Predicting Media Memorability Using Neural Network

Soumi Mitra
20210300
School of Computing
Dublin City University
Dublin, Ireland
soumi.mitra2@mail.dcu.ie

ABSTRACT

Memorability refers to the way or state of remembering something [1]. With the advancement of technology, the amount of video content generated per unit of time has increased exponentially. Some of these videos have a greater emotional effect than others, resulting in higher memorability ratings. Predicting memorability has several practical applications[2][3]. The organizers of MediaEval provided various pre-computed video based features and frame based features to calculate short-term and long-term memorability scores. In this project, I have used C3D and HMP features and sequential model (Neural Network) to calculate the scores.

1 INTRODUCTION

Social networking sites, media advertising, information processing, and recommendation systems all deal with increasingly growing data on a daily basis. To make multimedia events more relevant in our daily lives, we need new ways to coordinate – and, in particular, retrieve – digital material. Memorability, like other video importance indicators like aesthetics or interestingness, can be considered useful in deciding between competing videos. This is especially true when considering the specific use cases of commercial production and educational content development. Since the impact of different multimedia material, such as photographs or videos, on human memory varies, advertising professionals must be able to predict the memorability level of a given piece of content. The Mediaeval memorability challenge requires participants to predict video memorability ratings, which represent the probability of a video being remembered. Participants are given a broad data set of videos that have been pre-extracted and have memorability annotations, relevant content, and state-of-the-art visual features.[4]

2 LITERATURE REVIEW

I've referred to various previous papers on this study. In this predicting memorability task, people worked using various features like Captions, C3D, HMP, Inception etc. and they have used various machine learning models like Random Forest, Decision Tree, Linear Regression, Neural Network etc. They used the features separately sometimes and combined them sometimes to build the models and calculate the short-term and long-term memorability scores. In most of the cases, they have used video dedicated pre-computed features like C3D and HMP to calculate the scores. And the Neural Network model gives better scores than other models while using C3D and HMP. Hence, I have used Neural Network in this project. Also, in most of the cases, merging the features provided better scores than using the features individually. But sometimes there were exceptions too. Hence, I've used the features individually first and then combined to calculate the scores. The scores were calculated using Spearman's Correlation Function.

3 DATASET

The dataset contains 8000 soundless videos that are licensed for use and dissemination in the sense of the mediaeval memorability challenge. This dataset is further divided into a development set (train set) of 6000 videos and a test set of 2000 videos. The videos were created from raw footage used by experts in content creation. They are varied and contain various scene forms, each lasting 7 seconds. Each video also has its own unique title. These titles are sometimes interpreted as a list of tags (textual metadata) that can be used to infer the videos' memorability. The organizers to Mediaeval Memorability challenge provided some pre-computed features as well. Two of them are video-dedicated features: C3D and HMP. And others are frame-based features like color histogram, HOG, LBP, ORD, InceptionV3 and aesthetics.[6]

In this project, I've used C3D and HMP features which will be described in detail in the next section.

4 FEATURE DESCRIPTION

Considering that videos do have an effect during their runtime, but not because of a specific image in the frame. As a result, some of the image features derived from the beginning and end of the video did not appear to be sufficient features for predicting memorability. Hence, I only used C3D and HMP features that were extracted directly from the video.

C3D (Convolution 3D): C3D is obtained by training a deep 3D convolutional network on a large annotated video dataset. The dataset contains various concepts encompassing objects, actions, scenes and other frequently occurring categories in videos [9].

HMP (Histogram of Motion Patterns): HMP is a static image template which helps in understanding the motion location and path as it progresses. The temporal motion data is condensed into a single image template, where intensity is a function of motion recency, with brighter values indicating more recent motion [9].

5 MODEL

I've used a Keras Sequential Neural Network model in this project.

Keras is a powerful and easy-to-use free open-source Python library for developing and evaluating deep learning models. It wraps the efficient numerical computation libraries Theano and TensorFlow and allows us to define and train neural network models in just a few lines of code.[10]

The sequential model is a layer stack that runs in a straight line. Each layer has weights that match the weights of the layer above it.

I've used this model on C3D and HMP individually and also combining them together.

I've also checked the fitting of the model and it comes with a good fit.

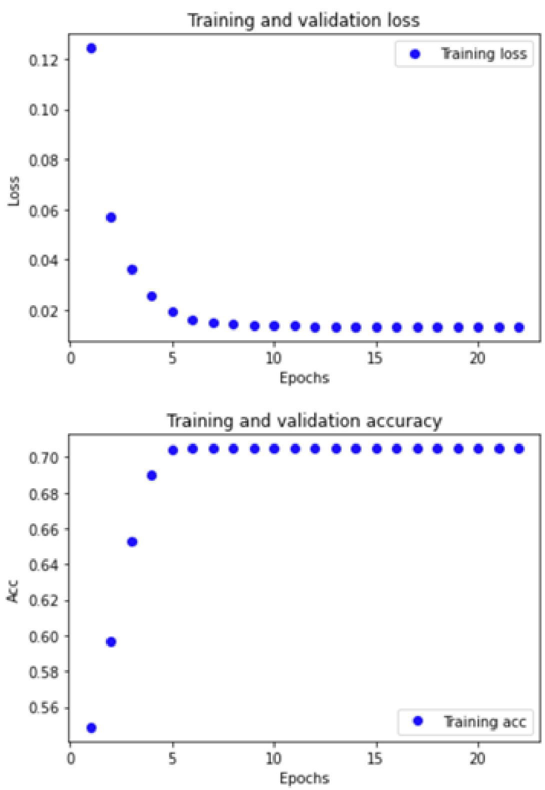

Figure 1: Checking fitting of the model

6 EVALUATION AND RESULT

The short-term and long-term memorability scores are calculated using Spearman's rank correlation coefficient. Spearman's rank correlation coefficient is a nonparametric measure of rank correlation (statistical dependence between the rankings of two variables) [8]. It assesses how well the relationship between two variables can be described using a monotonic function [8].

The formula to calculate the score is given by:

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

ρ = Spearman's rank correlation coefficient

d_i = difference between the two ranks of each observation

n = number of observations

Figure 2: Formula to calculate the Spearman's rank correlation coefficient

The results are tabulated as follows.

Feature	Short-term Memorability Score	Long-term Memorability Score
C3D	0.278	0.128
HMP	0.267	0.120
C3D + HMP	**0.292**	**0.146**

Table 1: Memorability scores for different features

7 CONCLUSION

This can be concluded from this project that combining the features gives better results than using them individually. Also, if we use the features individually, then C3D gives better scores than HMP.

However, for future work, I would like to experiment with frame-based features as well. And I would also like to use other Machine learning techniques to calculate the memorability scores.

ACKNOWLEDGMENTS

I would like to thank Prof. Tomas Ward and Mr. Eoin Brophy for all their support and guidance throughout the semester and this project implementation.

REFERENCES

[1] Definition of memorable | Dictionary.com. www.dictionary.com. Retrieved April 23, 2021 from https://www.dictionary.com/browse/memorability

[2] swathikiran86/Predicting-Video-Memorability-Scores. GitHub. Retrieved April 23, 2021 from https://github.com/swathikiran86/Predicting-Video-Memorability-Scores

[3] arunabellgutteramesh/VideoMemorabilityPredictionUsingML. GitHub. Retrieved April 23, 2021 from https://github.com/arunabellgutteramesh/VideoMemorabilityPredictionUsingML

[4] 2021. Predicting Media Memorability. MediaEval Benchmark. Retrieved April 23, 2021 from https://multimediaeval.github.io/editions/2020/tasks/memorability/

[5] Google Colaboratory. Retrieved April 23, 2021 from https://colab.research.google.com/drive/1X7l5MGrDZa2IdMCOxwgILCD5CzHKE8NF
[6] Romain Cohendet, Claire-Hélène Demarty, Ngoc Duong, Mats Sjöberg, Bogdan Ionescu, Thanh-Toan Do, and France Rennes. 2018. MediaEval 2018: Predicting Media Memorability Task. arXiv:1807.01052 [cs] (July 2018). Retrieved April 23, 2021 from http://arxiv.org/abs/1807.01052
[7] Spearman's Rank-Order Correlation - A guide to when to use it, what it does and what the assumptions are. Retrieved April 23, 2021 from https://statistics.laerd.com/statistical-guides/spearmans-rank-order-correlation-statistical-guide.php
[8] 2021. Spearman's rank correlation coefficient. Wikipedia. Retrieved April 24, 2021 from https://en.wikipedia.org/w/index.php?title=Spearman%27s_rank_correlation_coefficient&oldid=1013762125
[9] Google Colaboratory. Retrieved April 23, 2021 from https://colab.research.google.com/drive/1fNA8rIEGPYdbR5B4jipVPsirvWiEnBHf
[10] Jason Brownlee. 2019. Your First Deep Learning Project in Python with Keras Step-By-Step. Machine Learning Mastery. Retrieved April 23, 2021 from https://machinelearningmastery.com/tutorial-first-neural-network-python-keras/