

Prediction of Long-Term and Short-Term Video Memorability

CA684 – Machine Learning Assignment

Satyam Ramawat (19210520), MSc in Computing (Data Analytics),

Dublin City University, Dublin City, Ireland.
satyam.ramawat2@mail.dcu.ie

Abstract—This paper focuses on how machine learning algorithms solve real-life problems and also introduced how efficiently marketing strategy can be done to grow up the sales of various goods and products. Video memorability also helpful in understanding the threats sent over the video to reduce the risks like terrorism, violence, and the spread of fake humor. Henceforth, this project focuses on how much a machine can remember a video by analyzing them for a long and short period. The various semantic features and video features have been available to test the experiment whereas I have focused on Semantic: Caption and Video: HMP and sum of 4 experiments have been done which can be elaborated later in this paper. The Machine learning techniques used in the experiments is Regression modeling to predict video memorability score for both terms. Finally, the best-scored model/experiment has been used for final modeling to generate good-fit data.

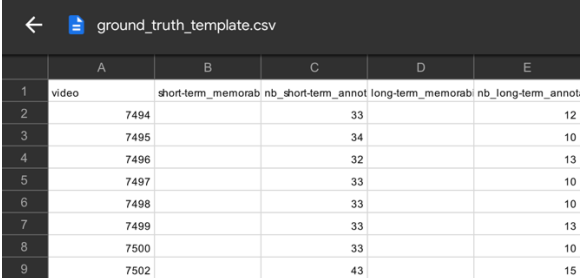
Keywords—Machine Learning, Video Memorability, Semantic Features, Captions, TFIDF-Vectorizer, HMP, Data Pre-Processing, Long-term memorability, short-term memorability.

I. INTRODUCTION

The Machine Learning technique enriches the ability to make things machines understand like human-being. Prediction is the most demanded and elegant feature provided by the machine learning technique to generate predicted data which can be used to reduce or strangle the risk and its affecting factors. In this study, prediction of video memorability score is done for long-term and short-term, i.e. how long a machine can remember the video by analyzing them from its semantic and video features. Whereas, four experiments have been performed,

1. Semantic: Caption
2. Semantic: Weighted Caption
3. Video: HMP
4. Semantic + Video: HMP + Caption

The best combination has been selected for the final modeling to generate predicted long-term and short-term memorability values for each video. In section Approach and Methodology, all regression models which are used for the experiment have been described along with the features in detail. Finally, results are saved in the Ground_Truth_Template.csv file of Test-Set, as shown in Fig. 1



	A	B	C	D	E
1	video	short-term_memorab	nb_short-term_annot	long-term_memorab	nb_long-term_annot
2		7494		33	12
3		7495		34	10
4		7496		32	13
5		7497		33	10
6		7498		33	10
7		7499		33	13
8		7500		33	10
9		7502		43	15

Fig. 1 Ground Truth Templat – Test Set.

II. LITERATURE REVIEW

[1] In the paper “Linear Models for video memorability Prediction using visual and semantic features” authors guided how simple linear regression models can we use in order to predict the values, where the author had used LASSO (L1) regularized Logistic Regression, Linear Support Vector Regression and ElasticNet (L1 and L2 Regularized Linear Regression). Where, Kush et al., has findings of the most positive coefficient and Negative coefficient which helped them to stimulate better results. The learning from this paper is to use the Caption feature whereas the positive words coefficient helped to give more weight to word bag corpus. [3] Here authors had introduced to ensemble models by predictions, despite running train sessions for big size vectors, where they have limited the memory by providing various modality parameters, such as captions with emotions that give extra weight to TFIDF vectorizer.

The author wensheng sun et al. [2], has predicted video memorability score for long-term and short-term by using a combination of video aesthetic and semantic video, where they have used spearman’s correlation in order to compare and select the best the model for further analysis.

From the above papers, it has been clearly understood that linear regression models and a combination of video features will provide more good score for long-term and short term memorability. Henceforth, approaches used in this work are based on the learning from the above research papers.

III. APPROACHES AND METHODOLOGY

A. Models

- Linear Regression Model
- Decision Tree Regression Model
- Random Forest Regression at Estimator 10.
- Random Forest Regression at Estimator 100.
- TFIDF Vectorizer

In this work, the above four models have been used to experiment with the semantic and video features for video memorability.

B. Features Experimented and Pre-Processing

- **HMP** is the Histogram of Motion Pattern video feature where it has been pre-processed and experimented alone with all three regression models.
- **CAPTIONS** is the semantic feature where it tells more details about the video, to process this Natural Language Processing has been used, where all the stop words and symbols have been removed and converted into lower case. Only captions feature to provide better Spearman's correlation score than only HMP feature
- **HMP & CAPTIONS**, this the combination of both semantic and video features, where it has generated better less than equal to better results than HMP.
- **WEIGHTED CAPTIONS**, this is the extra-ordinary experiment done where before fitting data into the TFIDF vectorizer, while pre-processing word bag dictionary created of positive coefficient terms, according to the paper and work [2] [3]. In the paper[2] it is described that some words or terms have a more positive effect on video memorability. The list of terms is additionally provided in the research paper with its impact. As opposed to the prevalent view, terms relating to the environment had an antagonistic impact that results out in less video memorability scores and terms relating to individuals or particular errands or activities had positive results which result in good video memorability scores. Leveraging the benefit of this work and taking it forward to the next level, The implementation has been done by providing the extra weights to the captions with the help of positive words vector, as per given in Figure 2

```
positive_words_dict = {'women':16,'woman':16,'eating':15,'putting':14,'lying':13,'girl':12,'selfie':11,
```

Fig. 2 Positive word Corpus

All the mentioned features have been experimented by using three simple linear regression models, where the benefit of exploring features has enabled to select the best feature and model based on Spearman's correlation scores. The Weighted Captions has performed well from all other features, and also

Random Forest Regressor has been outperformed from all other regression models.

Finally, for the prediction of video memorability score for long-term and short-term scores, the Machine Learning model has been built upon Random Forest Regressor at estimator 100 by using a weighted caption feature. The output of the predicted values has been saved in Ground_Truth_Template.csv file which is available in Test-Set.

IV. RESULTS

Features	Linear Reg.	Decision Tree Reg.	Random Forest Reg. 10	Random Forest Reg. 100
Weighted Caption	0.132	0.285	0.294	0.351
Only Caption	0.148	0.259	0.289	0.325
HMP	0.043	0.087	0.224	0.321
HMP + CAPTION	-0.008	0.093	0.224	0.318

Table 1: Short-Term Spearman's Correlation Coefficient Score.

Features	Linear Reg.	Decision Tree Reg.	Random Forest Reg. 10	Random Forest Reg. 100
Weighted Caption	0.030	0.093	0.158	0.193
Only Caption	0.036	0.127	0.148	0.154
HMP	-0.027	0.043	0.056	0.112
HMP + CAPTION	0.038	0.038	0.035	0.138

Table 2: Long-Term Spearman's Correlation Coefficient Score.

video	nb_short-term_annotations	nb_long-term_annotations	short-term_memorability	long-term_memorability
7494	33	12	0.847892088	0.765770185
7495	34	10	0.90673107	0.855989349
7496	32	13	0.896926361	0.82662024
7497	33	10	0.856702928	0.777344837
7498	33	10	0.857068392	0.780128407
7499	33	13	0.859100358	0.782812963

Fig. 3 Prediction Results saved in Ground Truth Test-Set

All evidences like python notebook along with all the csv has been provided along with the folder, also notebook shareable link is available at [4].

V. CONCLUSION

In this research, the work described the robust way to compute the memorability, the finding is the semantic feature is more capable and accurate to do a prediction for long-term and short-term video memorability score where addition support to word bag corpus leads result to outstanding spearman's score performance, thus found that weighted caption has the highest and significant score and this implies that caption is a good source to do furthermore investigation in the domain of video memorability.

REFERENCES

- [1] Rohit Gupta and Kush Motwani, "Linear Models for Video Memorability Prediction Using Visual and Semantic Features", MediaEval'18, 29-31 October 2018, Sophia Antipolis, France.
- [2] Wensheng Sun and Xu Zhang, "Video Memorability Prediction with Recurrent Neural Networks and Video Titles at the 2018 MediaEval Predicting Media Memorability Task", MediaEval'18, 29-31 October 2018, Sophia Antipolis, France.
- [3] Azcona, David and Moreu, Enric and Hu, Feiyan and Ward, Tom {\a}s E and Smeaton, Alan F,"Predicting Media Memorability Using Ensemble Models", MediaEval'19, 2019, CEUR-WS.
- [4] <https://colab.research.google.com/drive/1DVip9Irr6kO10d01swqvB2epaUwxYELw>