videos – 6000 test set and 2000 test set are given. The dataset contains no audio videos which are 7 seconds in time. Pre-extracted features such as C3D, HMP, captions, etc. are given along with the ground truth values of short-term and long-term memorability for the test set videos are given. Then the train set was divided into 4800 and 1200 videos for training and validation. The approach was to work with the given features on videos and train them with different models and then decide which one model to use for the test set. HMP features were used because it is said that it gives best results for image features [3].

Feature Extraction: The HMP features were used on the dev video which had 6075 values for each video and then made into a Dataframe for future use. The Ground truth values were also merged with features extracted videos.

Model Fitting: The dataframe which had all the feature values is used and the splitting of data is done. The model is trained using 80% of the dev set and 20% is used for testing. First, the training and testing is done using the dev set only.

After the we get predictions, we compare them with the ground truth values for getting the Spearman;s score. We have used regressor models because we need continuous values and not to classify them under categories. The models used are:

1. Random Forest Regressor
2. Decision Tree Regressor
3. K-Nearest Neighbor Regressor
4. AdaBoost Regressor
5. Neural Network
6. Gradient Boost Regressor

Some hyper tuning was done for the regressors to try to perform better like changes the estimators in KNN and the using different activation functions in NN, but the results did not vary much.

Now based on the correlation value obtained, the best model is used to predict values for the test set videos, which will be compared with ground truth values later. The CSV containing these values is given along with this paper.

IV. RESULTS

From the table 1, we can say that the short-term memorability predictions are more accurate than the long-term memorability predictions. The K-nearest Neighbor Regressor and the Gradient Boost Regressor have given significantly better results than the other regressors. The KNN model gives a short-term memorability correlation of 0.287 and long-term memorability of 0.155.

Spearman's Correlation Coefficient		
Model	**Short-term**	**Long-term**
Random Forest	0.215	0.028
Decision Forest	0.058	0.003
K Nearest Neighbor	**0.287**	**0.155**
Ada Boost Regressor	0.163	0.071
Neural Network	0.046	- 0.02
Gradiant Boost Regressor	0.225	0.105

Table 1: Comparison with various regressors

The NN gave the worst results with long-term memorability correlation at -0.02, maybe because a lot of hyper tuning and adding layers is needed which is not relevant because this paper just wants to show the difference between models which are simple. However, for the Random Forest Regressor, the short-term correlation is very high but the long-term memorability correlation is very low. The Decision Tree and AdaBoost Regressors are very low to even perform.

V. ANALYSIS AND DISCUSSION

From the above-mentioned results, it can be said that Short-Term memorability predictions are more credible and accurate than Long-Term predictions. Also, the HMP features give good results.

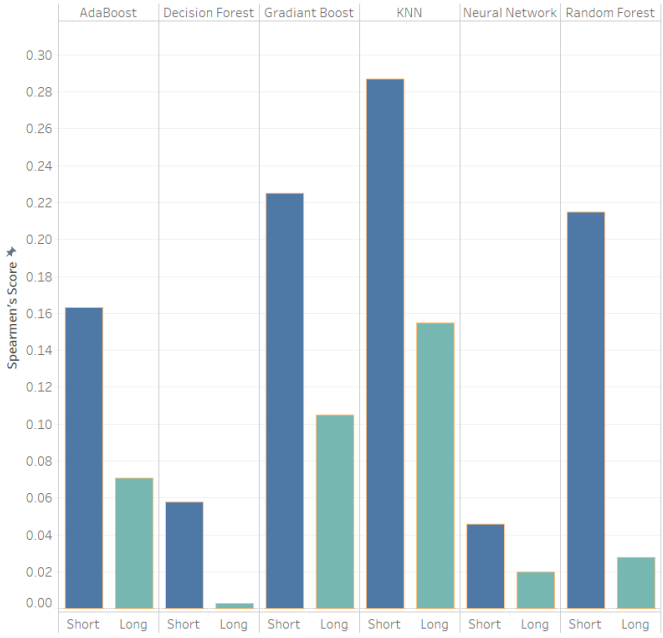

Figure 1: Bar graph showing difference between regressors

CONCULSION AND FUTURE WORK

This paper discussed about the memorability of videos, both short-term and long-term, how they differed between different regressor and which is the better one.

Most of the regressors are not hyper tuned. We have used only one feature for this paper, but we can use captions, C3D, Aesthetic and other features. We can also combine many features as one and then extract for prediction. The MediaEval Compition would be the best platform to show case any state-of-art results.

REFERENCES

[1] J. Han, C. Chen, L. Shao, X. Hu, J. Han and T. Liu, "Learning Computational Models of Video Memorability from fMRI Brain Imaging," in *IEEE Transactions on Cybernetics*, vol. 45, no. 8, pp. 1692-1703, Aug. 2015, doi: 10.1109/TCYB.2014.2358647.

[2] Azcona, D., Moreu, E., Hu, F., Ward, T.E. and Smeaton, A.F., 2020, September. Predicting media memorability using ensemble models. CEUR Workshop Proceedings.

[3] Bo, L. and Fox, D. (no date) ‘Hierarchical Matching Pursuit for Image Classification : Architecture and Fast Algorithms’, pp. 1–9.

[4] Smeaton, A.F., Corrigan, O., Dockree, P., Gurrin, C., Healy, G., Hu, F., McGuinness, K., Mohedano, E. and Ward, T.E., 2018. Dublin's participation in the predicting media memorability task at MediaEval 2018.