Utilizing Video Descriptions and Titles to Predict YouTube View Counts

Final Project for SI 630

Andrew Barber

Abstract

For YouTube content creators, the number of views their videos receive not only dictates the videos popularity, but also its ad revenue. For those with monetary investment in YouTube, the number of views a video receives directly correlates with its contribution to their livelihood, growth of their business or increase in their social media presence. By looking at the description of the video, it is possible to get a brief overview of the content and therefore estimate correlations between video content and number of views. This view-estimation algorithm can help guide creators on ideal content for maximum viewership. We used both the video description and the description + video title as input into a simple Long-Short Term Memory (LSTM) model in order to predict YouTube view counts. We found that when using both the video description and title in our model, we were able to predict view counts just as well as taking the median view counts per YouTube channel. This finding represents a jumping off point for future research in this area, as well as a powerful tool for new content creators looking to estimate the validity of their ideas.

1 Introduction

The main difficulty for content creators lies in how to tell a good idea from a bad one. Spending money and time on a video that is unlikely to reach a respectably large audience represents a waste of those resources. The ability to predict the number of views a video will receive would then be extremely valuable to content creators; not only would they be able to predict their future ad revenues before even releasing videos, they would also gain valuable insights regarding what type of content is successful and what is not. Natural language processing (NLP) makes this possible, looking at the video description and video title, we get a basic conception regarding the video content making it possible to represent videos through NLP.

The specific task of predicting video success using natural language processing is not often addressed in the literature. However, based on both the variable lengths of the YouTube video descriptions and titles along with the prevalence and success of these models historically, we decided that using a simple LSTM model would represent a logical first step for research on this topic. We also decided to implement a linear regression model as a baseline to compare the performance of neural networks to that of traditional machine learning techniques.

To establish a second strong baseline for this data, we used the median view count of a given YouTube channel to predict the view count of a given video belonging to that channel. When using only the video description as input, we found that neither the linear regression or the LSTM model could beat this baseline metric, with linear regression slightly outperforming LSTM. However, when using both the video description and title, the LSTM model achieved a similar score as the baseline, well outperforming the linear regression model. Our model demonstrates a positive outlook for future research in this area. It also could be used to predict view counts for novice content creators without established YouTube channels. Based on plausible video descriptions and titles, these creators would be able to ”test out” their ideas without expending valuable time and energy on ideas that are not likely to succeed.

2 Problem Definition and Data

In 2017, YouTube made 3.5 billion dollars in net advertising revenue (Statista, 2018). In 2018, that

Title	Description
Depeche Mode It's No Good	DANCA MUITO.
Farmer Walk 130kg Arnold Classic Australia	Arnold Classic Australia Qualifier 2016, Victorian Strongman Record Day, U105kg category. Best record, 21s 130kg
How to build a play gym for medium sized birds.	Hey, just wanted to help some of you guys out, if you r thinking of building a bird play gym. Ask questions in the comments. Hope this helps!

Table 1: Examples of video title and description found in dataset

number increased to 3.95 billion, a number that has been on the rise since YouTube's infancy. With all of that money on the table, video views are of the utmost importance to creators. More views mean more ad revenue, and so tailoring content to maximize viewership is the primary way creators can increase their own earnings.

The model we are proposing is meant to give content creators a useful tool to accurately predict and maximize their earnings. Beyond just earning more revenue, creators will also be able to plan out their finances. The content creator occupation is very volatile; our tool can therefore be used to plan out both their content and expected revenue. This can give them more financial security and peace of mind in their profession. By comparing the root mean square errors (RMSE) of our baseline models to our LSTM model, we can determine which is better at predicting YouTube view counts. Achieving a lower RMSE value than the median view count by channel ID means that content creators can get more accurate predictions with our tool than just by estimating future view counts based on the view counts typical for their videos. Beating this baseline will therefore give credence to our model while beating the linear regression baseline will demonstrate the utility of neural networks for this problem compared to traditional machine learning.

Our data for this project was adapted from another featured in *Towards Data Science* (Allen Wang and OFarrell, 2017) which in turn was taken from the YouTube 8m dataset (YouTube, 2017). The dataset we will use focuses on fitness-related videos - 92,459 videos in total. The video title, description, view count, like/dislike count and channel are some of the features included in the dataset. (Examples of video descriptions and titles can be seen in Table 1 above.) We plan to use both the description and title of these videos to predict the number of views the corresponding video will accrue. The log-distribution of view counts can be seen in Figure 1 below.

3 Related Work

The paper titled ”Predicting Sales from the Language of Product Descriptions by Pryzant et al. (Reid Pryzant and Jurafsky, 2017) shares many similarities with our study. Although they used a recurrent neural network (RNN) + gated feedback (GF) model in their approach to predicting sales, the idea behind their study is very similar: predicting the success of an item based on its description. The approach used in this paper is therefore especially relevant to our own methodology, with the RNN + GF model representing a possible future direction for YouTube view count prediction. More similar to our own methodology is ”Predicting Polarities of Tweets by Composing Word Embeddings with Long Short-Term Memory” (xwang, 2015). According to this paper, using RNNs and LSTMs are very popular approaches in the realm of NLP and have many possible applications including our own task of interest.

Although many studies, such as those listed above, suggest that RNNs are the best method for regression problems in NLP others insist that linear regression produces the best results. The paper titled ”Movie Reviews and Revenues: An Experiment in Text Regression” looked specifically at movies as a product(Mahesh Joshi and Smith, 2017). They decided to use machine learning as their method, specifically linear regression. Also using a machine learning approach, the article titled ”Improving Movie Gross Prediction through News Analysis” used regression and k-nearest-neighbors methods along with news data to predict the gross revenues of movies(Zhang and Skiena, 2009). They claimed that using news data alone, they were able to match the performance of models using IMDB data. This article shows a potential for external data, such as news articles, to predict the success of audio-visual media such as movies. These studies achieved excellent results using linear regression. Our study will therefore compare a simple LSTM model to a linear regression model in order to determine in neural networks are a promising direction for the task of video view count prediction.

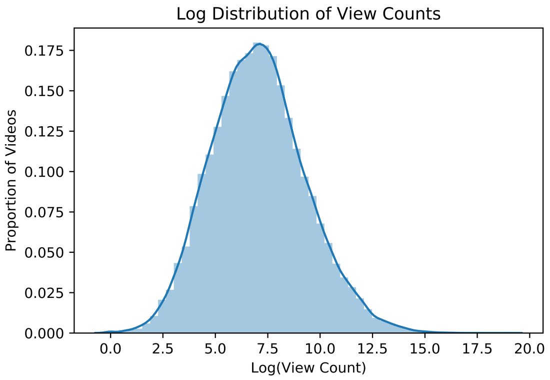

Figure 1: Log distribution of view counts for YouTube fitness video data.

4 Methodology

As mentioned above, we used both the video description and title as input for our model. In order to prepare these pieces of raw text, we tokenized the text. We also converted the text into lower case characters and removed all stop words. Our linear regression baseline model used an input of TFIDF vectors to represent the whole description and also the whole description + title. For the description + title, we simply concatenated the two strings.

For our LSTM model, we limited the number of tokens used as input to 25 - this created a fixed-length representation of the input which we believe was long enough to properly represent the description or description + title. In order to transform our description and title into usable input for our LSTM model, we created customized word embeddings using the vocabulary from both the video descriptions and titles. Input that fell below the 25-token limit was zero-padded. Description + title input was also limited to 25 tokens to maintain the speed of the model.

Our neural network model consisted of two LSTM layers and a hidden layer - representing a very simple LSTM implementation. Both LSTM layers utilized dropout at rates of 0.5. Our linear regression model was also very simple, with no hyper parameter tuning used. We hoped that this would more accurately compare the validity of either approach to our task of interest. Both models were trained on 80 percent of the data and tested on 20 percent.

5 Evaluation and Results

All models were evaluated on how well they could predict the natural log of the YouTube video view counts in the test data. The error in estimations were determined using the RMSE of the predictions compared to the true values.

The three models we evaluated include: the first baseline model calculated by using the median view counts of the channel a given video belongs to, the second baseline model using a simple linear regression model and the experimental model using a simple two-layered LSTM model. Upon evaluating these models, we found that the first baseline was very difficult to beat. When using only the video description as input, the LSTM model under-performed both baselines models. However, when utilizing both the description and title, we found that our experimental model outperformed the linear regression model and had a similar RMSE value to the first baseline (see Table 2). For a more detailed view of our model's performance see Figure 2 for a comparison of our model's predicted values versus true values.

6 Discussion

Our results show that it is possible to predict YouTube video view counts using a simple LSTM model just as well as using median view counts for a specific video's channel. This result gives a promising outlook for future research in this area. It is interesting, however, that positive results were only achieved when using both the video description and title as input for the model. It is especially interesting that although the linear regression model's performance increased by a small amount when using both the description and title as input, the LSTM model increased significantly with the added input. This may suggest that there is some information that the model learned through the combination of the video description and title or perhaps through just the title itself. We speculate that through the title information, the model was able to learn certain attributes that identify the YouTube channel. This would make sense, as the baseline model uses only channel-level information and performs very well, it would then follow that learning this feature would give the model a boost in performance. This assumption also seems viable given the nature of YouTube video titles that will usually include names of people appearing on the channel or even the name of the channel itself.

If the model is indeed learning channel-level information and that is the driving force behind its

Model	RMSE
Baseline - Median ViewCount by Channel	1.17
Baseline - Linear Regression: Description	1.94
Baseline - Linear Regression: Description + Title	1.86
LSTM: Description	2.03
LSTM: Description + Title	1.13

Table 2: Evaluation of models

boost in performance, this certainly dampens the utility of the model in real-world situations. By not using channel-level information in our model we hoped to increase its usefulness for content creators without an established online presence. If information regarding a video's channel ID is the main feature behind our model's success, then new content creators will not be able to expect as accurate a prediction for their video concepts as demonstrated by our model's performance. However, even if it is true that YouTube channel information plays a large part in our model's success, it can still be useful for novice content creators; if channel-level information is captured through the video's title, then this still means that video descriptions + titles that are predicted to do well closely match video descriptions + titles from successful channels and so are more likely to draw from that audience and in turn, draw in a higher view count. Although our model does not perform significantly better than determining the median performance of the other videos in a given channel, and so does carry significant utility for established content creators, we still believe that content creators without established channels can utilize our model to get an estimate of how popular their idea might be.

It should also be mentioned that based on the distribution of predicted values of our experimental model using both description + title as input seen in Figure 2, there is a pattern the errors follow compared to the true values indicated by the bisecting line plot. For true view counts that are lower, the model tends to overestimate the view count while with true view counts that are higher, the model tends to underestimate the view count. This should also be taken into consideration when using the model as newer creators are more likely to start off at lower view counts but the model will tend to overestimate their popularity.

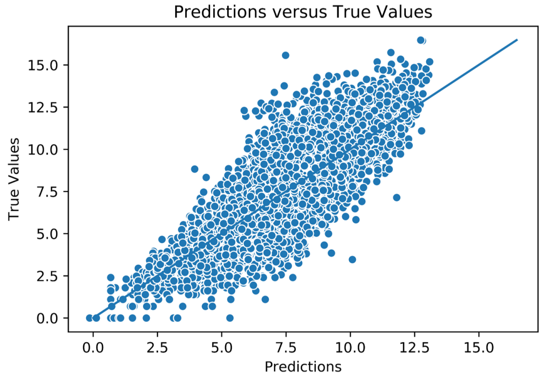

Figure 2: Predictions versus true values for best-performing experimental model

7 Conclusion

In this study we have demonstrated that a simple two-layer LSTM model using fitness-related YouTube video descriptions and titles as input can achieve an RMSE metric similar to that achieved by looking at median view counts per YouTube channel. This represents a promising jumping-off point for future research in this area as well as a promising model for content creators just starting out. Future research in this area may include looking at different models such as the RNN + GF model proposed by Pryzant et al. (Reid Pryzant and Jurafsky, 2017). Incorporating channel-level information is also a promising avenue and would be useful specifically for established content creators.

8 Other Things We Tried

The most significant variations we explore for this project include varying the length of the input and the dropout rate of the LSTM layers in our model. We first tried embedding lengths of 100 and 50 features, both of which took an extremely long time to train. We found that a length of 25 features represented a good preliminary model for this task and using longer inputs may represent an interesting future direction. While experimenting with the dropout rate of the LSTM layers, we found that a rate of 0.5 performed the best and so used that in our model.

Another avenue we explored had to do with available data. The original vision for the project was to take the closed captioning of videos and use that as input for our model. We also planned on looking at a broader number of categories than just fitness-related videos. However, available data, limited this type of exploration and so we decided to represent the videos using both their description and title.

9 What You Would Have Done Differently or Next

As mentioned above, we originally hoped to use closed captioning to represent YouTube videos rather than the video description + title. With available data, I believe that this would be a promising future direction for predicting video view counts. As I also mentioned, looking at different categories would represent an interesting avenue for research.

An unfortunate mistake may also have affected our methodology. Initially we calculated the baseline metric of average view count per channel to be much higher than in reality. This mistake artificially inflated our results and made our model seem more successful than it actually was. It is possible that our model still could have handily beat the baseline with more tweaking, however, this mistake was caught at a late stage in development and so stifled any further changes to our methodology.

10 Work Plan

Originally, the plan for our project was to predict book sales based on their contents. However, due to availability of pricing data, that idea quickly fell flat. We then decided to switch gears to predicting YouTube video view counts based on their content (in the form of closed captions); a similar idea but more feasible due to higher availability of YouTube data. However, this too was unfeasible due to unavailable data and we instead decided to represent YouTube videos with their description and title. Looking back, I believe that it is important to make sure to have all needed data for a specific project first, before developing a research idea any further. Even though a concept may seem feasible, it does not mean that all necessary resources can be easily accessed.

References

Aravind Srinivasan Kevin Yee Allen Wang and Ryan OFarrell. 2017. Youtube views predictor a comprehensive guide to getting more views on youtube backed by machine learning. *Towards Data Science* .

Dipanjan Das Kevin Gimpel Mahesh Joshi and Noah A. Smith. 2017. Movie reviews and revenues: An experiment in text regression. *Boston Globe* 461(154):116.

Young-joo Chung Reid Pryzant and Dan Jurafsky. 2017. Predicting sales from the language of product descriptions. *In Proceedings of the SIGIR 2017 Workshop on eCommerce (ECOM 17)* .

Statista. 2018. Net advertising revenues of youtube in the united states from 2015 to 2018 (in billion u.s. dollars). pages https://www.statista.com/statistics/289660/youtube–us–net–advertising–revenues/.

lyc cjsun-wangxl xwang. 2015. Predicting polarities of tweets by composing word embeddings with long short-term memory. *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* page 13431353.

YouTube. 2017. Youtube-8m dataset. Data received from: https://research.google.com/youtube8m/.

Wenbin Zhang and Steven Skiena. 2009. Improving movie gross prediction through news analysis. *Web Intelligence and Intelligent Agent Technologies* 1:301–344.