Name	Navaneethan Rajasekaran
Type of Document	Machine Learning Assignment
Student Id	18210481
Date	24.04.2019

Predicting memorability of the video using polarity of the sentiment in video

Navaneethan Rajasekaran
Dublin City University
navaneethan.rajasekaran2@mail.dcu.ie

ABSTRACT

This paper explains the approach developed to predict the short-term and long-term video memorability of videos at the 2018 MediaEval Predicting Media Memorability Task. This approach uses the features generated from Title of the video using NLP. The performance of the model for prediction was compared to using different combination of features.

1 INTRODUCTION

Predicting the memorability of the video was always a challenging task. There was a possible correlation between the objects and memorability, that if a system knows which objects an image contains, it can predict memorability. Although images are different from videos, the connection between the images and videos will be a driving factor for this approach.

The extraction of information from the title of the video will be less computational. The plan was to use extract the details such as sentiment's polarity, objects of the video from Titles by NLP techniques and use it as features for the memorability prediction. The Spearman correlation was used as a metric to evaluate the model.

2 LITERATURE REVIEW

Objects without semantics are not valid at predicting memorability. The author[2]has estimated the contribution of objects in the image to the memorability. The method sorted objects into an intuitive ordering: people, interiors, foregrounds, and human-scale objects tend to contribute positively to memorability; exteriors, wide-angle vistas, backgrounds, and natural scenes tend to add negatively to memorability.

Even though there is certain correlation between short and long-term memorability, results[3] have shown that short-term memorability is more predictable than long-term one since all models score higher in short-term than long-term memorability. Long-term scores range from 0.2 to 1 and exhibit higher variance than the short-term scores, which distribute from 0.4 to 1. And the author suspects that one possible reason is that the long-term memorability is more subjective and depends more on an individual's memory.

The investigation[1] on preference for certain scene types found that people prefer natural outdoor scenes rather than human-made scenes. While interestingness was higher for images containing sky, actual memorability decreases if the sky is present. Indeed, when comparing actual memorability and interestingness, they found them to be negatively correlated.

3 APPROACH

The approach was to extract information about emotions, objects of the video from their Titles using NLP Techniques.

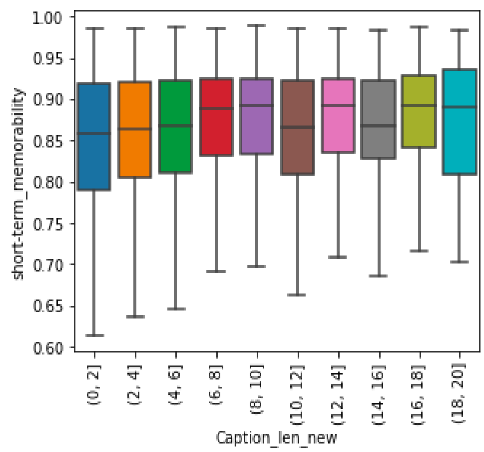

Figure 1: Caption's word length

As the short-term memorability was correlated with long-term memorability. The initial approach is to predict the short-term memorability, with inputs generated from Titles. And to use the predicted short-term memorability as one of the feature to predict the long-term memorability.

3.1 Machine Learning with Title's word counts

The approach is to count the number of words in Title and create a feature called captionlen. The word count of the title without stop words was used as one more features called captionlennew. Later, the captionlennew is divided into categories. The result illustrated in Figure 1, showed us the average short-term memorability score of some categories are similar. These similar categories are categorized and created as features called feature1 with values ranged from 0 to 3. Using the above features as input, two machine learning models-RandomForestRegressor, SVR was trained to predict the short-term memorability.

3.2 Machine Learning with Title's word count and sentiment score

In addition to the features generated from Title's word count, two more features such as sentiment score and a subject score of the Titles were created. Assuming that sentiment will correlate with the memorability. Using the above features as input, three machine learning models- RandomForestRegressor, SVR, Lightgbm were trained to predict the short-term memorability.

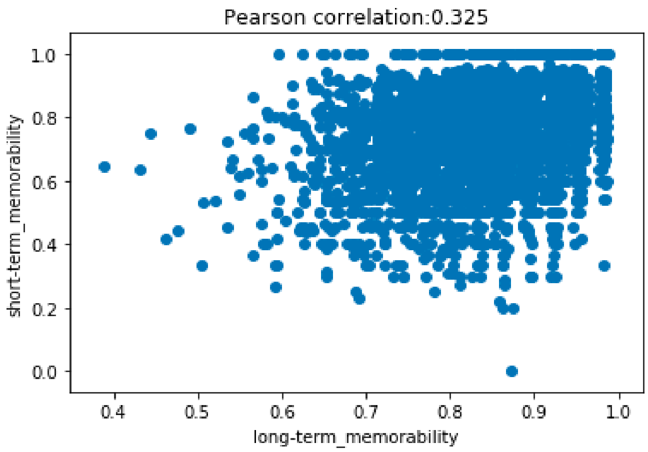

Figure 2: Correlation between short-term and long-term memorability

3.3 Machine Learning with Title's word count and polarity of the sentiment

The result of models was improved when the sentiment score was used as input. Later the polarity of the sentiment was calculated and created as four new features called positivity, negativity, neutral,compound score. From this approach, it is found that negativity score had an impact on short-term memorability score and improved the model performance in prediction.

3.4 Machine Learning with Title's word count,the polarity of the sentiment and Tf-idf

The tf-idf with trig-ram were calculated for the Titles. The result of this approach has produced more than 2000 columns. In addition to the above-generated features, these features were used as input in the model to predict the short-term memorability.

As you can see in figure 2, the short-term memorability was found to be correlated with long-term memorability . So with the inputs used above to calculate short-term memorability and predicted short-term memorability as input, the model was trained to predict the long-term memorability.

3.5 Predicting Long-term memorability using short-term memorability

As short-term memorability was correlated with long-term memorability, the short-term memorability was predicted initially with Title's word count, Polarity, Tf-idf as inputs. Later the predicted short-term memorability in addition to Title's word count, Polarity, Tf-idf were used as input to predict long-term memorability. This approach was found to be effective.

4 RESULTS AND ANALYSIS

From the different approaches in Table 1 and Table 2, the following conclusions were made.

Model	RandomForest	SVR	Lightgbm
Input			
Title's word count	0.147	0.137	0.160
Title's word count, Polarity	0.226	0.217	0.199
Title's word count, Polarity, Tf-idf	-	-	0.346

Table 1: Spearman correlation for short-term memorability

Model	Lightgbm
Input	
Title's word count, Polarity, Tf-idf	0.164
Title's word count, Polarity,Tf-idf + short-term memorability	0.297

Table 2: Spearman correlation for long-term memorability

The features generated from Titles of the video helped to extract details of objects present in the video. When compared to the other predefined features, inputs generated from video's Title provided a better result. The polarity of the sentiment has proven to be highly correlated with the memorability. Like the objects in the images also contribute to the memorability score of the humans. Using sentiment score with a subject score of the Title increased the performance of the models. In addition to this, encoding the Titles as vectors and added as a feature, has increased Spearman correlation of short-term memorability by 18 per cent. The negativity in the video affected memorability when compared with positivity and neutral. Further investigation of the sentiments of the video is needed.

Though the short-term memorability was correlated with long-term memorability, using the inputs of short-term memorability has not provided the same result to the long-term memorability. One reason could be the variance of long-term memorability was more compared with short-term memorability. By using short-term memorability as one of the features depicted in Table 2, the Spearman correlation of long-term memorability had increased.

It is observed that RandomForest, SVR had provided that similar result when compared to Lightgbm. But failed to perform better when high dimensional data was used. The lightgbm being a gradient boosting framework that uses tree-based learning algorithms provided better accuracy when using high dimensional data.

As the long-term memorability depends on the individual's memory. The performance of the model to predict long-term memorability was worse. But further investigation of the sentiments of the video can increase the model performance to predict long-term memorability. The approach of predicting short-term memorability initially and using it as an input for long-term memorability was found to be effective.

5 CONCLUSION AND FUTURE WORK

The approach was mainly focused on the sentiments and objects of the video. I have not tried a combination of different features. Hence i will try to investigate more on the sentiments of the video and use it with a different combination of features.

REFERENCES

[1] Gygli, M., Grabner, H., Riemenschneider, H., Nater, F., and Gool, L. V. The Interestingness of Images. In *2013 IEEE International Conference on Computer Vision* (Sydney, Australia, Dec. 2013), IEEE, pp. 1633–1640.

[2] Isola, P., Xiao, J., Torralba, A., and Oliva, A. What makes an image memorable? 8.

[3] Sun, W., and Zhang, X. Video Memorability Prediction with Recurrent Neural Networks and Video Titles at the 2018 MediaEval Predicting Media Memorability Task. 3.