Predicting Memorability of Videos using Various Features Like C3D and Captions

Palash Dange[1]
[1]Dublin City University
palash.dange2@mail.dcu.ie

ABSTRACT

Humans have a tendency to remember the things they visualize for some amount of time. Generally these visualization powers of a video are measured in terms of long and short term memorability. From MediaEval we have been given 8000 videos distributed in dev-test and test set. Both these datasets consist of short term and long term memory annotations. The task of this paper is to select models in which we are able to predict the values of short and long term memorability effectively. We selected various machine learning models including SVR, Random Forest RNN. Our best model is based on the features captions which use random forest method to predict the values. We have also tried ensemble of two models which had the highest Spearmen coefficient values. Spearman coefficient for the training set is 0.419. Other features such as C3D can also be used to predict the memorability values of the test dataset.

1 INTRODUCTION

Media Eval 2019 is a challenge which provides challenges in multimedia retrieval access and exploration. This video task is focused on the prediction of long-term memorability and short-term memorability of videos. Memorability of videos is actually the score of how well you can remember the video later after you see it once or twice. Spearman coefficient is used to determine what kind of models are more precisely predict the long term and short term memorability. The model, if designed correctly, could be of high value in the field like digital marketing. Also in the field of sales and marketing, where the customer is the main focus and we can design the videos which are more memorable and impressive to customers.

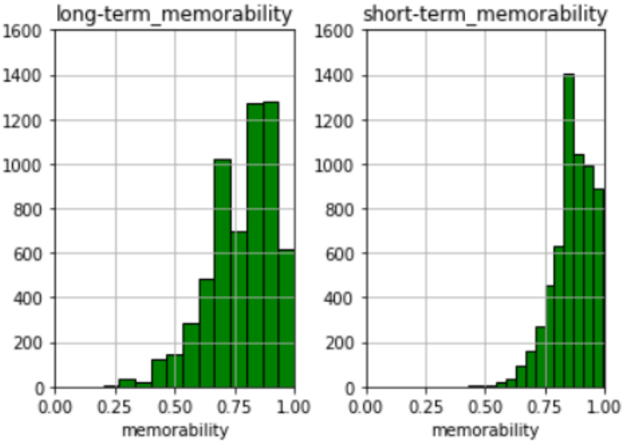
The dataset used in this project consists of 8000 samples of videos to apply algorithms on the data. The dataset is distributed as 75% for training set and 25% for the test set. The training set of 6000 videos are provided, where we can implement hit and trial of various models and check for the Spearman coefficient. Those models which have the highest value of the Spearman coefficient are then ensembled and used together to provide a better model. Model finalized from the above process is then used to test the 2000 videos dataset and predict the score of long term and short term memorability.

2 RELATED WORK

This paper explains the approach taken by me to predict the required task with given sources. The sources provided in the task include the Google Collab file which has the function to load all the features and a sample of Neural network model prediction. From the above resources, I have used the Google Collab file provided by Eion to load different features of data set and apply different algorithms. Doing a bit of research on the past submission result of MediaEval 2019, I found that the highest values of the coefficient are given by a few features which are C3d and Captions overall. I have also used a weighted ensemble approach to check the predictions but the predictions using these methods were not so great. Going to the machine learning models. I have used SVR (Support Vector Regression) and Random forest with C3D and Captions alone and then a combination of few to get the Spearman values. Later on, I went on predicting values of memorability both short term and long term.

Using the models having the highest value of Spearman. I have selected models to apply on the test-set. Please follow the Google Collab IPYNB file provided in the references to understand the process thoroughly. Below is the distribution graph for the memorability training samples provided.

Figure 1



3 APPROACH

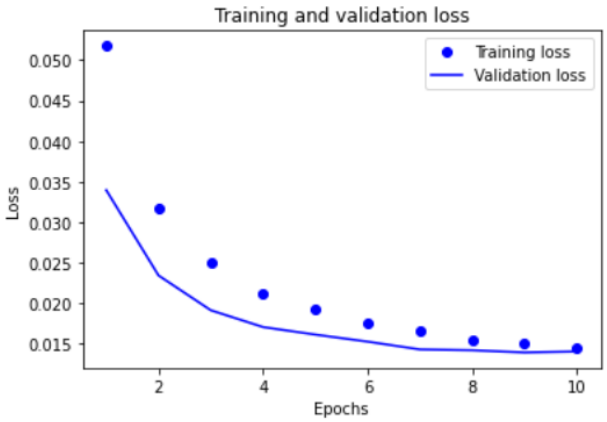
The main reason to select the features C3D and Captions were the past submissions done by the various good performing teams. The Media Eval data set was provided to us on which the research was carried out with the below steps.

Section 3.1 The dev-test data provided was imported and loaded in Google Collab. These features are loaded into a variable using python library pandas. Several pre-processing steps are applied to captions and C3D to create test data.

Section 3.2 These steps include counting each word in the caption file and then removing the stop words. Using count vectorizer, we have given weights to each caption according to the score given. These captions are split into words using tokenization.

Section 3.2 The C3d variables are loaded similarly to the captions and loaded into the dataset. After this, the dev-test data was divided into train and test to apply the machine learning model. Various models like Random Forest(RF), Support Vector Regression (SVR) and Recurrent Neural networks - RNN with multiple layers were applied to the training data to calculate the spearmen coefficient.

Figure 2: - RNN



Section 3.3 After applying the models individually on the data set, a combination of different features were used and models were applied on the data to calculate them different values of Spearman coefficient. Once we have values of the Spearman coefficient of individual feature models and models created using combination of features. We have used ensemble of models like weighted average and sample average models to check, but these models are not used for the prediction purpose.

Section 3.4 We have also implemented RNN which we have build using LTSM which generally performs well in NLP tasks. for this task we have used captions as one of the features, the RNN with LTSM is implemented with 150 hidden layers followed by 30 neurons The short term and long term memorability which is the last layer, for this we go through 2 neurons each. We have use the optimizer as admax and for the hyperparameters selection we use Gridsearch to find the best parameters. The model is trained on 10 epochs and the training and validation loss for the same can be seen in the Figure 2.

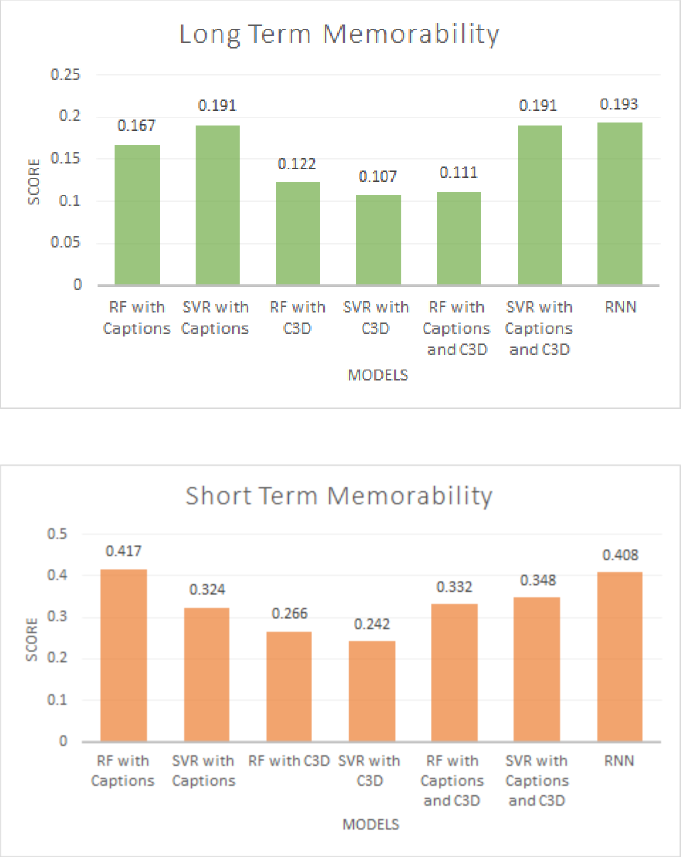
4 RESULTS AND ANALYSIS

The evaluation of multiple models implemented is done. The result set is obtained from the test set provided in the assignment. The test set includes different features with 2000 number of entries and for the same, the memorability values are provided for us to compare.

Observing the table values we choose to implement the random forest with captions on the data set as we can see the best values provided for short term memorability is from the random forest with captions. So I finalized the model and implemented the same on the test data set. Even the long term memorability values for RNN can be seen as the best one.

After implementing the combination of different features and one of the neural network models, we can see the results in the below graphs.

Figure 3: Memorability Score from different Models



5 DISCUSSION AND OUTLOOK

Based on the above result we concur that the best results are provided by the random forest with captions model for the short term and long term memorability. The results are printed in the Goolge Collab file and uploaded. As for future work, we can try and ensemble an algorithm like gradient, ada boosing approach to mix and match multiple models. The hit and trial of ensemble techniques can give use more better result. Even different neural networks with multiple hidden layers and large number of epochs can be implemented to check which one is the best prediction model. Also just not the model variation, but we can try a different combination of features like color histogram, inception, and HMP and many more and check the result for the same.

REFERENCES

[1] Cohendet, R., Demarty, C.H., Duong, N., Sjöberg, M., Ionescu, B. and Do, T.T., 2018. Mediaeval 2018: Predicting media memorability task. arXiv preprint arXiv:1807.01052.

[2] Tran-Van, D.T., Tran, L.V. and Tran, M.T., 2018. Predicting Media Memorability Using Deep Features and Recurrent Network. In MediaEval.

[3] Constantin, M.G., Ionescu, B., Demarty, C.H., Duong, N.Q., Alameda-Pineda, X. and Sjöberg, M., 2019, October. Predicting Media Memorability Task at MediaEval 2019. In Proc. of MediaEval 2019 Workshop, Sophia Antipolis, France.

[4] Savii, R.M., dos Santos, S.F. and Almeida, J., 2018. GIBIS at MediaEval 2018: Predicting Media Memorability Task. In MediaEval.