Name:	Nihar Sudhanshu Limaye
Student Number:	18210876
Email Address:	nihar.limaye3@mail.dcu.ie
Programme:	MCM – Machine Learning
Module Code:	CA684
Date of Submission:	24/04/2019
Title Of Project	Data value analytics for an online retailer
Name of Supervisor.	Prof. Tomas Ward

An report submitted to Dublin City University, School of Computing for module CA684 – Machine Learning, 2018/2019.

I understand that the University regards breaches of academic integrity and plagiarism as grave and serious. I have read and understood the DCU Academic Integrity and Plagiarism Policy.

I accept the penalties that may be imposed should I engage in practice or practices that breach this policy.

I have identified and included the source of all facts, ideas, opinions, viewpoints of others in the assignment references. Direct quotations, paraphrasing, discussion of ideas from books, journal articles, internet sources, module text, or any other source whatsoever are acknowledged and the sources cited are identified in the assignment references.

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.

By signing this form or by submitting this material online I confirm that 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.

By signing this form or by submitting material for assessment online I confirm that I have read and understood DCU Academic Integrity and Plagiarism Policy (available at
http://www.dcu.ie/registry/examinations/index.shtml

Name(s): Nihar Sudhanshu Limaye

Date: 24/04/2019

Predicting Media Memorability

Nihar Sudhanshu Limaye
School of computing,DCU,Ballymun Road,Dublin 9
nihar.limaye3@mail.dcu.ie

ABSTRACT

Memorability is defined as the quality of being worth remembering. In today's multimedia age lots of video and images are created, shared and they are the intrinsic part of daily life. But with use of gadgets like smartphones, tablets and laptop we store those multimedia in storage devices. If we ask some person to describe an image or video he will say some important features about it that means he stores only key points from that image which helps him to memorise. So in my paper I will work on designing a systems that automatically predict memorability scores for videos, which reflect the probability of a video being remembered along with the features provided. The proposed dataset comes with "short-term" and "long-term" memorability annotations.

1 INTRODUCTION

Following the rapid expansion of the research field of image memorability prediction [1,2] the challenge has recently been extended to videos [5,6,7]. It is important to work on video memorability prediction as it will help to organize and retrieve digital content, to make it more useful in our daily lives. The problem is arising as media platforms, social networks, search engines, and recommender systems deal with growing amounts of content data day after day. As we can see the importance of the video memorability which will help in a large number of applications, e.g., education and learning, content retrieval and search, content summarization, storytelling, targeted advertising, content recommendation and filtering. Despite of having an active area in today's research in the computer vision community, VM prediction suffers from two main obstacles that were described in [5]. Firstly, among the previous attempts at predicting VM [5,6, 7] no clear definition of VM has been established, nor does a common and unified protocol for its measurement exist. Secondly, no large dataset available for building the models. The purpose of this task is therefore to propose a public benchmark to assess the memorability of videos, based on a publicly released largescale dataset and on an objective and clear measurement protocol.

2 RELATED WORK

Work on video memorability has recently began to generate a lot of interest, and recent works [2] [3] also investigate the use of various and high level visual features such as deep learning based action recognition representations (C3D-Preds), and image and video captions for memorability prediction. The major findings on memorability from these papers are that models using captions give the best individual results. Additionally, researchers have found that high level semantic features learned by CNNs trained for image classification achieve state of the art performance on a variety of computer vision tasks [9]. As the world is changing we have such tools which provides powerful parallel machines (GPUs, CPU clusters), together with large amounts of training data, Convolutional neural networks (ConvNets) have made a comeback providing breakthroughs on visual recognition. ConvNets have also been applied to the problem of human pose estimation in both images and videos [7].In this paper we will focus on C3D feature for our analysis.

3 DATA DESCRIPTION

The dataset consist of two parts i.e. dev-set and test-set totally composed of 8000.These 8000 videos were split into 6000 videos for the development set and 2,000 videos for the test set. They were extracted from raw footage used by professionals when creating content. The dev set contains the memorability scores for each video for short and long term memorability and test set only contains video name. After using ML algorithms we have to find the memorability values for test set. Each video also comes with its original title. These titles can often be seen as a list of tags (textual metadata) that might be useful to infer the memorability of the videos.

4 APPROACH

4.1 Loading ground truth

The dev set ground truth file is loaded and then it is analyzed by taking it's count, mean and standard deviation. After taking the statistics the graph is plotted to see the memorability trend for short-term and long term.

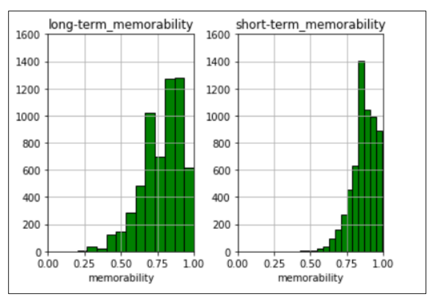


Fig 1. Graphical representation for short, long term memorability

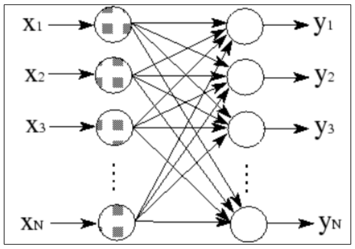
After that feature C3D is added into the data frame and on that basis model is trained to get results for test dataset.

4.2 Feature Decision

It gives best performance on different types of video analysis tasks. The feature from these 3D encapsulates information related to objects scenes and actions in a video, making them useful for various tasks without requiring to finetune the model for each task. C3D has the properties that a good descriptor should have it is generic, compact, simple and efficient.

4.3 Building network model

In this the keras library comes into the picture. The dataset is split into train and test for model training, A dense layer is just a regular layer of neurons in a NN. Each neuron receives input from all the neurons in the previous layer, thus densely connected.



The following is the docstring of class Dense from the keras documentation:

output = activation(dot(input, kernel) + bias)

where activation is the element-wise activation function passed as the activation argument, kernel is a matrix created by the layer, and bias is created by the layer.

4.3 Visualizing

The data is pass through the model for the analysis and then analyzed using graphs. As we can see the training loss is decreasing and validation loss is increasing at epoch 50.

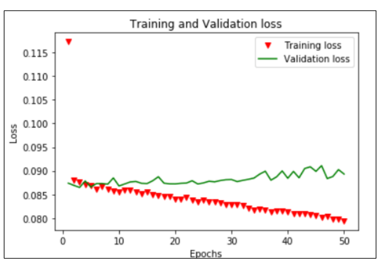


Fig 2. Train/Validation loss

In the next graph for the accuracy out training set gives more accuracy which is gradually increasing as compare to validation. So the model is almost predicting the values correctly.

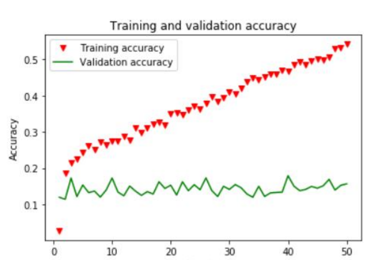


Fig.3 Training/validation accuracy

Let's examine the below graph of training value vs testing value graph. In training the values are more between 0.8-1.0 to 0.6-1 for short and long term respectively.

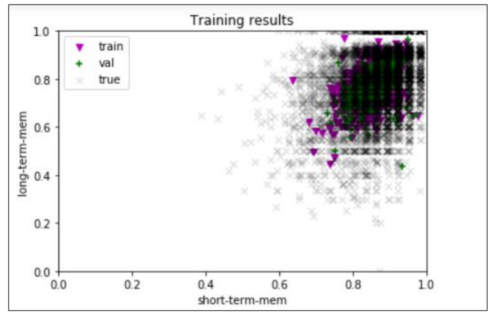


Fig 4. Training results

But after passing through model both short and long term values gives similar trend compare with training results. Hence we can say our model is predicting values correctly.

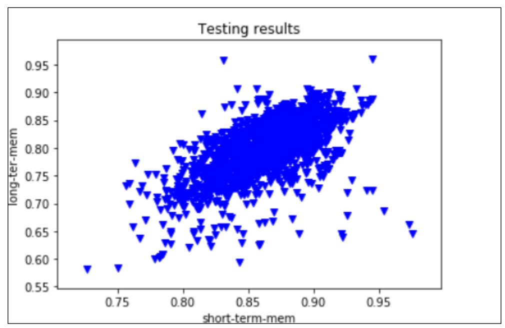


Fig 5. Testing Results

5 RESULTS AND ANALYSIS

After running the model the values which I got Spearman correlation is as follows.
Short-Term Spearman's correlation is : 0.256
Long-Term Spearman's correlation is : 0.099
As compare with previous models my model gives better result for dataset.

6 CONCLUSIONS

In this paper, I used C3D feature with keras to analyze the dataset provided to me. Especially, I found that the combination of the C3D and model gives impressive results, which demonstrated the effectiveness of the method.

REFERENCES

[1] Yoann Baveye, Romain Cohendet, Matthieu Perreira Da Silva, and Patrick Le Callet. 2016. Deep Learning for Image Memorability Prediction: the Emotional Bias. In Proc. ACM Int. Conf. on Multimedia (ACMM). 491–495.

[2] Jiri Fajtl, Vasileios Argyriou, Dorothy Monekosso, and Paolo Remagnino.2018. AMNet: Memorability Estimation with Attention.In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 6363–6372

[3] Phillip Isola, Jianxiong Xiao, Antonio Torralba, and Aude Oliva. 2011.What makes an image memorable?. In Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR). IEEE, 145–152.

[4] Aditya Khosla, Akhil S Raju, Antonio Torralba, and Aude Oliva. 2015.Understanding and predicting image memorability at a large scale. In Proc. IEEE Int. Conf. on Computer Vision (ICCV). 2390–2398.

[5] Romain Cohendet, Karthik Yadati, Ngoc Q.K. Duong, and Claire-Hélène Demarty. 2018. Annotating, understanding, and predictinglong-term video memorability. In Proc. of the ICMR 2018 Workshop,Yokohama, Japan, June 11-14.

[6] Junwei Han, Changyuan Chen, Ling Shao, Xintao Hu, Jungong Han,and Tianming Liu. 2015. Learning computational models of video memorability from fMRI brain imaging. IEEE Transactions on Cybernetics 45, 8 (2015), 1692–1703.

[7] Sumit Shekhar, Dhruv Singal, Harvineet Singh, Manav Kedia, and Akhil Shetty. 2017. Show and Recall: Learning What Makes Videos Memorable. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2730–2739.