Predicting Media Memorability Using Regression Models and Ensemble Learning

Aakash Khair
School of Computing
Dublin City University
Dublin, Ireland
aakash.khair2@mail.dcu.ie

***Abstract*—Many images and videos are exposed to people on a daily basis, but they do not remember the details, even if many look very similar. This Video Memory (VM) is mainly because of the distinct and fine representation of frames from video that people tend to recall. Videos have an abundance of frame data that can be used to retrieve features. The aim of this challenge is to anticipate how people can recall a video in terms of short and long-term memorability. We're forecasting video memorability scores, which represent the likelihood of a video being remembered or not [1]. In this research our baseline approach is to predict video memorability using state of the art machine learning algorithm like XGBoost and also by hyperparameter tuning of ensemble method like random forest. This research also focuses on using stacking based ensemble approach to combine best outcome prediction from our baseline model to improve the performance of our short term and long term memorability score**

***Index Terms*—Random Forest, XGBoost, Stacking Based Ensemble, Hyperparameter tuning, video memorability, C3D, HMP, Aesthetic Features**

I. Introduction

Video is increasingly growing user engagement, search engine exposure, the rate of social media, the influence of email marketing, and audience perception of brands. Marketing videos are increasingly being used in a variety of contexts, including digital displays, out-of-home advertisements, and social networking [2]. The objective of Media memorability task includes determining how well a video can be recalled in both short and long term periods. The dataset provided includes the ground truth values of 6000 videos collected by the mediaEval team. We were also provided with some precomputed video features like C3D and HMP and image features like ORB, ColorHistogram, HOG, InceptionV3.

In this research the approach used for predicting the short and long term memorability focuses on using multiple features like precomputed video features C3D and HMP and a combination of them were trained using Regression Models like Bayesian Ridge and Elastic Net and also used Boosting algorithm like Gradient Boosting, LightGBM, and state of the art algorithm XGBoost. Using XGBoost with combined features of C3D and HMP gave outstanding results. I have also used ensemble model like Random Forest and using hyperparameter tuning of different parameters to find best optimal parameters. I have then stacked models using some of models which gave good results to increase prediction accuracy.

II. Related Work

Previous research has demonstrated that image memorability is inherent to visual content, but the issue of modeling video memorability has not been adequately discussed. Isola et.al in their research found that object and scene semantics are the key explanations for memorability [3]. The author identified new techniques for modeling video clip memorability and automatically predicting how memorable those videos are using brain functional magnetic resonance imaging (fMRI) [4]. Compared to esthetic feature-based approaches using SVR and ANN, semantics are more efficient than those used, and a basic function is used to analyze the probability of predicting highly subjective media memorability. The best one of the five models is the RNN and semantics [5]. The multimodal method was used to predict video memorability, and the author concluded that combining visual and textual elements yielded better results [6].

III. Approach

A. Dataset and Feature Description

The dataset provided comprises of 8000 videos, 2000 test set videos and development set of 6000 videos. Ground truth labels of development set were only provided we had to explicitly generate ground truth labels for the test set. Video features like C3D and HMP and a combination of them were used. Additionally, I have also used the Aesthetic image features to see if the intrinsic properties of images have any effect on predicting the memorability. C3D features in the form of text file were provided, we then used a function to extract those features in an array which was converted into numpy array. Dimensions of C3D dataset for development set were (6000,101). Similar process was used for loading HMP feature. Dimension of HMP Feature were (6000,6075).

B. Model Selection

The approach followed in this research was to train the traditional machine learning models and then using ensemble method named stacking which combines the base model using a meta model [7]. Following approaches were used namely, **Linear Regression Models and Bagging and Boosting Algorithm**

1) Lasso Regression
2) Elastic Net Regression
3) Random Forest
4) Gradient Boosting Regressor
5) XGBoost Regressor

Deep Learning Model namely,

1) Multilayer Perceptron

Each model was trained on each feature and then using stacked ensemble method selected the base model like Gradient Boosting repressor, Random forest and XGBoost as the meta model to check for improved prediction. The models were evaluated using Spearmen's rank correlation coefficient.

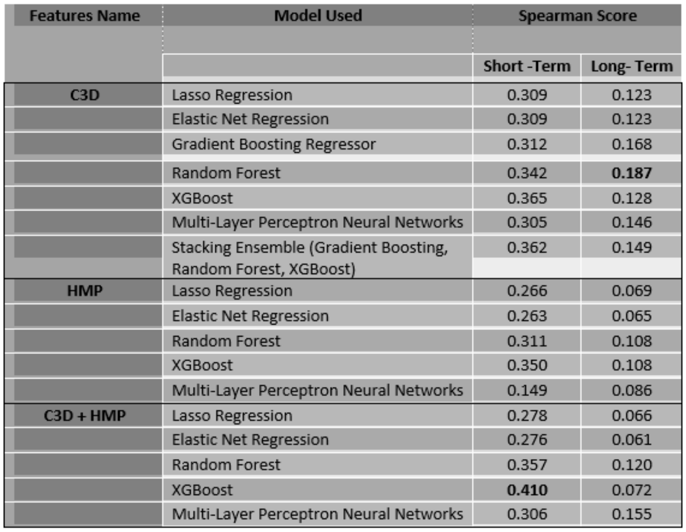
Linear Regression models like Elastic Net and Lasso Regression gave almost similar short term and long term memorability score on predicting on C3D Feature. I then trained the dataset using Bagging algorithm named Random Forest on C3D feature which was used with its default parameter and gave satisfactory results. To select the best parameters, I used hyperparameter tuning techniques like RandomizedSearchCV which gave us the domain set of parameters and further to narrow our range of values for the parameters used GridSearchCV and compared its result with the base random model which gave an improvement in predicting the scores.

Boosting Algorithm were then used namely, Gradient Boosting Regressor and XGBoost on each feature and a combination of them and found that XGBoost gave outstanding results of Short term memorability score but gave less score for long term memorability. The results can be seen in below Table 1.

Stacking named ensemble was used to predict for improved prediction. Models like Random Forest regressor and Gradient Boosting regressor where used as base models which learns from heterogeneous weak learners and combined them with meta model XGBoost for training. Multi-Layer Perceptron Neural Network gave almost similar results when trained using C3D and a combination of C3D and HMP.

IV. Results and Analysis

Experiments results on different combinations of features with models are given in Table 1. Best prediction of Long Term and Short term score are presented in the table. For Short term score using feature C3D + HMP, XGBoost is the best model whereas for Long term score on hyperparameter tuning and finding the best parameters of Random Forest gave good results. Stacked Ensemble method used the weak learners and gave average Short term memorability score whereas yielded poor results for Long term memorability score.



Features Name	Model Used	Spearman Score	
		Short -Term	Long- Term
C3D	Lasso Regression	0.309	0.123
	Elastic Net Regression	0.309	0.123
	Gradient Boosting Regressor	0.312	0.168
	Random Forest	0.342	**0.187**
	XGBoost	0.365	0.128
	Multi-Layer Perceptron Neural Networks	0.305	0.146
	Stacking Ensemble (Gradient Boosting, Random Forest, XGBoost)	0.362	0.149
HMP	Lasso Regression	0.266	0.069
	Elastic Net Regression	0.263	0.065
	Random Forest	0.311	0.108
	XGBoost	0.350	0.108
	Multi-Layer Perceptron Neural Networks	0.149	0.086
C3D + HMP	Lasso Regression	0.278	0.066
	Elastic Net Regression	0.276	0.061
	Random Forest	0.357	0.120
	XGBoost	**0.410**	0.072
	Multi-Layer Perceptron Neural Networks	0.306	0.155

Table 1: Spearman Short and Long Term Memorability Score

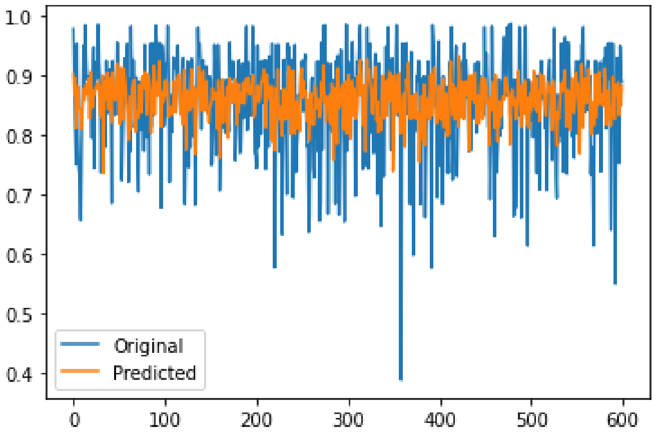


Fig. 1.Actual vs Predicted Memorability Score using XGBoost Model

V. Conclusion

Boosting algorithm XGBoost outperform traditional linear regression model and deep learning model Multilayer perceptron. Fine tuning XGBoost model to get the best parameters could improve the score of the prediction and help yield better result. Previous research of image memorability shows that training XGBoost with pre-trained model like ResNet152 and VGG16 gave model prediction near to human level memorability.

I will further explore the image features like LBP, HOG, InceptionV3 to see if there is any improvement in prediction. We can then use stacking ensemble method and use powerful boosting algorithm like XGBoost and CatBoost, AdaBoost can find best parameters using automated hyperparameter techniques like Bayesian Optimization and Optuna and build robust model to avoid overfitting for better predictions.

References

[1] Mediaeval. [Online]. Available: http://www.multimediaeval.org/mediaeval2019/memorabilit

[2] Spectrio. 13 video marketing stats that prove you're missing out by not using video. [Online]. Available: https://blogs.spectrio.com/the-13-video-marketing-stats-you-need-to-know

[3] P. Isola, J. Xiao, A. Torralba, and A. Oliva, "What makes an image memorable?" in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 145–152.

[4] J. Han, C. Chen, L. Shao, X. Hu, J. Han, and T. Liu, “Learning computational models of video memorability from fmri brain imaging,” *IEEE Transactions on Cybernetics*, vol. 45, no. 8, pp. 1692–1703, 2015.

[5] W. Sun and X. Zhang, “Video memorability prediction with recurrent neural networks and video titles at the 2018 mediaeval predicting media memorability task,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 4–6.

[6] ——, “Multimodal approach to predicting media memorability,” in *CEUR Workshop Proc*, 2018, pp. 29–31.

[7] J. Rocca. Ensemble methods: bagging, boosting and stacking. [Online]. Available: