IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 05, Issue 07 (July. 2015), ||V4|| PP 26-30

Detecting spammers in YouTube: A study to find spam content in a video platform.

P. Sai Kiran
(Computer Science and Engineering, Vidya Jyothi Institute of Technology(JNTU-H), TG, India)

Abstract:- *Social networking has become a popular way for users to meet and interact online. Users spend a significant amount of time on popular social network platforms (such as Facebook, MySpace, or Twitter), storing and sharing personal information. This information, also attracts the interest of cybercriminals.*
In this paper, a step further is taken by addressing the issue of detecting video spammers and promoters.

Keywords: *YouTube, Spammers, video spam, social network, supervised machine learning, SVM.*

I. INTRODUCTION

Over the last few years, social networking sites have become one of the main ways for users to keep track and communicate with their friends online. Sites such as Facebook, MySpace, and Twitter are consistently among the top 20 most-visited sites of the Internet. Moreover, statistics show that, on average, users spend more time on popular social networking sites than on any other site [1]. Most social networks provide mobile platforms that allow users to access their services from mobile phones, making the access to these sites ubiquitous. The tremendous increase in popularity of social networking sites allows them to collect a huge amount of personal information about the users, their friends, and their habits. Unfortunately, this amount of information, as well as the ease with which one can reach many users, also attracted the interest of malicious parties. In particular, spammers are always looking for ways to reach new victims with their unsolicited messages. This is shown by a market survey about the user perception of spam over social networks, which shows that, in 2008, 83% of the users of social networks have received at least one unwanted friend request or message [2].

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 [3]. A video response spam is defined 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 (often as an advertisement) or compromise the system reputation.

Ultimately, users cannot easily identify a video spam before watching at least a segment of it, thus consuming system resources, in particular bandwidth, and compromising user patience and satisfaction with the system. Thus, identifying video spam is a challenging problem in social video sharing systems.

paper is addressed on the issue of detecting video spammers and promoters. To do it, a large user data set from YouTube site is crawled, containing more than 260 thousands users. Then, the creation of a labeled collection with users "manually" classified as spammers and non-spammers is taken place. Using attributes based on the user's profile, the user's social behavior in the system, and the videos posted by the user as well as the target (responded) videos, I investigated the feasibility of applying a supervised learning method to identify polluters. I found that my approach is able to correctly identify the majority of the promoters, misclassifying only a small percentage of legitimate users. In contrast, although I was able to detect a significant fraction of spammers, they should to be much harder to distinguish from legitimate users.

The rest of the paper is organized as follows. The next section discusses background. Section 3 describes the crawling strategy and the test collection built from the crawled dataset. Section 4 discusses about the spam metrics. Section 5 describes the classification. Finally, Section 6 offers conclusions.

II. BACKGROUND

Mechanisms to detect and identify spam and spammers have been largely studied in the context of web [4, 5] and email spamming [6]. In particular, Castillo et al [4] proposed a framework to detect web spamming which uses social network metrics. A framework to detect spamming in tagging systems, which is a type of attack that aims at raising the visibility of specific objects, was proposed in [7]. Although applicable to social media sharing systems that allow object tagging by users, such as YouTube, the proposed technique exploits a specific object attribute, i.e., its tags. A survey of approaches to combat spamming in Social web sites is

presented in [8]. Many existing approaches are based on extracting evidence from the content of a text, treating the text corpus as a set of objects with associated attributes and using these attributes to detect spam. These techniques, based on content classification, can be directly applied to textual information, and thus can be used to detect spam in email, text commentaries in blogs, forums, and online social networking sites. Complementary to my effort, the characterization of the traffic to online video sharing systems, in particular YouTube, has also been the focus of some studies. An in-depth analysis of popularity distribution, popularity evolution and content characteristics of YouTube and of a popular Korean video sharing service is presented in [9]. The authors also analyze mechanisms to improve video distribution, such as caching and peer-to-peer distribution schemes. Gill et al [10]present a characterization of the YouTube traffic collected from the University of Calgary campus network and compare its properties with those previously reported for web and media streaming workloads. Both studies focus on traffic and video characterization. I are not aware of any effort to characterize video spamming. Towards this end, this paper presents a characterization of user and video attributes that can be used to distinguish spammers from legitimate users in YouTube.

III. USER TEST COLLECTION

In order to evaluate the proposed approach to detect video spammers and promoters in online video social networking systems, I needed a test collection of users, pre-classified into the target categories. To the best of my knowledge, no such collection is publicly available for any video sharing system, thus requiring me to build one. Examples of video spam are: (i) an advertisement of a product completely unrelated to the subject of the responded video, and (ii) 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*. As an example, I found promoters in the dataset who posted a long sequence (e.g., 100) of (unrelated) video responses, often without content (0 second) to a single video.

3.1 Crawling YouTube:

YouTube includes several social networking features. Since my focus is on video response spamming, I am interested in sampling information about users who participate in video based interactions. In other words, my crawling strategy is driven by users who have responded to other users by uploading videos. This type of interaction is enabled by the video response feature, which allows a registered YouTube user to post a video as response to a pre-existing YouTube video. A YouTube video is a responded video if it has at least one video response. Similarly, a YouTube user is a responded user if at least one of its contributed videos is a responded video. Finally, a YouTube user is a responsive user if it has posted at least one video response. A very natural user graph emerges from video response interactions. At a given instant of time t, let X be the union of all responded users and responsive users. The set X is, of course, a subset of all YouTube users. I denote the video response user graph as the directed graph (X, Y), where (x1, x2) is a directed arc in Y if user $x1 \in X$ has responded to a video contributed by user $x2 \in X$.

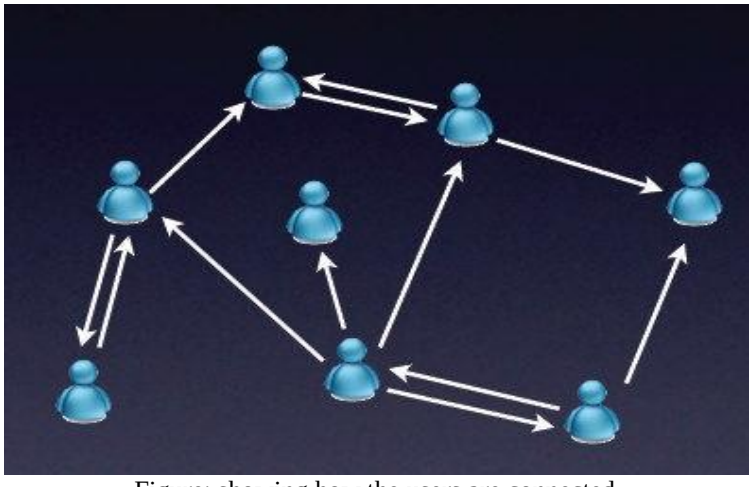

Figure: showing how the users are connected

Below is the algorithm that was used to construct a list of users.

input: Link to the Youtube's home page
1.1 randomly select any video from the home page
1.2 append the user in the user list

```
1.3 foreach instance I in likes_of_video do
1.4     if I is in userlist then
1.5         continue
1.6     end
1.7     if I is not in userlist then
1.8         append the user I in userlist
1.9     end
1.10 end
```

Algorithm-1: User collection

In order to obtain a large set of responded and responsive users, and ultimately build a video response user graph, I used the sampling procedure described in Algorithm below.

```
input : A list of users
 2.1 foreach User U in the crawler list do
2.2     Collect U's info using the YouTube API;
2.3     Collect U's video list using the API;
2.4     foreach Video V in the video list do
2.5         Copy the HTML of V ;
2.6         if V is a responded video then
2.7             Copy the HTML of V 's video responses;
2.8             Insert the responsive users in the crawler list;
2.9         end
2.10        if V is a video response then
2.11            Insert the responded user in the crawler list;
2.12        end
2.13    end
2.14 end
```

Algorithm 2: Crawler Algorithm for YouTube responses.

In total, my test collection contains 592 users, out of which 473 were classified as legitimate users and 119 as spammers. Throughout the rest of the paper, I will refer to these users simply by legitimate users and spammers, taking my manual classification as baseline of comparison for evaluating the effectiveness of my spammer detection mechanism. The users in my test collection posted a total of 16,611 video responses to 8,710 different videos.

IV. DETECTING SPAMMERS AND LEGITIMATE USERS

My spammer detection method relies on a machine learning approach for classifying my dataset. In this approach, the classification algorithm "learns"(supervised learning[14]) with part of the data and then applies its knowledge to classify users into two types: legitimate or spammers.

4.1 spam metrics:

In order to define the metrics used to evaluate the proposed heuristics, I have considered the following measures:

		Prediction	
		Legitimate	Spammer
True Label	Legitimate	a	b
	Spammer	c	d

Let us represent the number of legitimate users correctly classified as legitimate, b the number of legitimate users falsely classified as spammer, c the spammers falsely classified as legitimate, and d the number of spammers correctly classified as spammers. In order to evaluate the classification algorithms, we consider the following metrics, commonly used on Machine Learning and Information Retrieval [13]:

• True positive rate T P, or recall: $R = d/(c+d)$.

• True negative rate: T N = a/(a+b) .
• False positive rate: F P = b/(a+b) .
• False negative rate: F N = c/(c+d) .
• Accuracy = (a+d)/(a+b+c+d)
• F-measure: F = (2 · P ·R)/(P +R) , where P is the precision P = d/(b+d).

V. CLASSIFICATION

The SVM [11] (Support vector machine) methods are well known class of algorithms for data classification. I chose to use the SVM methods as the classifier for my dataset.

Basically, SVM performs classification by mapping input vectors to an N-dimensional space. The goal is to find the optimal hyper plane that separates the data into two categories, each one constructed on each side of the hyper plane. I use a binary non-linear SVM with RBF kernel to allow SVM models to perform separations with very complex boundaries. I chose to use the implementation of SVM provided with libSVM [12], an open source SVM package that allows searching for the best classifier parameters (i.e. cost and gamma) in order to define the best SVM configuration for the dataset.

Metric	User	Video	SN	ALL
TP	0.064	0.420	0.335	0.469
TN	0.998	0.946	1	0.991
FP	0.007	0.078	0	0.023
FN	0.976	0.574	0.625	0.571
Accuracy	0.822	0.851	0.874	0.890
F-measure	0.094	0.484	0.590	0.558

SVM classification results

In my analysis, I built the classifier using each set of correlated features (i.e. user-based, video-based, social network, and using all features together). The results are shown above. Analyzing the results for the classifier using all features, I observe that SVM obtained 0.469 for true positive rate, meaning that 46.9% of the spammers are correctly classified as spammers and could be correctly removed from the system. For the legitimate users, 99.1% are classified correctly. The accuracy obtained is 0.90, meaning that my approach classified erroneously 10% of all users. Clearly, the majority of the users classified erroneously are spammers.

VI. CONCLUSIONS

In this paper I studied video spam in a popular online social video network, namely YouTube. My study relies upon a dataset collected from YouTube. I crawled the YouTube site to obtain an entire component of the video response user graph. By manual inspection, I created a test collection with users classified as spammers or legitimate.

I provided a characterization of the users on this test collection which raises several attributes useful to characterize the social or anti-social behavior of users.

Using a classification technique, I proposed a video spam detection mechanism which is able to correctly identify significant fraction of the video.

REFERENCES:

[1]. Alexa top 500 global sites. http://www.alexa.com/topsites.
[2]. Harris Interactive Public Relations Research. A study of social networks scams. 2008
[3]. M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system. In Proc. of IMC, 2007.
[4]. C. Castillo, D. Donato, A. Gionis, V. Murdock, and F. Silvestri. Know your neighbors: Ib spam detection using the Ib topology. In Int'l ACM SIGIR, pages 423–430, 2007.
[5]. Z. Gy¨ongyi, H. Garcia-Molina, and J. Pedersen. Combating Ib spam with trustrank. In Int'l. Conf. on Very Large Data Bases, pages 576–587, 2004.
[6]. L. Gomes, F. Castro, V. Almeida, J. Almeida, R. Almeida, and L. Bettencourt. Improving spam detection based on structural similarity. In Proc. of SRUTI, 2005.
[7]. G. Koutrika, F. Effendi, Z. Gy¨ongyi, P. Heymann, and H. Garcia-Molina. Combating spam in tagging systems. In Proc. of AIRIb, 2007.
[8]. P. Heymann, G. Koutrika, and H. Garcia-Molina. Fighting spam on social Ib sites: A survey of approaches and future challenges. IEEE Internet Computing, 11(6):36–45, 2007.

[9]. M. Cha, H. Kwak, P. Rodriguez, Y. Ahn, and S. Moon. I tube, you tube, everybody tubes: Analyzing the world's largest user generated content video system. In Proc. of IMC, 2007.
[10]. P. Gill, M. Arlitt, Z. Li, and A. Mahanti. Youtube traffic characterization: A view from the edge. In Proc. of IMC, 2007.
[11]. I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large margin methods for structured and interdependent output variables. Journal of Machine Learning Research (JMLR), 6:1453–1484, 2005.
[12]. R. Fan, P. Chen, and C. Lin. Working set selection using the second order information for training svm. Journal of Machine Learning Research (JMLR), 6:1889–1918, 2005.
[13]. R. Baeza-Yates and B. Ribeiro-Neto. Modern Information Retrieval. ACM Press / Addison-Wesley, 1999.
[14]. https://en.wikipedia.org/wiki/Supervised_learning