✚ IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

AN PRACTICAL APPROACH ON PARAMETER BASED VIDEO SPAM DETECTION

Pooja Kamboj*, Er. Chandna Jain
M.Tech, CSE, JCDM College of Engineering, Sirsa, India.
Asst Professor, CSE, JCDM College of Engineering, Sirsa, India.

ABSTRACT

Video sharing websites are most popular services for share video among the number of users. These websites also connect with different kind of users known as social networking Web Services. But in these services, there are different kinds of pollution related to the video such as uploading of malicious, copyright violated and spam video or content. The users upload the video for the particular category and viewers/responders response the video by commenting, share video related to the uploaded video but malicious users does not share related video. The users share the unrelated information, abusive or pornographic, commercial video or it can be automatically scripts. These videos are the responsive videos. This research work presents a method for detection of such kind of responses. The first phase of detection is to divide the problem in different scenarios and define the keywords in database which can be related to commercial and pornographic. The sample dataset will contain all the information and video for analysis of the videos and categorized as per content type. The different kind of techniques will identify the video and with some features such as presence of certain terms in the title or description, Likes, elapsed time, uploaded time, etc. The different kind of videos data collection will be considered for analysis the algorithm and corresponding result will be generated. The research will be demonstrated using MATLAB TOOL which will contains the sample videos for Analyze the spam detection, porn video and commercial videos. The website will also save the keywords and these words will be update frequently and remove outdated words for improved detection of videos.
 .

KEYWORDS: Elapsed time, Token, Description.

INTRODUCTION

Spam on Web is a well known phenomenon for all the users that, on regular basis, browse the Web by the means of a search engine. It is an annoying experience since it forces users to load pages whose content is often completely unrelated with the search they submitted to the search engine. Nevertheless there does not exist a common definition, over which the scientific community agrees. It is indeed difficult to decide till which point we can consider "licit" the efforts devoted to increase pages rankings in the list of a search engine results. There are many and relatively easy-to-implement techniques used to attract and/ore redirect traffic likes attracted picture, comments etc. We have observed a number of typical aspects that characterize Web Spam pages:
– Many unrelated keywords and links.
– Many keywords used in the URL.
– Redirection of the user to another page.
– Creation of links in comments.
– More Likes to unrelated videos.

Social networking web Services create a scenario in which people can socialize and communicate throughout the world, examples of such social networking sites are Twitter, Facebook and MySpace. Users of these sites are able to add a wide variety of information to pages, to pursue common interests, and to connect with others. There are also business related web sites such as LinkedIn and these kinds of services used for business connections. For videos service, the YouTube is most popular web service by Google.

There are number of users who are interested in games and videos, and such kind of websites also has the section for video and games. But the industries related to their business and having great revenue by taking advantage of pornography and videos over the World Wide Web. For the revenue purpose, the industries are posting unrelated videos with respect to the related videos.

The users are searching something on internet but the searched content is not related to the user's requirement. For example, User is searching for the game of child but in another section, user is getting the porn video or such kind of advertisement which comes under the area of Spam. Being such a popular video sharing site, it become a platform for spammers and promoters to post unrelated and irrelevant content either as video response or as related video to the most popular videos either to gain popularity or to promote their sites or products. The term Spam can be defined, that is some inappropriate message posted over the internet web services, especially to large number of users with the intention of either getting publicity or to spread viruses, malwares.

Spam in domains such as emails, web-pages, blogs, social networking websites, online discussion forums, wikis and video sharing websites is prevalent and naturally has several negative impacts such as undesirable consumption of computing resources, lowering the reputation or value of the targeted legitimate web application, impacting search engine rankings, overwhelming moderators and administrators, and obstructs and misleads genuine usage of legitimate users and community.

The study shows that the interested user of particular video is getting the irrelevant video which is actually posted for the purpose of attraction towards the video which can be porn, commercial video and helpful for the company's profit. Video responses are the videos which are in response to the searched video content and related video which are matched with the searched content. The video responses can be detected as spam by analyze the title, description, uploading time, duration, number of Likes, etc in video responses and also in the related videos. The number of detection approached will be used for find the spam video which will analyze the video by counting of porn or commercial words and calculate its percentage. The complete approaches are explained in this paper described later.

Figure 1: Example of Spam on YouTube

Recently, online social networking services such as Facebook, Wikipedia and YouTube are experiencing a dramatic growth in terms of popularity. In particular, video content is becoming a predominant part of users' daily lives on the Web. By allowing users to generate and distribute their own multimedia content to large audiences, the Web is being transformed into a major channel for the delivery of multimedia. Video pervades the Internet and supports new types of interaction among users, including political debates, video chats, Video Responses, comments, video mail, and video blogs. A number of Web services are offering video-based functions as alternative to text-based ones, such as video reviews for products, video ads and video responses. In particular, the video response feature allows users to converse through video, by creating a video sequence that begins with an opening video and then followed with video responses from fans and detractors.

By allowing users to publicize and share their independently generated content, social video sharing systems may become susceptible to different types of malicious and opportunistic user actions, such as self-promotion, video aliasing and video spamming. We define a video response and related video spam as a video posted as a response to an opening video, but whose content is completely unrelated to the opening video. Video spammers are motivated to spam in order to promote specific content, advertise to generate sales, disseminate pornography or compromise the

system reputation. Spamming has been observed in several different contexts, including email, Web search engines and blogs. A numbers of spam detection techniques exploit characteristics present in the Video (e.g., Title, Description). Moreover, users of such systems can quickly learn to identify some Video spams e.g., by hyperlink, likes, duration etc.

Detection approach
The detection approach consists of first retrieving video's detail marked with count function. Count functon will be calculated by different scenarios. This calculated by parameters such as words percentage in Description, title like, duration. The next step consists of computing the values of variables indicating the spam intention of user (as spammer). We define different indicators and describe our intuition behind the proposed indicators. The value of the following indicators is then used to score a give user as Commercial and Pornographic Spammer.

User Test Collection
In order to evaluate our proposed approach to detect video spammers and promoters in online video social networking systems, we need a test collection of users, pre-classified into the target categories, namely, spammers, promoters and, in lack of a better term, legitimate users. However, to the best of our knowledge, no such collection is publicly available for any video sharing system, thus requiring us to build one.

Before presenting the steps taken to build our user test collection, we introduce some notations and definitions. We say a YouTube video is a responded video or a video topic if it has at least one video response. Similarly, we say a YouTube user is a responsive user if she has posted at least one video response, whereas a responded user is someone who posted at least one responded video.

Moreover, we define as spammer a user who posts at least one video response that is considered unrelated to the responded video (i.e., a spam). Examples of video spam are:
1. An advertisement of a product or website completely unrelated to the subject of the responded video,
2. Pornographic content posted as response to a cartoon video. A promoter is defined as a user who posts a large number of video responses to a responded video, aiming at promoting this video topic
A user that is neither a spammer nor a promoter is considered legitimate.

PROPOSED METHODOLOGY
1. Problem Statement
In Video Sharing Websites, there are number of videos and of different types such as child Videos, Pornography, Commercial Videos, Fun Video, etc. These websites provides the user interface for the clients to search the required video but the problem is of video response. When the user searches the videos, it also shows the related videos as well as response videos. The related videos are the same as searched videos. But the response videos are posted by the user as reply or comment of searched video. The users respond to the particular video and this response may or may not be related to the video which will be count as Spam. The spam detection of response videos is major issue. User responds the videos for increase the attraction to particular content or for the commercial purposes.
a. Study of Commercial and Porn Words Dictionary.
b. Study of identification Parameters for porn and Commercial Videos.
c. Detect Spam for the Searched Videos.
d. Analyze the Porn Video Response.
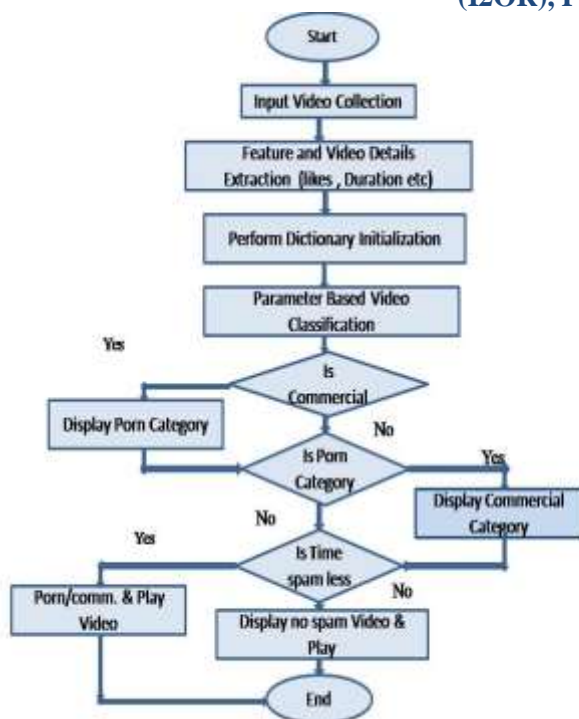e. Analyze the Commercial Video Reponses.

Figure 2: Proposed Flow Chart

Methodology
a.   Study of Existing Spam Detection Techniques.
b.   Identify the benefits of the Existing Methods.
c.   Research of new Spam Detection Technique.
d.   Initialize the words Dictionary includes Porn and Commercial.
e.   Analyze Title, Description, Uploaded Video time, Timestamp, etc. of response video with searched video and
      compare words with database dictionary.
f.   Identify the Commercial Spam and Porn Spam based on Words Percent.
g.   The result will be generated and display the spam videos.

Table show Description of video. Which user take as sample for representation of result. Table show video with s.no which is used as ID, Title of each video, Description of video (comments, hyperlinks etc), numbers of Likes by users, Duration of video, Date and Time of video uploading. These all video taking as sample to check video is spam video or not with elapsed time of each video.

S.no	Video Title	Video Description	Likes	Duration	Date/Time
1	Rum Whisky Baby	Baby video on request please watch to refund your amount click here account refund. Failure to store http://	10	02:06	30/09/2015 12:10 AM
2	Mar Jayian XXX	Wizarad re Wizard http://	8	00:11	30/09/2015 12:12 AM
3	sweet girl	as from Pani da Rang http://	10	02:20	24/11/2015 10:10 AM
4	cargo Aloochaat	this song form aloo chat title song http://	20	02:03	27/11/2015 11:01 AM
5	Anwar Maula Mere	Maula by Roop Kumar Rathod	15	02:01	27/11/2015 11:30 AM
6	Chandigarh	Babble Rai Offical Video	11	03:58	28/11/2015 10:00 AM
7	Dance India Dance	Audition 1. With Pack Your Suitcase Pimp	25	1:51	28/11/2015 11:15 AM
8	Leja Leja	Ustad Sultan Khan & Shreya Ghoshal	50	04:26	28/11/2015 11:30 AM
9	I Love You invoice	Bodyguard solo dance Compose this song within a week to store this song click on Download http://	22	02:34	29/11/2015 12:00 AM
10	Koi Tum sa nahi Three-way	failure to play Krish. Play for this video click on request Button http://	13	05:06	30/11/2015 09:00 PM
11	Tera Dedar Hua	Song on Three-way road. Man with Suitcase Snowballing Squirts Thumbs http://	20	2:05	30/11/2015 09:25 PM
12	Ye jo Mohbbat hai	To purchase this song click PAY. failure to pay. extension of payment time for this song http://	30	3:36	30/11/2015 09:45 PM

Figure 3: Video Records with Title, Description, Likes, Duration etc.

Objective
There is presence of spam on the most popular video sharing site i.e. YouTube and has several disadvantages. It requires researcher's attention to solve the video response spam problem on YouTube. The research aim of the work presented in this report is the following:
a. Identify the Spam on Video Sharing Websites.
b. Recognize the Porn, Commercial Videos.
c. Design web portal with Proposed Spam Detection.
d. Reduce Bandwidth Usage by Detection Un-related Video.

PARAMETER USED
Elapsed Time: Time taken to search keyword from the database and show the time and resulting video.

Token Passed: Token are keyword which we describe in porn and commercial tables e.g (Title, Description, like etc).

RESULTS
In spam detection, we have taken the collection of videos of different kind with multiple parameters such as Title, Description, likes, duration and uploaded time. The Porn, commercial detection has already explained and correspondingly the queries have also shown. Based on proposed techniques, the results have been generated which shows the accurate results.

Example 1:
Input: Input the number in digit. After enter the digit Eg. (1,2,3,4,5,6). These digit are save in database file name (a.xls) and in this input we enter 1 so the file play is 1.avi.

Figure 4: Video Search using video ID 1.avi

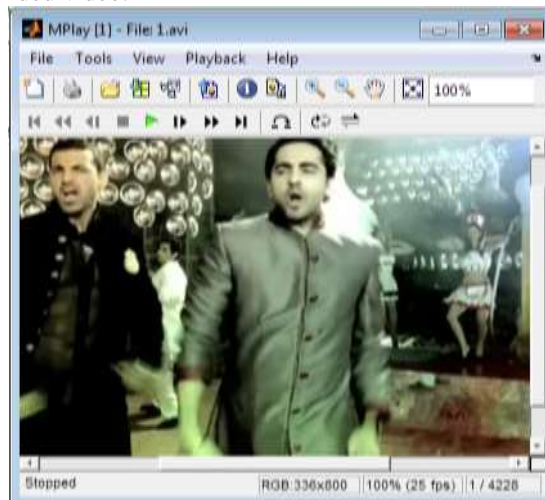**Video file**: Then play the responded video.

Figure 5: Responded Video according to search.

Output: The output shows the parameter for file 1.avi. the elapsed time of video and the categories of video. Elapsed time and categories are shown in figure.
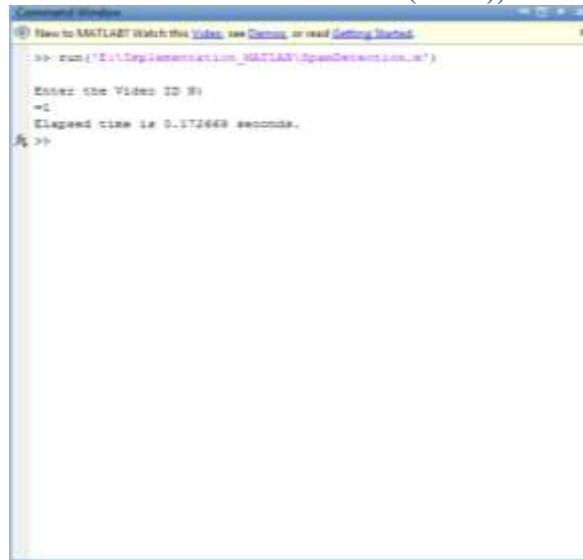
Figure 6: Show result video with elapsed time of search video

Example 2:
Input: Input the number in digit. After enter the digit Eg. (1,2,3,4,5,6). These digit are save in database file name (a.xls) and in this input we enter 2 so the file play is 2.avi.

Figure 7: Video Search using video ID 2.avi
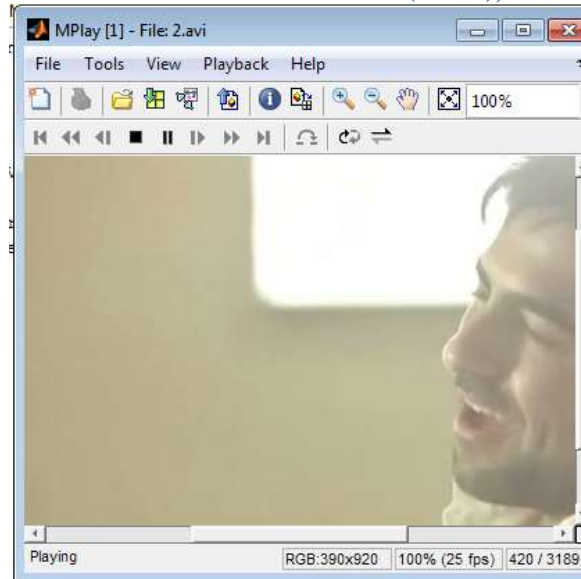
Video file: Then play the responded video.

Figure 8: Responded Video according to search.

Output: The output shows the parameter for file 2.avi. the elapsed time of video and the categories of video. Elapsed time and categories are shown in figure.
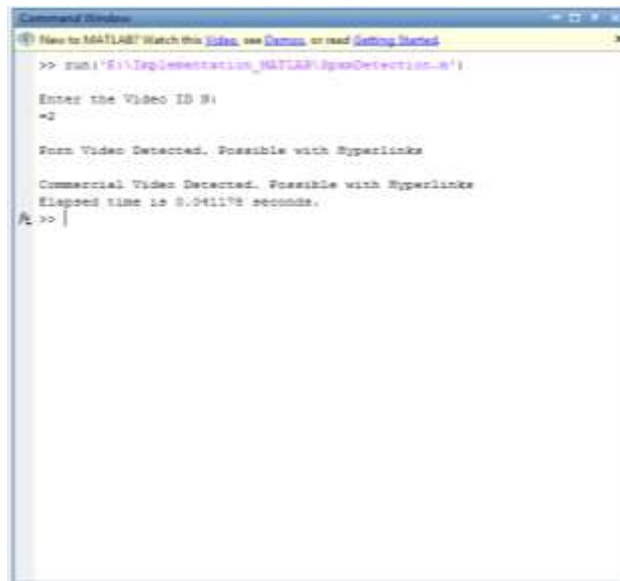

Figure 9: output show elapsed time of search video and category.

ELAPSED TIMING

This Table show the elapsed timing of each sample video which is used by user. This Table show Responded video ID and elapsed timing of each video. Which show how much timing is taken by video searching and output taken. Each video taken different elapsed time because video show spam video or not according to matching video from Dictionary using keywords used in Title, comment, Likes etc.

RESPONDING VIDEO ID	ELAPSED TIME
1.AVI	0.053744 seconds
2.AVI	0.070192 seconds
3.AVI	0.022028 seconds
4.AVI	0.027622 seconds
5.AVI	0.077109 seconds
6.AVI	0.076629 seconds
7.AVI	0.076629 seconds
8.AVI	0.079845 seconds
9.AVI	0.040255 seconds
10.AVI	0.062345 seconds
11.AVI	0.087481 seconds
12.AVI	0.03287 0 seconds

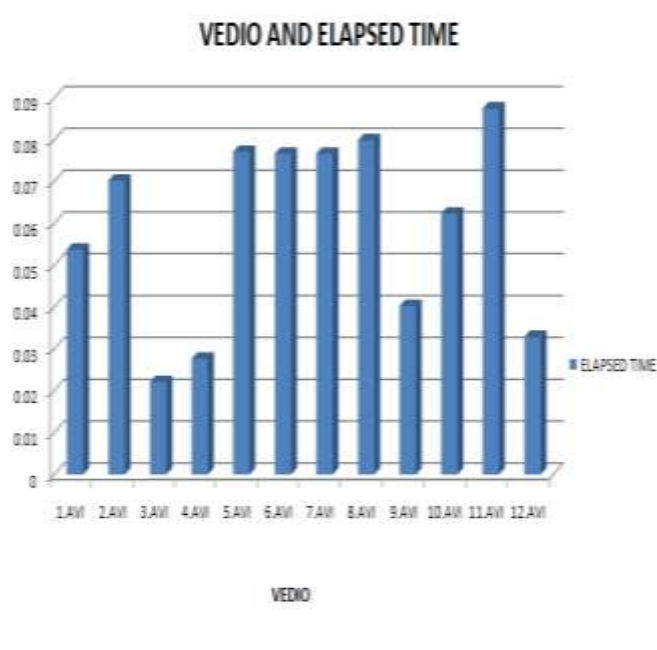*Figure 10: Table to show Responded Video & Elapsed Time*

Figure 11: Bar Graph of Video. Show elapsed time.

This Table show Elapsed Timing of each video when we use many number of keyword. Keyword used either from porn category Dictionary, or from commercial keyword Dictionary or from both. Elapsed time Different with different number of keyword.

S.NO	KEYWORDS	TIME
1.	7	0.054386 seconds.
2.	3	0.070192 seconds.
3.	2	0.076629 seconds.
4.	5	0.040255 seconds
5.	4	0.062345 seconds
6.	5	0.087481 seconds
7.	6	0.032870 seconds

Figure 12: Keywords &Time of Responded Video according to search.
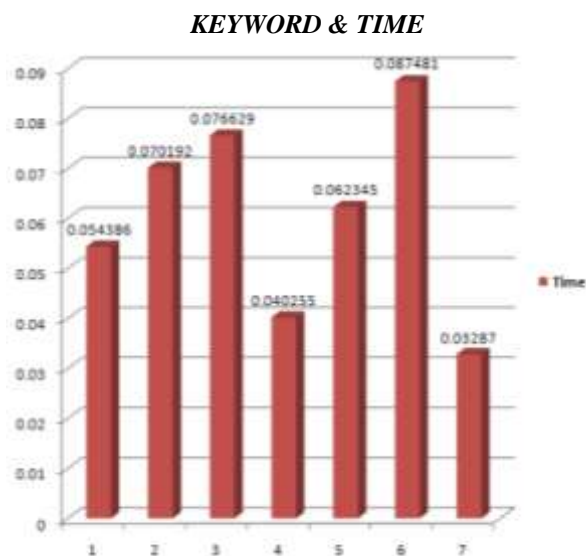
KEYWORD & TIME

Figure 13: Bar Graph of Videos. Show elapsed time when using many keywords.

CONCLUSION

In this paper, we presented a method based on the type specification of video for detection of the video responses spam on video sharing web sites. The proposed method detects the spam for responded videos and categorize accordingly such as porn, commercial and botnet videos. The detection is based on some parameters such as title, description, uploaded time, likes, etc. The survey shows that the likes for commercial and porn videos are approximately zero because of privacy and upload time is less for botnet video because it can be uploaded by the script. The web sites has been developed for implement the spam methods and results shows the un-related and related videos. This filtration has been performed based on features of video and responds video. The dictionary for commercial and porn words has been generated in database and compared it with the video features. The proposed method is efficient because the words dictionary can be update and will be helpful for spam detection and will be prevented by uploaded of irrelevant data.

REFERNCES

[1] Andras A. Benczur, Karoly Csalogany, Tamas Sarlas, Mate Uher (2002), "SpamRank-Fully Automatic Link Spam Detection".

[2]   Helen Costa UFOP, Fabricio Benevenuto, Luiz H. C. Merschmann(2003), "Detecting Tip Spam in Location-based Social Networks".
[3]   Gilad Mishne, David Carmel, Ronny Lempel (2005), "Blocking Blog Spam with Language Model Disagreement".
[4]   Luca Becchetti, Carlos Castillo Debora, DonatoAdriano Fazzone, (2006), "A Comparison of Sampling Techniques for Web Graph Characterization".
[5]   Seungyeop Han, Yong-yeol Ahn, Sue Moon, Hawoong Jeong(2006), "Collaborative Blog Spam Filtering Using Adaptive Percolation Search".
[6]   Yuan Niu Yi-Min Wang Hao Chen Ming Ma Francis Hsu(2006), "A Quantitative Study of Forum Spamming Using Context-based Analysis".
[7]   Yuan Niu, Yi-Min Wang (2007), "A Quantitative Study of Forum Spamming Using Context-based Analysis", Network & Distributed System Security.
[8]   Alan Mislove Massimiliano Marcon Krishna P. Gummadi (2007), "Measurement and Analysis of Online Social Networks", ACM.
[9]   Haifeng Yu, Michael Kaminsky, Phillip B. Gibbons, Member, IEEE, and Abraham D. Flaxman(2008), "SybilGuard: Defending Against Sybil Attacks via Social Networks".
[10] Fabricio Benevenuto Fernando Duarte (2008) ,"Understanding Video Interactions In Youtube", ACM