Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X
IMPACT FACTOR: 6.199

IJCSMC, Vol. 8, Issue. 4, April 2019, pg.103 – 107

Video Analysis with Image Recognition in TensorFlow

Avijeet Jha[1]; Ashish Jha[2]; Amarjeet Kushwaha[3]; Deepak Aeloor[4]

[1]Department of Computer Engineering, St. John College of Engineering and Management, India
[2]Department of Computer Engineering, St. John College of Engineering and Management, India
[3]Department of Computer Engineering, St. John College of Engineering and Management, India
[4]Department of Computer Engineering, St. John College of Engineering and Management, India
[1] avijeet.jha41@gmail.com; [2] ash98jha@gmail.com; [3] amark1995.jeet@gmail.com; [4] deepakaeloor@gmail.com

Abstract— Video analytics, loosely defined as autonomous understanding of events occurring in a scene. The use of deceptive techniques in user-generated video portals is ubiquitous. Unscrupulous up loaders deliberately mislabel video descriptors aiming at increasing their views and subsequently their ad revenue. This problem, usually referred to as "click-bait," may severely undermine user experience. In this work, we study the click-bait problem on YouTube by collecting metadata for 206k videos. To address it, we devise a deep learning model based on variation auto-encoders that supports the diverse modalities of data that videos include. Every click earns them display advertisement revenue. Social media users who are tricked into clicking may experience a sense of disappointment or agitation, and social media operators have been observing growing amounts of click bait on their platforms. As largest video-sharing platform on the web, YouTube, too, suffers from click bait. Thus, it is susceptible to recommending misleading videos to users.
Keywords— Video analysis, Clickbait, Sentiment analysis, Machine Learning, Transfer Learning

I. INTRODUCTION

As we all know YouTube has become now the largest source for movie, songs, news streaming [1]. The Cable TV has now become secondary source for entertainment. This is because YouTube offers a vast number of videos, which are always available on demand. However, because videos are generated by the users of the platform, known as You Tubers, a plethora of them are of dubious quality. The ultimate goal of You Tubers is to increase their ad revenue by ensuring that their content will get viewed by millions of users. Several You Tubers deliberately employ techniques that aim to deceive viewers into clicking their videos. These techniques include: (i) use of eye-catching thumbnails, such as depictions of abnormal stuff or attractive adults, which are often irrelevant to video content; (ii) use of headlines that aim to intrigue the viewers; and (iii) encapsulate false information to either the headline, the thumbnail or the video content. We refer to videos that employ such techniques as click-baits. The continuous exposure of users to click-baits cause frustration and degraded user experience (see Fig. 1). The click-bait problem is essentially a peculiar form of the well-known spam problem [2], [3], [4], [5], [6]. In spam, malicious users try to deceive users by sending them misleading messages mainly to advertise websites or perform attacks (e.g., phishing) by redirecting users to malicious websites. Nowadays,

the spam problem is not as prevalent as a few years ago due to the deployment of systems that diminish it. Furthermore, users have an increased awareness of typical spam content (e.g., emails, etc.) and they can effortlessly discern it. However, this is not the case for click-bait, which usually contains hidden false or ambiguous information that users or systems might not be able to perceive. Comments that were found in click-bait videos. The users frustration is apparent (we omit users' names for ethical reasons).There are basically two models for obtaining the final result in our system (I) Using thumbnail of the video and matching [8] it with the video frames divide on per second rate of the video length and (ii) sentiment analysis on the comments of the video extracted from the video URL [9].We propose a deep generative model that allows for combining data from as diverse modalities as video headline text, thumbnail image and tags text, as well as various numerical statistics, including statistics from comments. Most importantly, the proposed model allows for successfully addressing the problem of learning from limited labelled samples and numerous unlabelled ones (semi-supervised learning). This is achieved by postulating a deep variation auto-encoder that employs a finite mixture model as its encoder. In this context, mixture component assignment is regulated via an appropriate gating network; this also constitutes the eventually obtained classification mechanism of our deep learning system. We provide a large-scale analysis on YouTube; we show that, with respect to the collected data, its recommendation engine does not consider how misleading a recommended video is. Hence, it is susceptible to recommending click-bait videos to its users.

II. LITERATURE REVIEW

2.1. Impression Network for Video Object Detection

The paper [1] is inspired by how human utilize impression to recognize objects from blurry frames, they propose Impression Network that embodies a natural and efficient feature aggregation mechanism. These methods treat single image recognition pipeline as two stages: 1. the image is passed through a general feature network; 2. the result is then generated by a task-specific sub-network This enables successful detection on low-quality frames, but the aggregation cost can be huge thus further slows down the framework. In this work, we combine the advantages of both tracks.

2.2. An 10Sent: A Stable Sentiment Analysis Method Based on the Combination of Off-The-Shelf Approaches

The authors of paper [2] propose to combine several very popular and effective state-of-the-practice sentiment analysis methods, by means of an unsupervised bootstrapped strategy for polarity classification. Their main goals is to reduce the large variability (lack of stability) of the unsupervised methods across different domains (datasets). The experimental results demonstrate that their combined method (aka, 10SENT) improves the effectiveness of the classification task, but more importantly, it solves a key problem in the field.

2.3. Should I use TensorFlow

The paper [3] is based on evaluation of TensorFlow and its potential to replace pure Python implementations in Machine Learning The rapidly growing field of Machine Learning has been gaining more and more attention, both in academia and in businesses that have realized the added value. By utilizing TensorFlow, one can thus overall describe a model more expressively, with less effort and with less potential errors.

2.4. Sentiment Analysis of Review Datasets using Naïve Bayes' and K-NN Classifier

This paper [4] mainly describes Bayesian network classifiers are a popular supervised classification paradigm. A well-known Bayesian network classifier is the Naïve Bayes' classifier is a probabilistic classifier based on the Bayes' theorem, considering Naïve (Strong) independence assumption. It was introduced under a different name into the text retrieval community and remains a popular(baseline) method for text categorizing, the problem of judging documents as belonging to one category or the other with word frequencies as the feature.

III. METHODOLOGY

By Youtube's Data API between December and January 2018-19 we collected data published between 2010 to 2018. Specifically, we collected the following data descriptors for 206k videos: (i) thumbnail; (ii) comments from users; (iii) statistics (e.g., views, likes, etc.); and (iv) related videos based on YouTube's recommendation system. We started our retrieval from a popular (400M views) click-bait video [1] and iteratively collected all the related videos as were recommended by YouTube. Note that this approach enables us to study interesting aspects of the problem, by constructing a graph that captures the relations (recommendations) between videos. An important prerequisite for the construction of a valid corpus is to draw a representative sample of documents from the underlying population. YouTube offers little to no help in this regard, since neither its web front end nor its APIs allow for enumerating all videos available, nor tapping into the stream of videos uploaded every day. If not for the recently released YouTube 8M dataset (Abu-El-Haija et al. 2016), which has been constructed

by researchers working at YouTube, we would be left with no choice but to crawl YouTube ourselves. Below, we briefly review the construction and original purpose of the YouTube 8M dataset, describe our efforts to augment the dataset with the meta data necessary for clickbait detection (rendering the dataset also useful for tackling other research questions), and give a brief overview of the corpus statistics. Altogether, the resulting corpus compiles (if available) the meta data, comments, thumbnails, and subtitles ("captions") of 6,192,353 videos in a unified format, which we make available to other researchers on request.2 Table 1 gives an overview of the corpus.

Category

Table 1 reports the categories we find on the videos. In total, we find 15 categories but we only show the top five in terms of count for brevity. We observe that most clickbait's exist in the Entertainment and Comedy categories, whereas non-click baits are prevalent in the Sports category. This indicates that, within this dataset, YouTubers employ clickbait techniques on videos for entertainment.

Headline

YouTubers normally employ deceptive techniques on the headline like the use of exaggerating phrases. To verify that this applies to our ground truth dataset, we perform stemming to the words that are found in clickbait and non-clickbait headlines. Fig. 2 (a) depicts the ratio of the top 20 stems that are found in our ground truth clickbait videos (i.e., 95% of the videos that contain the stem "sexy" are clickbait). In essence, we observe that magnetizing stems like "sexy" and "hot" are frequently used in clickbait videos, whereas their use in non-clickbait's is low. The same applies to words used for exaggeration, like "viral" and "epic".

Thumbnail

To study the thumbnails, we make use of Imagga [19], which offers descriptive tags for an image. We perform tagging of all the thumbnails in our ground truth dataset. Fig. 2(b) demonstrates the ratio of the top 10 Imagga tags that are found in the manually reviewed ground truth. We observe that clickbait videos typically use sexually-appealing thumbnails in their videos in order to attract viewers. For instance, 81% of the videos' thumbnail of which contains the "pretty" tag are clickbait's.

Tags

Tags are words that are defined by YouTubers before publishing and can dictate whether a video will emerge on users' search queries. We notice that clickbait's use specific words on tags, whereas non-clickbait's do not. Fig. 2 (c) depicts the ratio of the top 20 stems that are found in clickbait's. We observe that many clickbait videos use tags like "try not to laugh", "viral", "hot" and "impossible"; phrases that are usually used for exaggeration.

Statistics

Fig. 2 (d) shows the normalized score of the video statistics for both classes of videos. Interestingly, clickbait's and non-clickbait's videos have similar views; suggesting that viewers are not able to easily discern clickbait videos, hence clicking on them. Also, non-clickbait videos have more likes and less dislikes than clickbait's. This is reasonable as many users feel frustrated after watching clickbait's.

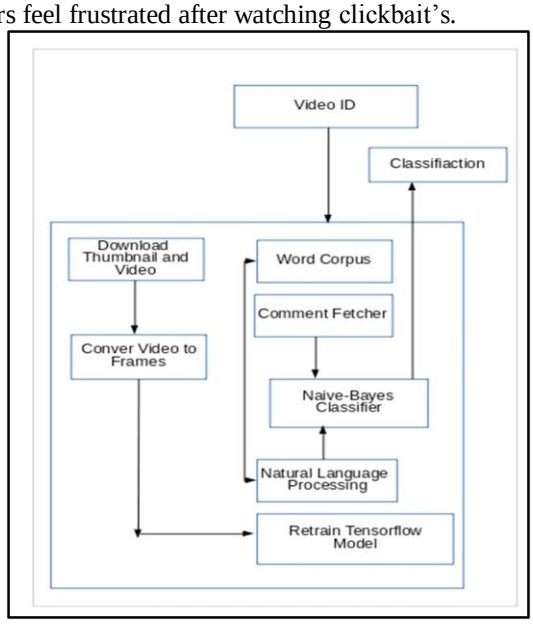


Fig 1 Architecture Diagram

Video ID
This is unique video id which is given by user.

Comment Fetcher
This is used to fetch comments of the given video for sentiment analysis.

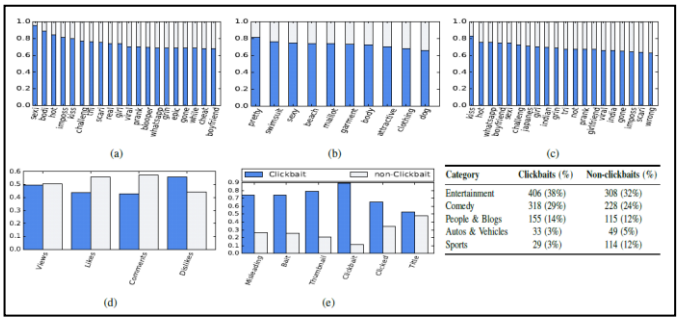
Word Corpus
It is document of huge wordlist containing positive and negative words to train the model.

Retrain TensorFlow Model
This is used to retrain the TensorFlow model using frames from video.

Natural Language Processing
This is used to perform many Natural Language Processing operation like tokenizing, vectorization and lemmatization.

Image recognition
The TensorFlow model is used to recognize the thumbnail against the video. The accuracy of prediction will be more if thumbnail is part of video. If thumbnail is not in video, then content and thumbnail are different.

IV. RESULTS



Category	Clickbaits (%)	Non-clickbaits (%)
Entertainment	406 (38%)	308 (32%)
Comedy	318 (29%)	228 (24%)
People & Blogs	155 (14%)	115 (12%)
Autos & Vehicles	33 (3%)	49 (5%)
Sports	29 (3%)	114 (12%)

Fig 2 & Table 1

Analysis of the manually reviewed ground truth dataset. Normalized mean scores for: (a) stems from headline text; (b) tags derived from thumbnails; (c) stems from tags that were defined by uploaders; (d) video statistics; and (e) comments that contain words for flagging suspicious videos. Table 1 shows the top five categories (and their respective percentages) in our ground truth dataset.

Source	Destination	Norm. Mean
clickbait	clickbait	4.1
clickbait	non-clickbait	2.73
non-clickbait	clickbait	2.75
non-clickbait	non-clickbait	3.57

Table 2

V. CONCLUSION

In this work, we have explored the use of variational auto-encoders for tackling the click-bait problem on YouTube. Our approach constitutes the first proposed semi-supervised deep learning technique in the field of click-bait detection. This way, it enables more effective automated detection of click-bait videos in the absence of large-scale labelled data. Our analysis indicates that YouTube recommendation engine does not take into account the click-bait problem in its recommendations.

VI. ACKNOWLEDGEMENT

We thank our guide, Mr. Deepak Aeloor who has extended all valuable guidance and help through various stages for the development of the project. Her Valuable suggestions were of immense help throughout the project work.

We convey our sincere regards to our respected principal Dr. G.V. Mulgund and Head of Department Dr. G.A. Walikar for their valuable support.

REFERENCES

[1] Piperjaffray. Survey, 2016. http://archive.is/AA34y.
[2] G. Stringhini, M. Egele, A. Zarras, T. Holz, C. Kruegel, and G. Vigna. B@bel: Leveraging Email Delivery for Spam Mitigation. In USENIX Security, 2012. U. Yabas, H. Cankaya, T. Ince. 2012. "Customer churn prediction for telecom services." 2012 IEEE 36th Annual computer software and application conference, Nov.2012.
[3] G. Stringhini, T. Holz, B. Stone-Gross, C. Kruegel, and G. Vigna. BOTMAGNIFIER: Locating Spambots on the Internet. In USENIX Security, 2011.
[4] G. Stringhini, C. Kruegel, and G. Vigna. Detecting spammers on social networks. In CSA, 2010.
[5] H. Gao, Y. Chen, K. Lee, D. Palsetia, et al. Towards Online Spam Filtering in Social Networks. In NDSS, 2012.
[6] N. Jindal and B. Liu. Review spam detection. In WWW, 2007.
[7] S. Zannettou, S. Chatzis, K.Papadamou. The Good, the Bad and the Bait: Detecting and Characterizing Clickbait on YouTube, 2018. [Online] [Accessed:Nov 15,2018]
[8] C. Hetang,H. Qin,S. Liu,J. Yan . Impression Network for Video Object Detection, 2017.
[9] P. Melo, M. Gonçalves, F. Benevenuto.10Sent: A Stable Sentiment Analysis Method Based on the Combination of Off-The-Shelf Approaches, 2018.
[10] Click-bait video. https://www.youtube.com/watch?v=W2WgTE9OKyg. [Online] [Accessed: Aug 7, 2018]

