Khan et al LGURJCSIT 2018

LGU International Journal for
Electronic Crime Investigation

LGU (IJECI)
ISSN: 2522-3429 (Print)
ISSN: 2616-6003 (Online)

Vol. 2 Issue 4, October - December 2018

Spammer Detection: A Study of Spam Filter Commentson YouTube Videos

Rafaqat Alam Khan
Lahore Garrison University
rafaqatalam@lgu.edu.pk

Abstract:

This paper presents a methodology to find out the spam comments on YouTube videos. The purpose of this research is to find out the comments of those spam users, who comment for their own promotional intentions or to detect users whose comments that have no relevancy with the given video.The monetization policy introduced by YouTube for its user's channel and advertisement of different ads on YouTube videos has attracted a large number of users. This increase in a large number of users has also lead to an increase in malicious users whose job is to create automated bots for commenting and subscription to different YouTube channels. These malicious users'comments hurt the channel publicity and also affect the normal user's experience. YouTube is also working on this issue by using different methods to limit these kinds of automated bots malicious comments by blocking those comments. These kinds of methods are ineffective so far as spammers have found out different methods to bypass those heuristic approaches. Different machine learning approaches provide somehow better classification accuracy with the introduction of new approaches to solve it better than that. In this work, different techniques used for classification of spam comments with those of normal user comments to improve the classification and recent trend going on in this area are briefly analyzed to tackle this major issue.

Keywords: Spam, YouTube, Classification, Comment, Social Media

1 Introduction:

YouTube a video sharing website was started back since 2005. In 2006, Google bought YouTube and nowadays it is on Google Subsidiaries. After it came under the umbrella of Google YouTube growth as a YouTube, video sharing has increased significantly. The users using YouTube can create their own Gmail login and through this Gmail login, they can create their own channel. Once the channel is created, YouTube allows users to publish his own video, rating, Comments, likes or dislikes, reporting and subscribing to your favorite channels. According to the recent statistics of YouTube, it has achieved the marked of 1 billion users log in. The global research [1] statistics of YouTube claims that around 1.9 billion users visit within a month, watching billions of hours per day and in turn generating billion of views per day.The 70 % of these watch hours come from mobile devices. Around $ 2 billion dollars, YouTube has paid over the last five years to its YouTube channel owners.

Out of different YouTube features for its channel users, YouTube commenting feature is one of the important features in which users can able to comment on other channels uploaded videos. This powerful feature of YouTube allows the interaction of YouTube channel owner with other users. Introducing of such feature has also allowed other malicious users to promote their promotional content to know as a spam comment. These spam comments are usually irrelevant and are generated by mostly automated bots. The capability of bots to perform spam comments on YouTube videos has been discussed in [2].

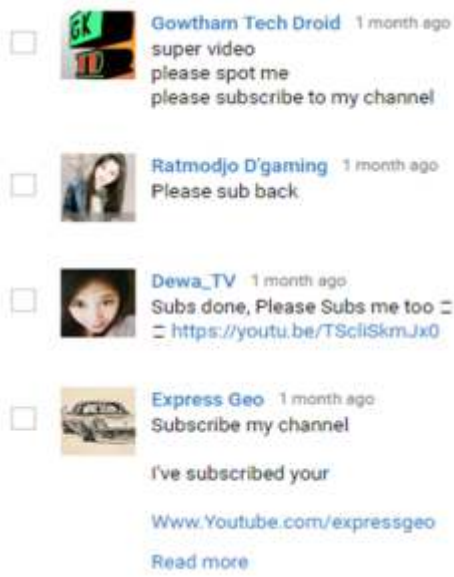

Fig.1. Example of Spam Comments on YouTube VideosA recent report [3] regarding spam comments has been published in BBC, in which YouTube claims that during the period of July-September 2018 they have almost deleted 224 million malicious comments out of billion of comments published in this 3 months period. They have also accepted that this a really hard work to tackle this issue and they are investing in this area to get rid of that malicious content.

The spamming can be basically divided into two categories i.e. link-based spam and channel-based promotion spamming. Link-based spamming is that type of spamming in which users comments different links comment on published YouTube videos, on which other users click than it redirected to malicious webpage's. While on the other hand, channel-based spamming is used for promotion of user channels to get subscribers, views, and likes for his channel as well as for his channel videos. YouTube generally use comment blocking [4] on the basis of HTTP URL, but this form of filtering has to lead the spammer a new approach to publish their links by insertion of white spaces in their link and instead of .com they typed dot com in the form of text to publish their link on the given video. These problems have to lead to the machine learning approaches to resolve such issues. The contribution in this research is finding out of better spam classification using different machine learning approaches with high F1 score and better classification accuracy.

2 Related Work

Chowdhury Rashid et al. [5] generated a lift chart by using three different data mining models. This lift chart finds out the lift score when compared to a random guess. The predicted probability for the three different data mining models i.e. Naïve Bayes, Decision Tree and Clustering is calculated. From the result, they have concluded that for most cases Naïve Bayes and decision tree performed better than that of clustering.

Tulio C. Alberto et al. [6] used different classification algorithms i.e. Naïve Bayes, Decision tree, SVM, Random forest and logistic regression on five different datasets. They have achieved a confidence level of almost 99 % on all these classifiers. Based on these classification results they have developed their own online tool known as TubeSpam that automatically detects the spam message on the fly.

SaumyaGoyal et al. [7] spam message detection on real twitter social media dataset is applied using KNN and decision tree. The WEKA tool is used and the metrics used for classification are precision-recall, F measure and class, FP rate and TP rate.

SimranKanodia et al. [8] suggested a Markov Decision process for YouTube spam message detection and the result is compared to other data mining tools used in this field. The Markov Decision accuracy is 78.82 % which is quite better than those of other data mining algorithms out of which the maximum accuracy is obtained through the random forest which is 72.52 %.

Abdullah O. Abdullah et al [9] WEKA tool and python code is used for the employment of different classification algorithms on YouTube dataset that was generated using YouTube API. All the different 9 algorithms used have almost 90 % and above than that accuracy. Out of these different 9 algorithms, accuracy Adaptive Genetic Algorithm has performed quite well and achieved an accuracy of 99.1 %.

ShreyasAiyar et al. [10] in this they have used different machine learning algorithms along with custom approaches i.e. N-Grams. For automated detection of spam messages on YouTube videos they have suggested that the character gram approach performed better result as compared to word gram for obtaining better classification accuracy.

3. Outline of Spam Detection

S.No	Research Article	Techniques Used	Available Dataset	Results
1	N-Gram Assisted YouTube Spam Comment Detection	RF,SVM, Naïve Bayes,N-Gram	13000 Comments https://developers.google.com/youtube/v3/docs/commentThreads#Retrieve_comments [11]	N-Gram Outclassed Other Machine Learning Algorithms
2	A Comparative Analysis of Common YouTube Comment Spam Filtering Techniques	AGA, ICA-Amuse, ELM - AE, ANN-BP, SVM-K, K-NN, LR, NBC, DT	100 Channels having 10,000 Samples https://developers.google.com/youtube/v3/[12]	Adaptive Genetic Algorithm Performed better than other 8 Algorithms
3	A Novel Approach for YouTubeVideo Spam Detection using Markov Decision Process	Markov Decision Process, Decision Tree, Naïve Bayes, KNN, Random Forest, Ripper, Clustering	50 Videos, 2054 Instances out of which 824 was Spam Comment and 1230 normal	Markov Decision Process accuracy 78.52 % better than the best RF which was 72.82 %
4	Spam Detection Using KNN and Decision Tree Mechanism in Social Network	KNN, Decision Tree	FED Real Dataset http://mashable.com/2012/12/18/twitter-200-million-active-users/ Accessed July 22, 2016 [13]	Precision Call, F Measure and Class, FP and TP rate.
5	TubeSpam: Comment Spam Filtering on YouTube	Naïve Bayes, Decision Tree, Logistic Regression	5 Different YouTube Video Datasets Datasets / YouTube ID / # Spam/ # Ham / Total 1) Psy9bZkp7q19f0 175 175 350 2] KatyPerryCevxZvSJLk8 175 175 350 3) LMFAO KQ6zr6kCPj8 236 202 438 4) Eminemuel Hwf8o7 U 245 203 448 5) Shakira pRpeEdMmmQ0 174 196 370 http://dcomp.sor.ufscar.br/talmeida/youtubespamcollection/[14]	A confidence level of 99 % on all algorithms, Suggested their own App. Tube Spam.
6	A Data Mining Based Spam Detection System For YouTube	Naïve Bayes, Decision Tree, Clustering	Self Generated Data using Tube Kit http://www.tubekit.org/ [15] 1. No of Videos 1719 2. No of distinct users 1428 3. No of comments 10102865 4. No of ratings count 23013568 5. No of different categories 16	Lift Score is Calculated, Naive Bayes and Decision Tree performed better having an accuracy of 80.20 % and 82.11 % respectively

4 YouTube Data Generation Technique

This section discussed how YouTube comments data is generated from YouTube videos. The toolkit TubeKit [15] is used for customized YouTube Crawlers. This toolkit collects a variety of data from YouTube videos. The working design of the crawler is shown in the below fig 2.

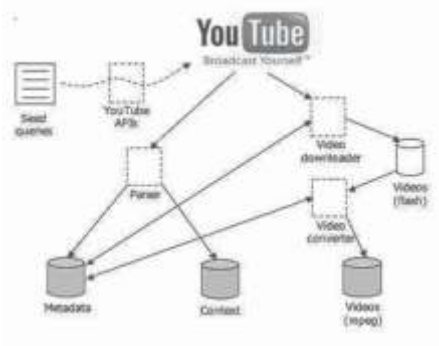

Fig.2. Scheme for Query Based YouTube Crawling [16]

5 Conclusion and Future

WorkYouTube a social networking feature website providing one of the largest video content publication. Recently a vast number of increases in his popularity and its new monetization policy for new channels have increased the number of users with low-quality information which is known as spam, which is posted solely for their channel promotion or to post the malicious video link.

The automatic moderation of comment is still an unexplored field and reason is the lack of automatic tool detection for spam filter messages. Due to which popular channelson YouTube try to avoid communication with their fans on this platform and use another platform for communication purpose.

The main goal behind this research was to find out the different techniques and strategies to find out the undesired comments i.e. spam messages and to describe different dataset available for the user working in this area. The results obtained through this research i.e. about techniques and available datasets description would further enhance the future comparison. Since for future work I would suggest that as there is not a single method that performs well on all the available dataset, so cascading of different machine learning classifiers along with best feature selection algorithms can be used to further enhance the result. Also preprocessing of the comment text can be done using natural language processing for text normalization. The reason behind text normalization is that the words used in comments are usually slang, idioms, emoticons, symbols, and abbreviations.

3 References

[1] https://www.youtube.com/intl/en-GB/yt/about/press/

[2] O'Callaghan, Derek et al. “Identifying Discriminating Network Motifs in YouTube Spam.” CoRR abs/1202.5216 (2012): n. pag.

[3] https://www.bbc.com/news/newsbeat-46559772

[4] https://support.google.com/youtube /answer/111870?hl=en/

[5] Chowdury, Rashid &Monsur Adnan, Md.Nuruddin& Mahmud, G.A.N. & Rahman, Mohammad. (2013). A data mining based spam detection system for YouTube. 373-378.

[6] C. Alberto, Tulio&Lochter, Johannes & Almeida, Tiago. (2015). “TubeSpam: Comment Spam Filtering on YouTube.” 138-143. 10.1109/ICMLA.2015.37.

[7] Goyal, Saumya& K. Chauhan, R &Parveen, Shabnam.(2016)“Spam detection using KNN and decision tree mechanism in social network”522-526.10.1109/PDGC.2016.7913250.

[8] S. Kanodia, R. Sasheendran and V. Pathari, "A Novel Approach for Youtube Video Spam Detection using Markov Decision Process," 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 2018, pp. 60-66. doi: 10.1109/ICACCI.2018.8554405.

[9] A. O. Abdullah, M. A. Ali, M. Karabatak and A. Sengur, "A comparative analysis of common YouTube comment spam

filtering techniques," 2018 6th International Symposium on Digital Forensic and Security (ISDFS), Antalya, 2018, pp. 1-5. doi: 10.1109/ISDFS.2018.8355315

[10] ShreyasAiyar, Nisha P Shetty, "N-Gram Assisted Youtube Spam Comment Detection", Procedia Computer Science, Volume 132,2018, Pages 174-182, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2018.05.181.

[11] Youtube, YoutubeDataAPI(v3), https://developers.google.com/youtube/v3/docs/commentThreads#Retrieve_comments

[12] Data API. Available at: https://developers.google.com/youtube/v3/ (Accessed on 22 November 2017)

[13] S. Fiegerman, "Twitter now has more than 200 million monthly active users," Mashable. [Online]. Available. http://mashable.com/2012/12/18/twitter-200-million-active-users/. Accessed July 22, 2016

[14] http://www.dt.fee.unicamp.br/~tiago//youtubespamcollection/

[15] TubeKit, http://www.tubekit.org/. , last accessed 24th August 2013.

[16] C. Shah, "Supporting Research Data Collection from YouTube with TubeKit". Proceedings of YouTube and 2008 Election Cycle in the United States, Amherst, MA: April 16-17, 2009.