Detecting of Spammers and Content Promoters

[1]K.Ramamohan Rao, [2]Mr. Hari Krishna. Deevi

[1]Student, Nova College of Engineering and Technology, Ibrahimpatnam, Krishna Dist, Andhra Pradesh, India

[2] Assistant Professor, Nova College of Engineering and Technology, Ibrahimpatnam, Krishna Dist, Andhra Pradesh, India

Abstract: In many video social networks, users are permitted to post video responses to other users' videos. Even social networking services are a fast-growing business in the Internet. A response can be legitimate or can be a video response spam. A prototype of a large-scale search engine that makes heavy use of the structure present in hypertext. A number of online video social networks provide features that allow users to post a video as a response to a discussion topic. Spammers may post an unrelated video as response to a popular one aiming at increasing the likelihood of the response being viewed by a larger number of users. Opportunistic users - promoters - may try to gain visibility to a specific video by posting a large number of responses to boost the rank of the responded Video. Malicious users may post video response spam for several reasons including marketing advertisements, increase the popularity of a video, and distribute pornography. We propose, to go a step further by addressing the issue of detecting video spammers and promoters. We manually build a test collection of real YouTube users towards the end by classifying them as promoters, legitimates and spammers. We provide a characterization of social and content attributes that may help distinguish each user class. The feasibility of using a state-of-the-art supervised classification algorithm to detect spammers and promoters, and assess its effectiveness in our test collection. Although we are able to detect a significant fraction of spammers, they showed to be much harder to distinguish from legitimate users.

Keywords: Video**,** distribute pornography, youtube, Wikipedia

I. INTRODUCTION

Video content is becoming a predominant part of user's daily lives on the Web. Recently, online social networking services such you tube, face book and Wikipedia. Social networking services (SNSs) are one successful example of Internet has been a vessel to expand our social networks in many ways. SNSs provide an online private space for individuals and tools for interacting with other people in the Internet. The Web is being transformed into a major channel for the delivery of multimedia, by allowing users to generate and distribute their own multimedia content to large audiences.

Video pervades the Internet and supports new types of interaction among users including video charts, video blogs and video mails. A number of Web services are offering video-based functions as alternative to text-based ones. Most part of this huge success of multimedia content is due to the change on the user perspective from consumer to creator. The design of effective video content classification mechanisms seems crucial for automatic identification of videos with malicious content such as copyright protected. Content classification based solely on the *bare* content can be a challenging research problem due to the typically low quality of user-generated videos and the multitude of strategies one can make use of to publicize content in a video. Online video social networks may become susceptible to different types of malicious and opportunistic user actions. These systems usually offer three basic mechanisms for video retrieval:

- A search system
- Ranked lists of top videos
- Social links between users and/or videos

Although appealing as mechanisms to ease content location and enrich online interaction into the system. Video search systems can be fooled by malicious attacks in which users post their videos with several popular tags. Opportunistic behavior on the other two mechanisms for video retrieval can be exemplified by observing a YouTube feature, which allows users to post a video as a response to a video topic. *Spammers* may post an unrelated video as response to a popular video topic aiming at increasing the likelihood of the *response* being viewed by a larger number of users. Polluted content may compromise user patience and satisfaction with the system since users cannot easily identify the pollution before watching at least a segment of it. Promoters can further negatively influence system aspects. Promoted videos that quickly reach high rankings are strong candidates to be kept in caches or in content distribution networks.

We address the issue of detecting video spammers and promoters. We created a labeled collection with users “manually” classified as legitimate, spammers and promoters. We conducted a study about the collected user behavior attributes aiming at understanding their relative discriminative power in distinguishing between legitimate users and the two different types of polluters envisioned. We investigated the feasibility of applying a supervised learning method to identify polluters. Our approach is able to correctly identify the majority of the promoters, misclassifying only a small percentage of legitimate users.

II. RELATED WORK

A number of detection and combating strategies have been proposed for Content pollution has been observed in various applications like web search, blogs and e-mails. Most of them rely on extracting evidences from *textual descriptions* of the content and treating the text corpus as a set of objects with associated attributes. A malicious behavior that aims at increasing the visibility of an object by fooling the search mechanism. Our proposal is complementary to these efforts for two reasons:

a. Instead of classifying the content itself, it aims at detecting *users* who disseminate *video* pollution. Content-based classification would require combining multiple forms evidences extracted from textual descriptions of the video and from the video content itself
b. Video content would require more sophisticated multimedia information retrieval methods that are robust to the typically low quality of user-generated videos. We explore attributes that capture the feedback of users with respect to each other or to their contributions to the system

We analyzed the properties of the social network created by video response interactions in YouTube for finding evidence of pollution. In additional, we preliminarily approached this problem by creating a small test collection composed of spammers and legitimate users and applying a binary classification strategy to detect spammers. The present work builds on this preliminary effort by providing a much more thorough richer and solid investigation of the feasibility and tradeoffs in detecting video polluters in online video sharing systems. Our approach is complementary to these efforts as it aims at detecting video spammers using a combination of different categories of attributes of both objects and users.

Our study is also complementary to other studies of the properties of social networks and of the traffic to online social networking systems. An in-depth analysis of popularity distribution and evolution, the content characteristics of YouTube and of a popular Korean service. Our approach to detect video spammers consists on classifying users and relies on a set of attributes associated to the user actions and social behavior in the system as well as attributes of their videos.

III. CRAWLING A SOCIAL NETWORK

We visit pages on the YouTube site and gather information about video responses and their contributors, to collect data. We say a YouTube video is a *responded video* if it has at least one video response. We say a YouTube user is a *responded user* if at least one of its contributed videos is a responded video. Judge mentally, we say that a YouTube user is a *responsive user* if it has posted at least one video response. Consider a natural graph emerges from the video responses. For a't' time instance let us consider *X* be the union of all responded users and responsive users. We denote the video response user graph as the directed graph (*X*, *Y*). Since YouTube does not provide a means to systematically, visit all the responded videos. The sampled graph (A, B) obtained from this seed set is the graph analyzed on the next sections. Our second seed set consists on users obtained from the random sampling technique.

3.1. Crawling YouTube

Our strategy consists of collecting a sample of users who participate in interactions through video responses. These interactions can be represented by a video response user graph G=(X, Y)

Where X = union of all users who posted or received video responses until a certain instant of time

Y = directed arc of (x1, x2) in Y

In order to obtain a representative sample of the YouTube video response user graph. We build a crawler that implements Algorithm 1.

Algorithm 1: Video Response Crawler

```
Input: A list L of users (seeds)
for each User U in L do
    Collect U's info and list of videos (responded and responses);
    for each Video V in the video list do
        Collect info of V ;
        if V is a responded video then
            Collect info of V 's video responses;
            Insert the responsive users in L;
        end if
        if V is a video response then
            Insert the responded user in L;
        end if
    end for
end for
```

The crawler follows links of responded videos and video responses gathering information on a number of different attributes of their contributors

3.2. Building a Test Collection

The main goal of creating a user test collection is to study the patterns and characteristics of each class of users. The desired properties for our test collection include the following:

- Having a significant number of users of all three categories
- Including spammers and promoters that are aggressive in their strategies and generate large amounts of pollution in the system

- Including a large number of legitimate users with different behavioral profiles

We argue that these properties may *not* be achieved by simply randomly sampling the collection. Randomly selecting a number of users from the crawled data could lead us to a small number of spammers and promoters. Compromising the creation of effective training and test data sets for our analysis. Research has shown that the sample does not need to follow the class distribution in the collection in order to achieve effective classification. It is natural to expect that legitimate users present a large number of different behaviors in a social network. Selecting legitimate users randomly may lead to a large number of users with similar behavior not including examples with different profiles. In order to minimize the impact of human error, three volunteers analyzed all video responses of each selected user in order to independently classify her into one of the three categories. Volunteers were instructed to favor legitimate users. Video responses containing people chatting or expressing their opinions were classified as legitimate.

IV. DETECTING SPAMMERS AND PROMOTERS

We investigate the feasibility of applying a supervised learning algorithm along with the attributes for the task of detecting spammers and promoters. One of the each attribute for each user is represented by a vector of values. The algorithm learns a classification model from a set of previously labeled and then applies the acquired knowledge to classify new users into three classes:

- legitimate
- spammers
- promoters

To assess the effectiveness of our classification strategies we use the standard information retrieval metrics of recall, Micro-F1, precision. The recall (r) of a class X is the ratio of the number of users correctly classified to the number of users in class X. The precision (p) of a class X is the ratio of the number of users classified correctly to the total predicted as users of class X. The F1 metric is the harmonic mean between both precision and recall and is defined as F1 = 2pr/(p + r). Micro-F1 is calculated by first computing global precision and recall values for all classes, and then calculating F1. We use a Support Vector Machine (SVM) classifier that is a state-of-the-art method in classification and obtained the best results among a set of classifiers tested.

A SVM performs classification by mapping input vectors into an N-dimensional space and checking in which side of the defined hyper plane the point lies. The SVMs are originally designed for binary classification but can be extended to multiple classes using several strategies. We use a non-linear SVM with the Radial Basis Function (RBF) kernel to allow SVM models to perform separations with very complex boundaries. An open source SVM package that allows searching for the best classifier parameters using the *training* data. we use the *easy* tool from libSVM including normalization of all numerical attributes.

Classification experiments are performed using a 5-fold cross validation. The original sample is partitioned into 5 sub-samples out of which four are used as training data and the remaining one is used for testing the classifier. Process is then repeated 5 times with each of the 5 sub-samples used exactly once as the test data. The entire 5-fold cross validation was repeated 5 times with different seeds used to shuffle the original data set. The results reported are averages of the 25 runs. The confusion matrix obtained as the result of our experiments with the flat classification strategy as shown in the table1.

	Predicted		
	Promoter	Spammer	Legitimate

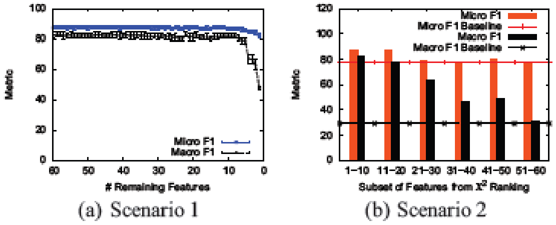
	96.13%	3.87%	0.00%
Promoter	1.40%	56.69%	41.91%
True	0.31%	5.02%	94.66%
Spammer			
Legitimate			

Table 1: Flat Classification

The numbers presented are percentages relative to the total number of users in each class. No promoter was classified as legitimate user that has only a small fraction of promoters were erroneously classified as spammers. We found that the videos that they targeted actually acquired certain popularity by manually inspecting these promoters. Significant fraction spammers were misclassified as legitimate users. these spammers exhibit a dual behavior sharing a reasonable number of legitimate videos and posting legitimate video responses. This dual behavior masks some important aspects used by the classifier to differentiate spammers from legitimate users.

V. RESULT ANALYSIS

Once we have understood the main tradeoffs and challenges in classifying users into spammers, legitimate and promoters. We now turn to investigate whether competitive effectiveness can be reached with fewer attributes. Evaluating the impact on the classification effectiveness of gradually removing attributes in a decreasing order of position in the X^2 ranking. There is no noticeable impact on the classification effectiveness when we remove as many as the 40 lowest ranked attributes. All social network attributes are among them is in the 30th position is the best positioned of these attributes. The Figure also shows that the effectiveness drops sharply when we start removing some of the top 10 attributes from the process as shown in the below fig.



(a) Scenario 1

(b) Scenario 2

Figure 1: Impact of Reducing the Set of Attributes

Evaluating our classification when subsets of 10 attributes occupying contiguous positions in the ranking are used. The fig shows the Micro-F1 and Macro-F1 values for the flat classification and for the baseline classifier that considers all users as legitimate, for each such range. Our classification provides gains over the baseline for the first two subsets of attributes whereas significant gains in Macro-F1 are obtained for all attribute ranges. This confirms the results of our attribute analysis that shows that even low-ranked attributes have some discriminatory power. Significant improvements over the baseline are possible even if not all attributes considered in our experiments can be obtained.

VI. CONCLUSION

Promoters and Spammers can pollute video retrieval features of online video social networks but also system resources and aspects such as caching. We propose an effective solution to the problem of detecting these polluters that can guide system administrators to spammers and promoters in online video social networks. Our proposed approach poses a promising alternative to simply considering all users as legitimate or to randomly selecting users for manual inspection. The system administrators could be more tolerant to misclassifications than in the second case, we proposed using the different classification tradeoffs. We found that our classification could produce significant benefits even if only a small subset of less expensive attributes is available. Expect that spammers and promoters will evolve and adapt to anti-pollution strategies. Some attributes may become less important whereas others may acquire importance with time consequently. We envision two directions towards which our work can evolve. we aim at reducing the cost of the labeling process by studying the viability of semi-supervised learning methods to detect polluters. we intend to explore other refinements to the proposed approach such as to use different classification methods.

VII. REFERENCE

[1] Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *Int'l World Wide Web Conference (WWW)*, 2007.

[2] G. Koutrika, F. Effendi, Z. Gyöngyi, P. Heymann, and H. Garcia-Molina. Combating spam in tagging systems. In *Int'l Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*, 2007.

[3] A. Mislove, M. Marcon, K. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Internet Measurement Conference (IMC)*, 2007.

[4] 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, 2005.

[5] comscore: Americans viewed 12 billion videos online in may 2008. http://www.comscore.com/press/release.asp?press=2324.

[6] The new york times: Search ads come to youtube. http://bits.blogs. nytimes.com/2008/10/13/search-ads-come-to-youtube.

[7] Youtube fact sheet. http://www.youtube.com/t/fact_sheet.