CA684: Machine Learning

Name	Yashaswi Verma
Student Number	19211007
Program	Master of Science (data analytics)
Module Code	CA684
Assignment Title	Machine Learning
Submission date	29th April,2020
Module coordinator	Prof. Tomas Ward

I declare that this material, which I now submit for assessment, is entirely my own work and has not been taken from the work of others, save and to the extent that such work has been cited and acknowledged within the text of my work. I understand that plagiarism, collusion, and copying are grave and serious offences in the university and accept the penalties that would be imposed should I engage in plagiarism, collusion or copying. I have read and understood the Assignment Regulations set out in the module documentation. I have identified and included the source of all facts, ideas, opinions, and viewpoints of others in the assignment references. Direct quotations from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged, and the source cited are identified in the assignment references. This assignment, or any part of it, has not been previously submitted by me or any other person for assessment on this or any other course of study.

I have read and understood the referencing guidelines found at http://www.dcu.ie/info/regulations/plagiarism.shtml, https://www4.dcu.ie/students/az/plagiarism and/or recommended in the assignment guidelines.

Name: Yashaswi Verma Date: 29th April 2020

Predicting Video Memorability with Extracted Feautures and Captions

Yashaswi verma (19211007)
M.Sc. in Computing (2019-20)
Dublin City University
Dublin, Ireland
yashaswi.verma2@mail.dcu.ie

***Abstract* – Prediction has been an essential part of Machine Learning since the term has introduced. Predicting Memorability of any video is a hustle of going through lots of features and memory of the video. This application is widely used in various fields. This paper describes how short term and long-terms video memorability is predicted using pre-extracted features like caption, C3D combined with Machine Learning Algorithms. Performance is assessed using the Spearman correlation or score.**

I. INTRODUCTION

Brains have a strong propensity to memorize the video content, but a Machine memorizes the video content via the bits and bytes stream of the video. Video Memorability is an advancement to the Machine Learning field, which has a unique approach to this task. It is crucial to know what video feature can be useful so that people can save the video information for an extended period. The feature used in this report is C3D and Captions, which are available to predict the memorability of video. Regression approach is used to calculate the memorability of the video. Different Regression model is used such as KNN, Random forest Regressor for Computation. Caption Processing for a massive no of videos takes sizeable computational power, so different Natural Language Processing techniques are Used to cut the cost of Computation. After Prediction, Spearmen's Correlation is used to check the model for its efficiency.

Google Collaboratory has been used for the computational purpose.

II. RELATED WORK

In recent times, various researchers have shown interest in this field and they came up with assorted ways to automatically predict the memorability of the videos.

Most of the video memorability is estimated using neural networks on the video streams. Images and videos have long been used as data to test memory output and show that humans have adequate long-term visual memory.[1].

In contrast to previous work on image memorability – where memorability was measured a few minutes after memorization – memory performance is measured twice: a few minutes and again 24-72 hours after memorization. Hence, the dataset comes with short-term and long-term memorability annotations.[2]

It would be hard to generalize the form. Shekhar et al. investigated several characteristics, including C3D, semantic characteristics obtained from some video captioning phase, saliency characteristics, dense trajectories, and color characteristics, before developing their memorability predictor.[3]

III. APPROACH

Pre-extracted features are available for prediction such as Caption, C3D, HMP, etc out of these "Caption" and "C3D" is selected and stored in the data frame. Two regression model was used one on each to calculate the spearmen score.

A. Feature Extraction Method

The caption feature had 2 columns having video name and the caption of that video. The Dev-set and Test-Set caption were extracted in a data frame. Both the data frames were combined together so that we can clean both the data frames. Captions include noise and words, which could affect our computation.

Stemming, removal of stop words, and an essential function that helped in training the model using this feature, which is the TF-IDF vectorization method, which converted our caption into a vector form, which is stored in an array for training purposes.

C3D file is available for each video in Our dataset, extraction of the C3D feature used tqdm and glob library of python to extract it from all video files and load its value to a dictionary. Values in dictionaries can not be used for training, so we stored it in a data frame and then converted it into a series.

Both the feature was then merged into a single data frame having the tf-idf vectorized feature of caption and the C3D feature of the video.

B. Model Selection

As predicting memorability is a regression approach, multiple regression gives the best performance and accuracy.

The TF-IDF vector in the caption and the columns having the C3D feature act as independent variables. The values in Short- and long-term memorability are considered as the dependent variables here.

Linear regression is used as a simplest regression for predicting the memorability.

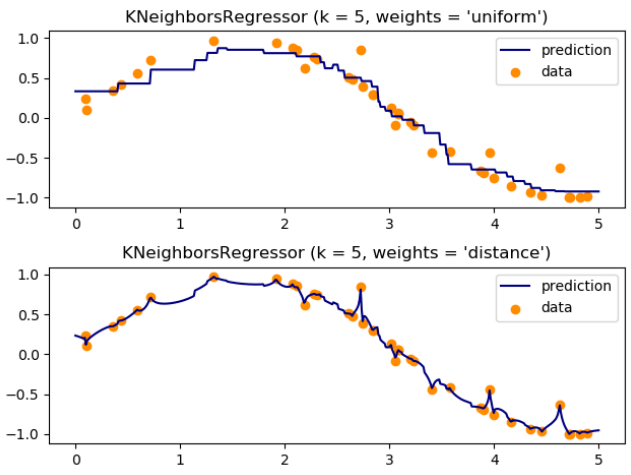
KNN-regressor is used as a regression model for C3D feature ,followed by an ensemble model RandomForest regressor, which is used for caption and predicting memorability of the Test-set.

C. Model description

The Prepared Data is Split in 80% and 20% ratio for training and testing the model . Splitting of test and train data is done manually.

RandomForest regressor is used and is an ensemble technique used both for regression and classification purposes. Random Forest approach involves training each decision tree on a particular sample of data in which replacement and sampling is performed. No of estimators is passed as an argument which defines no of decision trees in the model. Different no. of estimators are passed to check the optimal value and computed in the ipynb file.

K Nearesh Neighbour is a simple algorithm that stores all available cases and predicts the numerical target based on a measure of similarity (e.g., distance functions). KNN is used as a non-parametric tool in statistical estimation and pattern recognition.The neighbours are set in the model to 5 (minimum) to get our model accurate.



Source:*https://scikitlearn.org/stable/auto_examples/neighbors/plot_regression.html*

IV. Results and Analysis

The result is evaluated using the spearmen score which is stated in the below table separately for each feature and model used.

Regression Models	Captions *TF-IDF*		HMP *Sequence*	
	Long Term	Short Term	Long Term	Short Term
Random Forest	0.374	0.144	0.254	0.101
KNN	0.323	0.107	0.265	0.131
Linear Regression	0.135	0.026	0.289	0.098

Table 1

The data above describes ensemble model as best for *captions(tf-idf)* having the short-term memorability score of 0.374 and but not for long term memorability score i.e 0.144.

In Comparison to RF regressor KNN has less score for short term memorability i.e 0.265 but has similar score to long term memorability about 0.131.

Linear regression is used as a basic model the results are not much close to both the models.

These model Root Mean Square Error is also calculated to see the efficiency of the of model.

V. Conclusions

The evaluation of model is done by comparing their spearmen score which tells that the models used prediction for short term memorability is more accurate that calculating the long term memorability. Hence short term memorability .The accuracy of the model increase when we use caption (tf-idf) against the C3D feature of the video.

References

[1] Timothy F Brady, Talia Konkle, George A Alvarez, and Aude Oliva. Visual long-term memory has a massive storage capacity for object details. Proceedings of the National Academy of Sciences, 105(38):14325–14329, 2008.

[2] *Openaccess.thecvf.com*,2020.[Online]. Available:http://openaccess.thecvf.com/content_ICCV_2019/papers/Cohendet_VideoMem_Constructing_Analyzing_Predicting_ShortTerm_and_LongTerm_Video_Memorability_ICCV_2019_paper.pdf.

[3] KSumit Shekhar, Dhruv Singal, Harvineet Singh, Manav Kedia,and Akhil Shetty. Show and recall: Learning what makes videos memorable. In Proceedings of the International Conference on Computer Vision and Pattern Recognition (CVPR), pages 2730–2739, 2017, 2018.