Tiernon Riesenmy, Duc Trinh, Eli Lleshi

Professor Molontay

Data Science

02/27/2020

Analyzing and Predicting the Popularity of Memes

With our project, we want to better understand the humor of the internet. The basis for most of the internet’s humor is the making and sharing of memes. What makes this project so interesting is that we are able to better qualify what makes a meme humorous through data science and machine learning techniques.

We aim to predict the popularity of a meme based upon its contents. The program will take a meme as input and convert it into some more digestible data form, then predict the level of popularity it will receive. It is expected that we will find that meme popularity will depend upon the time it is posted and there exists a correlation to what other memes are popular in a certain time frame.

We plan to use datasets from sources such SNAP from Stanford and Kaggle. An example of a dataset we have found that could be used: [Reddit Memes Dataset](https://www.kaggle.com/sayangoswami/reddit-memes-dataset).

We will load and operate on the data using Pandas. There are tens of thousands of meme images (with likes/upvotes, timestamp) to be used online that we could gather. There are also extensive data sets of images classified by their contents and sentiments which may be cross-referenced to make connections and predictions.

There are tens of thousands of meme images (with likes/upvotes, timestamp) to be used. There are also extensive data sets of images classified by their contents and sentiments which may be cross-referenced to make connections and predictions.

For the preparation of the data, we want to make sure there are not too many missing data fields in our datasets. We also want to make sure that the links to the images/memes work. We can also reduce the size of the images/memes before the training step.

From our research, it seems most image classification tasks are done using convolutional neural networks. We probably would need to do external research to do this project.

We could have a small UI that lets you insert a new image, and we will tell you how many upvotes we predict it will get.