

CSE445

Machine Learning

Meme Review: Dead or Alive?

- Predicting Reader Response to Timely Memes using Machine Learning

21st October 2018

Abir Rahman

*Dept. of Computer Science and Engineering
North South University
Dhaka, Bangladesh*

*abir.rahman01@northsouth.edu
ID - 1530412042*

Abdullah-Al Nahian Siraj

*Dept. of Computer Science and Engineering
North South University
Dhaka, Bangladesh*

*nahian.siraj@northsouth.edu
ID - 1520526042*

Jawad Aziz Khan

*Dept. of Computer Science and Engineering
North South University
Dhaka, Bangladesh*

*jawad.khan1@northsouth.edu
ID - 1530457042*

Taufiq Rahman

*Dept. of Computer Science and Engineering
North South University
Dhaka, Bangladesh*

*rahman.taufiqur@northsouth.edu
ID - 1521285042*

Meme Review: Dead or Alive?

Abir Rahman

*Dept. of Computer Science and Engineering
North South University
Dhaka, Bangladesh
abir.rahman01@northsouth.edu*

Abdullah-Al Nahian Siraj

*Dept. of Computer Science and Engineering
North South University
Dhaka, Bangladesh
nahian.siraj@northsouth.edu*

Jawad Aziz Khan

*Dept. of Computer Science and Engineering
North South University
Dhaka, Bangladesh
jawad.khan1@northsouth.edu*

Taufiq Rahman

*Dept. of Computer Science and Engineering
North South University
Dhaka, Bangladesh
rahman.taufiqur@northsouth.edu*

Abstract

In an era where internet community has become a staple, meme culture has seemingly dominated the populus' preference for expression in, everywhere from, forums to social media and has even gotten as far as influencing big companies to use memes to market their brand. But their short lifespan makes them an ever changing lexicon. Teaching a machine to estimate the timely relevance of a given meme could prove to be a novel method of predicting human behavior.

I. Introduction

Memes are a modern cultural phenomena that helps this generation express a vast array of sentiments albeit having short relevance life spans.

Our project is a classification problem where meme images are to be recognized by :-

- Running supervised model training
- Grouping images according to visual similarity and sorted by the month/year that it was popular
- Which can then be linearly regressed to predict what type of a meme might gain popularity / be relevant.

But the issues include outlier exceptions as memes are very diverse visually with slight format similarities and the fact that regular consumer GPUs are not as powerful for processing image matrices.

II. Problem Definition

With a growing number of social media platforms and social media influencers, feedback on content created by content creators have become crucial. Getting feedback after publishing is already done nowadays but what if there was a way to provide feedback based on a working formula even before the content is placed in front of the world?

Data is key in this ambition. With businesses racing to make their products and services 'viral'. The internet is scattered with such data that already exists. Some ventures have become massive success stories and others have flopped beyond recovery. This massive dataset that already exists with machine learning algorithms can be turned into a tool useful to many of the modern day content creators. ¹

III. Motivation

In today's modern day age, the internet is a massive ocean of digital data starting from important emails to cat videos. The pioneers of the internet could not have predicted the popularity of the term 'meme' to be taking over the throne. Humans have changed over time adjusting to their surroundings and have found several ways to entertain themselves. Memes have become a multi-million dollar entertainment platform where businesses such as '9GAG' have made a fortune. They are not only a matter of entertainment but rather

mainstream corporations have started using them as marketing tools to reach this new millennial generation. With this motivation, we plan to create a way to help this new generation of ‘meme’ creators to find the best outreach for their content.

IV. Background

Meme is an activity, concept, catchphrase, or piece of media that spreads, often as mimicry or for humorous purposes, from person to person via the Internet. An Internet meme usually take the form of an image, GIF or video. These small movements tend to spread from person to person via social networks and media.

Machine Learning is a subset of artificial intelligence where computer algorithms are used to autonomously learn from data and information. In machine learning, computers don’t have to be explicitly programmed but can change and improve their algorithms by themselves.

V. Data & Data Source

The data is composed of more than 3,000 meme images of varying resolutions and types downloaded from Reddit. An attached JSON Array file containing the images’ respective metadata is also found with the dataset. The JSON metadata contains vital information like title, time published, number of upvotes and

downvotes. The image dataset and its data were collected from Kaggle.com, uploaded by user sayangoswami on 17th April 2018.²

VI. Pre-processing

The 3,000 images come in varying resolutions so a resize preprocess is necessary for more accurate model training. Although that might cause stretching it will result in more easily comparable negatives that can be fed to the network. Finally the accompanying JSON metadata provides valuable information. Corresponding filenames are included in the metadata as random alphanumeric strings that should be renamed and indexed by some sorting method. And the time provided is in the UTC (Coordinated Universal Time) time standard and is in a UNIX-based timestamp which needs to be converted to an easily readable dd/mm/yy format.

VII. Algorithms

As this is an image classification problem we have considered taking advantage of the OpenCV platform for traditional computer vision but that proved to be too challenging as image analysis requires large datasets and very intense computation, best powered by a GPU. Hence finally TensorFlow³ is preferred as various Convolutional Neural Networks can be constructed with relative ease and has GPU support to power heavy computes. Image classification is best done by Neural Networks

or Deep Learning and hence that is what this project will attempt. Widely used networks are well implemented by TensorFlow thus the 3 Network algorithms below are planned to be implemented:

- Inception v3, Google's state of the art Image Recognition Convolutional Neural Network ⁴
- AlexNet, a CNN developed by Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton for the 2012 ImageNet Challenge. It is built to run on CUDA GPUs. ⁵
- VGGnet, is a deep convolutional network developed by Oxford's renowned Visual Geometry Group in 2013. ⁶

VIII. Contribution

Contributions of each member during the preparation of this report is equally valuable. Jawad worked on researching machine learning algorithms that would be useful to this project. Abir, Nahian and Taufiq sorted through numerous datasets to sort list a useable dataset perfect for this particular purpose. Everyone contributed to the writing of this report.

IX. References

1. A Study of Meme Propagation: Statistics, Rates, Authorities, and Spread - *Onkar Dalal, Deepa Mahajan, Ilana Segall, Meghana Vishvanath*

2. Reddit Memes Dataset: *A collection of the latest memes from the various meme subreddits* - *Sayan Goswami*
3. Is Google Tensorflow Object Detection API the Easiest Way to Implement Image Recognition? - *Priyanka Kochhar, Deep Learning Consultant, kdnuggets.com*
[https://www.tensorflow.org/tutorials/imag es/image_recognition](https://www.tensorflow.org/tutorials/imag/es/image_recognition)
4. Object Recognition with Google's Convolutional Neural Networks: Classifying Images Using Google's Pre-Trained Inception CNN Models - *William Koehrsen, medium.com*
5. ImageNet Classification with Deep Convolutional Neural Networks - *Alex Krizhevsky, Ilya Sutskever and Geoffrey Hinton, UofT NIPS*
6. Very Deep Convolutional Networks For Large-scale Image Recognition - *Karen Simonyan & Andrew Zisserman, Oxford arXiv and Google DeepMind*