Media Memorability prediction using Neural Network, Linear model and Ensemble Method

Kiran Negi
Student ID: 19210510
Dublin City University, Ireland
Kiran.negi2@mail.dcu.ie

ABSTRACT

The quality of being worth remembering is the definition of memorability from dictionary. In current era with the plenitude of video substance on the web and social media, forecast of media memorability has the numerous conceivable application. This paper, with the given visual, semantic and video caption highlights from Medieval 2018 organizers for the video, utilized in models like Linear regression, support vector regression and dense Neural network to foresee its memorability.

1 INTRODUCTION

Video memorability depends on different highlights such as semantics, color, saliency, etc. [1]. The primary task was to anticipate the Media Memorability score which will depict how vital the video will be.[2] In this paper, the investigation was to the different given semantic and visual highlights to anticipate the media memorability for the video and to perform the profundity examination on the given highlights to create the vigorous indicators for the media memorability. We have been given with different visual features such as HMP, LBP and Color Histogram visual features and InceptionV3-Predictions & C3D-Predictions semantic features and video captions, where semantic or visual highlight didn't contribute much towards score as that of video caption commitment [3]. I have worked with C3Dpreds feature and captions for the videos to construct the model. The models were assessed utilizing Spearman's rank relationship as the metric. Below are the key discoveries from this work-

i) Short-term memorability forecast score was higher as compared to the long-term memorability forecast.
ii) Model based on captions outperformed Models based on C3D feature.
iii) Model based on captions outperformed Model-based on both features combined, C3D & caption.

2 RELATED WORK

Numerous analysts have taken interest in the field and looking at it as the potential application of Machine Learning in this region. In recent work [2] [3] [4], deep learning based semantic highlight representation (C3D-Preds), different levels of visual features and video captions were utilized for the memorability forecast. The researchers also noticed that the CNN model, which is one of the state-of-the-art technique, prepared with high-level semantics features for image classification has displayed best performance on numerous computer vision task .[5]

3 APPROACH

3.1 Model:

The potential concern I faced during model selection was high variance and overfitting since most of the features were high dimension. I went for two linear models and one dense neural network model. I also tried using Ensemble Technique of Simple average.

i) Linear Regression
ii) Support vector Regression
iii) Dense Neural Network
iv) Ensemble (Simple Average)

I have seen from the recent work in the area that Caption was most prominent feature as compared to other features provided to us. Hence, I performed some Natural language Processing. Also, extracted C3d-pred feature and applied to above three models to predict long and short-term score. I also tried using Single Average Ensemble technique.

3.2 Feature and Data Processing

Being a generic, compact feature for video, C3D-Pred is efficient to compute. As this feature was provided to us, I tried by loading the extracted feature for each video file and then merged it with ground truth file.

Other features such as Captions which is text for video was provided to us, I pre-processed by removing punctuations, stop words and converting them into lowercase. Followed by converting these cleans captions into bag of words by using methods such as CountVectorizer and TfidfVectorizer using Scikit-learn library. TfidfVectorizer displayed better performance. I also extracted length of the captions and counted noun words as additional features. All the above models used the extracted features, text and visual. I used parameters such as C, alpha, epsilon, tolerance, dropout, regularizer for finetuning and regularization.

4 RESULTS AND ANALYSIS

Below shown table displays the summary of the experimental results.

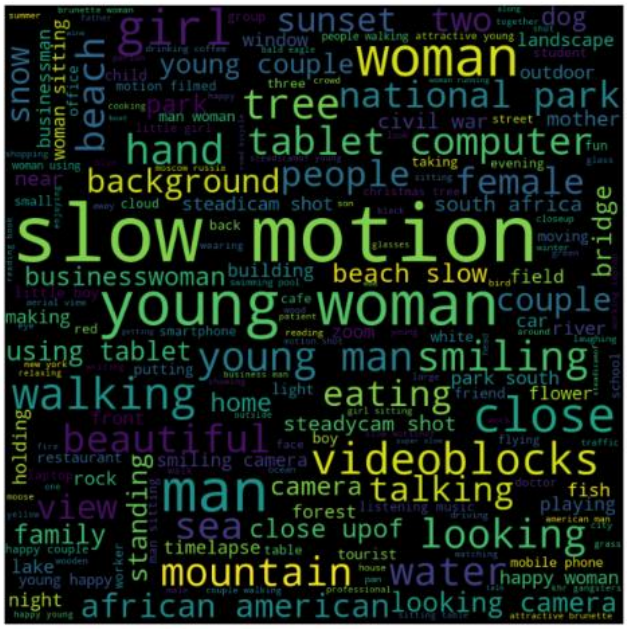
Table 1: Short-Term Memorability Prediction

Model	Features		Spearman
LR	Captions	CountVector+	0.291
		TfidfVector+	0.316
	Captions + C3D	TfidfVector+	0.306
DNN	Captions + C3D	TfidfVector+	0.073
		One hot encod	0.283
SVR	Captions	TfidfVector+	**0.408**
	C3D		0.248
	Captions + C3D	TfidfVector+	0.405

Table 2: Long-Term Memorability Prediction

Model	Features		Spearman
LR	Captions	CountVector+	0.091
		TfidfVector+	0.087
	Captions + C3D	TfidfVector+	0.101
DNN	Captions + C3D	TfidfVector+	0.077
		One hot encod	**0.160**
SVR	Captions	TfidfVector+	0.141
	C3D		0.039
	Captions + C3D	TfidfVector+	0.135

As part of exploratory analysis, I tried plotting word cloud below for the interpretation of caption feature and to display most occurring words in it. The bigger in size word means more occurring in the caption feature and the smallest size represents least occurring word.



5 CONCLUSION AND FUTURE WORK

From the MediaEval video memorability prediction has given me an opportunity to deep dive into the complex models and play with them by continuous training and testing of data. Models used by me did not perform that well since the high dimensionality of the fetched features. If given more dataset I would have tried attempting the problem with advanced state of art techniques such as CNN, LSTM which are known to get good accuracy. I tried with Ensemble Simple average but the prediction variables used in all my the models were of different shape and when I could find two models prediction variables with same shape, in the end when I tried getting the spearmen's score it showed it was out of indices. As part of future scope, I would like to tweak my models by overcome this challenge. I will keep exploring more ensemble techniques to achieve better score.

REFERENCES

[1] L.-V.-T. Duy-TueTran-Van, "PredictingMediaMemorability UsingDeepFeaturesandRecurrentNetwork," HoChiMinhCity, 2018.

[2] C.-H. S.-T. RomainCohendet, "MediaEval2018: PredictingMediaMemorabilityTask.InProc. of the MediaEval 2018 Workshop," 2018.

[3] K. M. Rohit Gupta, "Linear Models for Video Memorability Prediction Using Visualand Semantic Features," Conduent Labs, India, 2018.

[4] D. S. a. H. S. Sumit Shekhar, "Show and Recall: Learning What Makes Videos Memorable," in *Computer Vision and Pattern Recognition. 2730–2739*, 2017.

[5] H. A. J. S. a. S. Ali Sharif Razavian, "CNN Features Off-the-Shelf: An Astounding Baselinefor Recognition. InThe IEEE Conference on Computer Vision and PatternRecognition (CVPR) Workshops," 2014.

[6] A. Verma and S. Mehta, "A comparative study of ensemble learning methods for classification in bioinformatics," *2017 7th International Conference on Cloud Computing, Data Science & Engineering - Confluence*, Noida, 2017, pp. 155-158.