Prevention of Spammers and Promoters in Video Social Networks using SVM-KNN

Indira K [#1], Christal Joy E [*2]

[# 1]Assitant Professor, Department of Information Technology, Sathyabama University, Chennai, India
[# 2]Assitant Professor, Department of Information Technology, Sathyabama University, Chennai, India
[1] indira.it@sathyabamauniversity.ac.in, [2] christaljoy.it@sathyabamauniversity.ac.in

Abstract— As online social networks acquire larger user bases, they also become more interesting targets for spammers and promoters. Spam can take very different forms on social websites, especially in the form of videos and cannot always be detected by analyzing textual content. There are online video sharing systems that allow the users to post videos s response to any type of discussion topic. This feature encourages some of the users to post polluted content illegally as responses and there may be content promoters who try to promote them in the top listed search. Content pollution like spread advertise to generate sales, disseminate pornography, and compromise system reputation may threaten the trust of users on the system, thus weaken its success in promoting social interactions. As a solution for this problem, we classify the users as spammers, content promoters and legitimate users by building a test collection of real YouTube users using which we can provide a classification we use of content, individual and social attributes that help in characterizing each user class. For effective classification we use SVM-KNN which is an active learning approach. Our proposed approach poses a promising alternative to simply considering all users as legitimate or to randomly selecting users for manual inspection. In simple SVM training is very slow on whole dataset and not works very well on multiple classes. To overcome this problem and to provide efficient classification in fast manner we proposed new approach is SVM-KNN. Train a Support Vector Machine on K no of collections of nearest neighbours.

Keyword-Spammers, Promoters, Social Networks, SVM-KNN

I. INTRODUCTION

With internment video sharing sites gaining popularity at a dazzling speed, the web is being transformed into major channel for the delivery of multimedia content. Online video social networks (SNs), out of which YouTube is the most popular, are distributing videos at a massive scale. It has been reported that the amount of content uploaded to YouTube in 60 days is equivalent to the content that would have been broadcasted for 60 years, without interruption, by NBC, CBS, and ABC altogether. Moreover, YouTube has reportedly served over 100 million users only in January 2009 with a video upload rate equivalent to 10 h per minute. By allowing users to share their content, Video social networks are mostly targeted by spammers. Particularly, these systems usually offer three basic mechanisms for video retrieval: 1) a search system: 2) ranked lists of top videos; and 3) social links connecting users and/or videos. Although appealing as mechanisms to facilitate content location and enrich online interaction, these mechanisms open opportunities for users to introduce polluted content into the system. In many video social networks, including YouTube, users are permitted to post video responses to other user's videos. Such a response can be legitimate or can be a video response spam, which is a video response whose content is not related to the topic being discussed. Malicious users may post video response spam for several reasons, including increase the popularity of a video, marketing advertisements, distribute pornography, or simply pollute the system.

The most common technique involves people posting links to sites, most likely pornographic or dealing with online dating, on the comments section of random videos or people's profiles [9]. With the recent addition of a "Thumbs up/Thumbs down" feature, groups of promoters may constantly "Thumbs up" a comment, getting it into the Top Comments section and making the message more visible These pages may include their own or other user's videos, again often suggestive. Another kind is actual video spam, giving the uploaded movie a name and description with a well liked figure or event which is likely to draw consciousness or within the video has a certain image timed to come up as the video's thumbnail image to misdirect the viewer. The actual content of the video ends up being totally unrelated, sometimes offensive, or just features on-screen text of a link to the site being promoted [6]. In some cases, the link in question may lead to an online survey site or in extreme cases, malware.

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, which also consumes system resources, especially bandwidth.

To overcome this problem and to protect the online social networks from spammers and promoters, we proposed a new approach to classify the users as legitimate, spammer, or promoter by using SVM-KNN. Spammer means post an unrelated video as response to a popular video topic to increase the likelihood of the response being viewed. Promoters are post a large number of responses to boost the rank of the video topic. We found that our approach is able to correctly identify the majority of the promoters, misclassifying only a small percentage of legitimate users. In contrast, although we are able to detect spammers, they showed to be much harder to distinguish from legitimate users.

The rest of the paper is organized as follows. Section II discusses related work. Section III describes overall system architecture. In section IV, we described System implementation. It deals with crawling, user test data collection, user behaviour analysis, feature extraction, training phase, and classification using SVM-KNN approach. Experimental study conducted, evaluating performance metrics, and classification results were discussed in section V. Finally, Section VI provides Conclusion.

II. RELATED WORK

Social Systems are inherently open to users who generate, share and consume information like post a message, upload and watch a video. Many users mostly participate in social systems to engage in collaborative activities. And users, media and organization post information related to hot topics in real-time like updating live score of sports events. These positive aspects also lead to negative challenges. Some of the challenges are traditional attacks (Phishing, malware), spam (Comment spam, spam videos), misinformation, and crowdturfing. Web 2.0 is an area that’s gained much attention recently, especially with Google’s acquisition of YouTube [2]. The major drawback of this paper is mainly it is based on content classification and it requires multiple pieces of evidence from the textual descriptions of the video. The different kinds of existent pollution, their negative impact to users and system and possible strategies to minimize are analyzed [3]. The problem caused here is related to content retrieval; hence users would consume extra resources while accessing content if metadata is irrelevant. Three categories of potential countermeasures based on detection, demotion and prevention [4] are analyzed. This also depends on the pieces of evidence extracted from textual description of the content.

III.SYSTEM ARCHITECTURE

In this paper, we address the issue of detecting video spammers and promoters. To do it, first the user should be registered in our application and then they have to access social networks via our application. We crawl a larger user data set from YouTube site, containing more than thousands users. From this we collected test data and then we analysed user behaviours and attributes. Then, we create a labelled collection with users “manually” classified as legitimate, spammers, and promoters.

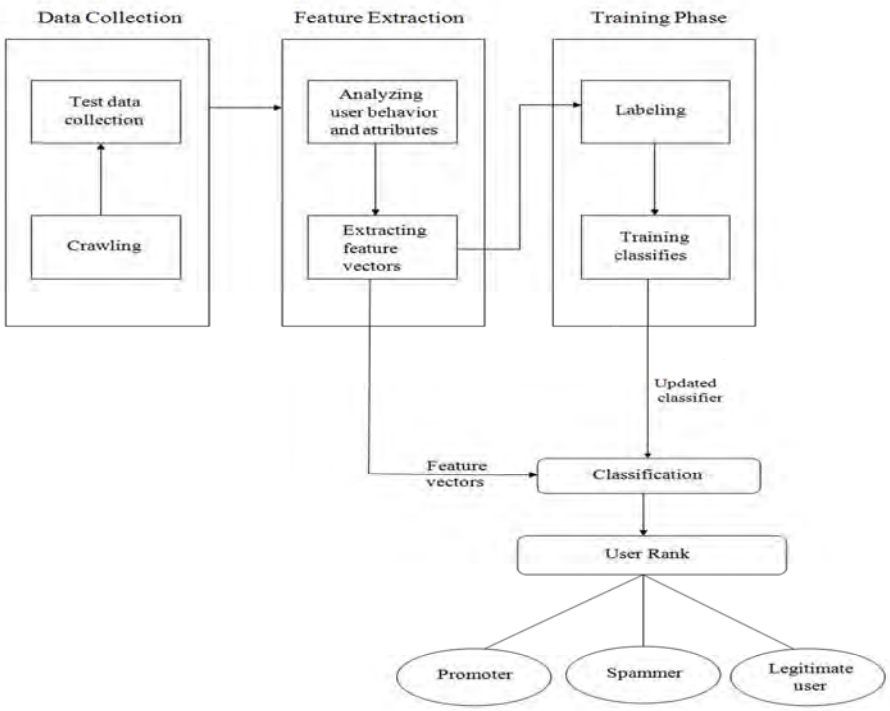

Fig. 1. System Architecture

After that, we will conduct a study about the collected user behaviour attributes aiming at understanding their relative discriminative power in distinguishing between legitimate users and the two different types of polluters envisioned.

Using attributes based on the user's profile, the user's social behaviour in the system, and the videos posted by the user as well as user target (responded) videos, we investigate the feasibility of applying a supervised learning method along with nearest neighbor classification to identify polluters. First apply the KNN approach on extracted feature vectors on that again we applied SVM. Then the users are classified into Legitimate, Spammers, and Promoters.

IV. SYSTEM IMPLEMENTATION

User should register in our application in order to accessing social networking Video Site.

A. User Test Collection

We say a You Tube video is a responded video or a video topic if it has at least one video response. Similarly, we say a You Tube user is a responsive user if user has posted at least one video response, whereas a responded user is someone who posted at least one responded video.

1) *Crawling:* Our strategy consists of collecting a sample of users who participate in interactions through video responses, i.e., who post or receive video responses. The sampling starts from a set of 88 seeds, consisting of the owners of the top-100 most responded videos of all time, provided by You Tube. The crawler follows links of responded videos and video responses, gathering information on a number of different attributes of their contributors (users), including attributes of all responded videos and video responses posted by user. User interactions can be represented by video response user graph.

$$G = (X, Y)$$

Where X is the union of all users who posted or received video responses. (x_1, x_2) is a directed arc in Y, if user x_1 has responded to a video contributed by a user x_2.

2) *Building Test Collection:* The main goal of creating a user test collection is to study the patterns and characteristics of each class of users. Our user test collection aims at supporting research on detecting spammers and promoters. Since the user classification labelling process relies on human judgment, which implies in watching significantly high amount of videos, the number of users in our test collection is somewhat limited. The collection should include the properties 1) having a significant number of users of all three categories; 2) including, but not restricting to large amounts of pollution; 3) including a large number of legitimate users with different behaviour. There are three strategies for user selection 1) different levels of interaction; 2) Aiming at the test collection with polluters; 3) randomly selected 300 users who responded video responses to the top 100 most responded videos; Then each selected user was classified manually into any one of three categories.

B. User Reviews

1) Post a Review: User can post their reviews about the videos. 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 is a responsive user if user has posted at least one video response, whereas a responded user is someone who posted at least one responded video.

2) Collecting User's Information: We consider three separated groups of videos owned by the user. The first group contains aggregate information of all videos uploaded by the user, being useful to capture how others see the (video) contributions of this user. The second group considers only video responses, which may be pollution. The group considers only the responded videos to which this user posted video responses (referred to as target videos). In order to obtain a representative sample of the YouTube video response user graph, we build a crawler by using video response crawling algorithm.

The sampling starts from a set of 88 seeds, consisting of the owners of the top-10 most responded videos of all time, provided by YouTube. The crawler follows links gathering information on a number of different attributes. (i.e) 264,460 users, 381,616 responded videos, and 701,950 video responses.

3) Analyzing User Behavior Attributes: The next step is to analyse a large set of attributes that reflect user behaviour in the system aiming at investigation their relative discriminatory power to distinguish one user class from the others. We considered three attribute sets, namely, video attributes (Duration, number of views, commentaries received, rating, number of times to be selected favourite, number of honor and external links), user attributes (number of friends, number of video uploaded, number of video watched, number of videos added as favourite, number of video responses posted and received, number of subscriptions and subscribers, average time between video uploads, maximum number of videos uploaded in 24 hours), and social network (SN) attributes (Clustering coefficient, betweenness, reciprocity, assortativity, user rank). For feature selection we used is chi squared method.

TABLE I
Feature Selection: $\varkappa^2$ Ranking

Attribute Set	Top10	Top20	Top30	Top40	Top50
Video	9	18	25	30	36
User	1	2	4	7	9
SN	0	0	1	3	5

C. *Detecting Spammers and Promoters*

1) *Calculating User Behavior:* The most discriminative user and social network attribute are the average time between video uploads and the user rank, respectively. In spite of appearing in lower positions in the ranking, particularly for the user rank attribute, these two attributes have potential to be able to separate user classes apart.

2) *Detecting and Blocking Spammers and Promoters:* Once we have understood the main tradeoffs and challenges in classifying users into spammers, promoters and legitimate, we now turn to investigate whether competitive effectiveness can be reached with fewer attributes. We report results for the flat classification strategy, considering two scenarios. In this approach, each user is represented by a vector values, one for each attribute. It is worth noting that some of the most expensive attributes such as user rank, which require processing the entire video response user graph.

D. *Classification Using SVM-KNN*

Each of the training data consists of a set of vectors and class label associated with each vector. In SVM, it will be either + or – (for positive or negative classes). It will not work well on multiple classes. Extension of SVM to multiple classes also is not natural as KNN. But KNN, can work equally well with arbitrary number of classes.

1) *Support Vector Machine (SVM)*: An SVM constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. H3 (green) doesn't separate the two classes. H1 (blue) does, with a small margin and H2 (red) with the maximum margin. Kernel function: K(x, y)=<Φ(x),Φ(y)>. The output of the kernel function is a kernel matrix whose element is the inner product of pairwise vectors in higher-dimensional space.

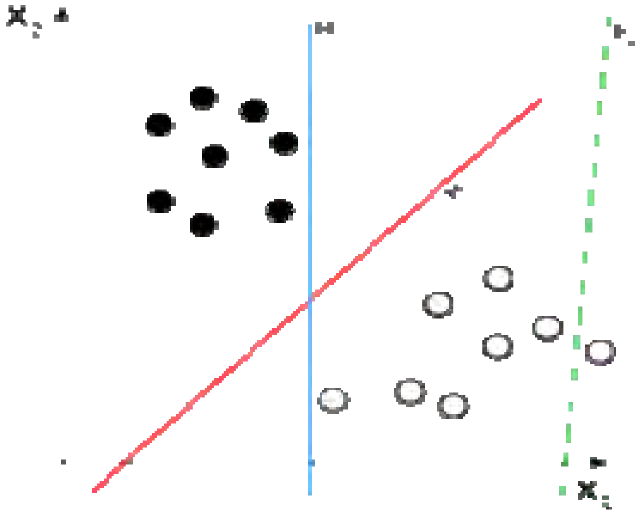

Fig. 2. SVM

2) *K-Nearest Neighbor (KNN)*: Assumptions in KNN 1) KNN assumes that the data is in a feature space. More exactly, the data points are in a metric space. 2) The data can be scalars or possibly even multidimensional vectors. 3) we are also given a single number "k". This number decides how many neighbours (where neighbor is defined based on the distance metric). This algorithm has different behavior based on k.

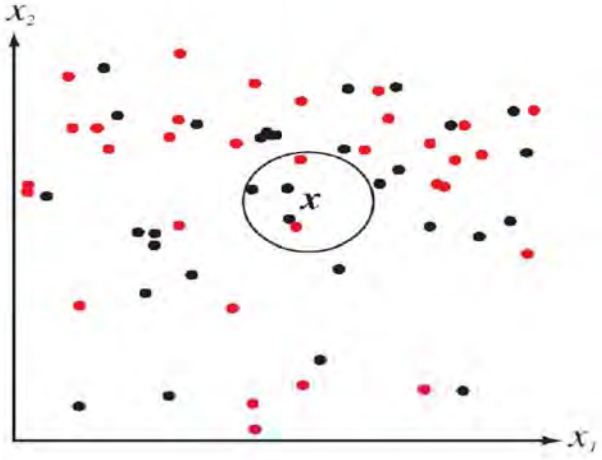

Fig. 3. KNN

Case 1: k=1, in this scenario let x be the point to be labelled. Find the point closest to x. Let it be y. Now assign the label of y to x. If the number of data points is very large, then there is a very high chance that label of x and y is same.

Case 2: k=K, we try to find the k nearest neighbour and do a majority voting. Typically k is odd when the number of classes is 2. Lets say k=5 and there are 3 instances of C1 and 2 instances of C2. In this case, KNN says that new point has to labeled as C1 as it forms the majority. We follow a similar procedure when there are multiple classes.

3) *SVM-KNN:* Train Support Vector Machine (SVM) on the collection of nearest neighbours. We use NN as an initial pruning stage and perform SVM on the smaller but more relevant set of examples that require careful discrimination. Kernel function: $K(x,y) = < \Phi(x), \Phi(y) >$. Map from Distance function to kernel function

$$K(x,y) = < x, y >$$

$$= \frac{1}{2}(< x,x > + < y,y > - < x-y, x-y >)$$

$$= \frac{1}{2}(d(x,0) + d(y,0) - d(x,y))$$

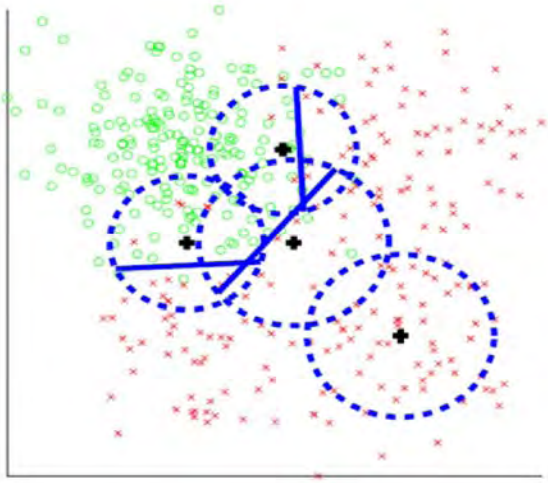

Fig. 4. SVM-KNN Local

Step by step procedure for SVM-KNN:

a) Compute distances of the query to all training examples and pick the nearest K neignbors;

b) If the K neighbours have all the same labels, the query is labelled and exit; else, compute the pairwise distances between the K neighbours;

c) Convert the distance matrix to a kernel matrix and apply multi class SVM;

d) Use the resulting classifier to label the query;

V. EXPERIMENTAL STUDY

A. *Evaluation Metrics*

Once we construct the predictive and actual results that can be evaluated by using standard

information retrieval metrics such as accuracy, recall, precision, micro-F1, macro-F1. To evaluating these metrics first we constructed confusion matrix.

A confusion matrix contains information about actual and predicted classification done by our classification system called SVM-KNN.

TABLE III
Example Confusion Matrix

		Predicted		
		Promoters	**Spammers**	**Legitimate**
True	**Promoters**	a	b	c
	Spammers	d	e	f
	Legitimate	g	h	i

TABLE IIIII
Actual Confusion Matrix

		Predicted		
		Promoters	**Spammers**	**Legitimate**
True	**Promoters**	97.72%	2.28%	0.00%
	Spammers	1.14%	65.90%	32.96%
	Legitimate	0.00%	4.54%	95.46%

1) *Accuracy (AC):* AC is the proportion of the total number of predictions that were correct. It is calculated as the sum of correct classifications divided by the total number of classifications.

$$AC = \frac{\sum TruePositive + \sum TrueNegative}{\sum TotalPopulation}$$

AC = 98.49%

2) *Recall (R):* R is the proportion of positives cases that were correctly identified.

Recall for Promoters:

$$R_{prom} = a / (a+b+c)$$

R_{prom} = 97.72%

Recall for Spammers:

$$R_{spam} = e / (d+e+f)$$

R_{sapm} = 65.9%

Recall for Legitimate Users:

$$R_{leg} = h / (g+h+i)$$

R_{leg} = 95.46%

3) *Precision (P):* P is the proportion of the predicted positive cases that were correct.

Precision for Promoters:

$$P_{prom} = a / (a+d+g)$$

P_{prom} = 99.37%

Precision for Spammers:

$$P_{spam} = e / (b+e+h)$$

P_{sapm} = 89.98%

Precision for Legitimate Users:

$$P_{leg} = i / (c+f+i)$$

$$P_{leg} = 67.3\%$$

4) *Micro-F1:* First computing global precision and recall values for all classes. Then calculating F1,

$$\text{Micro-F1} = 85.95\%$$

With per classes F1 values are,

$$F1_{prom} = 2P_{prom}R_{prom} / (P_{prom} + R_{prom})$$

$$= 98.54\%$$

$$F1_{sapm} = 76.08\%$$

$$F1_{leg} = 78.94\%$$

5) Macro-F1: First calculating F1 values for each class in isolation. Then averaging overall classes.

$$\text{Macro-F1} = 84.52\%$$

B. Classification Results

TABLE IVV
Performance Comparison

Classification Algorithms	Accuracy
KNN	84.89
Decision Tree	89.53
Random Forest	89.76
SVM-KNN	98.49

With 5-fold cross validation, compare to remaining methods listed in Table IV SVM-KNN approach produced more accuracy in flat classification.

VI. Conclusion

The main goal of creating a user test collection is to study the patterns and characteristics of each user class. Thus, the desired properties for our test collection are: 1) having a significant number of users of all three categories; 2) including, but not restricting to, spammers and promoters which are aggressive in their strategies and generate large amounts of pollution in the system; and 3) 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. The reasons for this are twofold.

We here proposed an effective solution that can help system administrators to detect spammers and promoters in online video SNs. our proposed approach poses a promising alternative to simply considering all users as legitimate or to randomly selecting users for manual inspection. And classify the users using super learning method along with nearest neighbor approach. Thus, we increase the detection rate and give better performance compared to SVM approach alone.

References

[1] F.Benevenuto, G.Magno, T.rodrigues, V.Almedia, “Detecting Spammers on twitter,” in Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference (CEAS) 2010.

[2] S.Boll., “Multitube- Where web 2.0 and multimedia could meet,” IEEE MultiMedia, vol. 14, no. 1, pp. 9-13, Jan.-Mar. 2007.

[3] F.Benevenuto, T.Rodrigues, V.Almeida, J.Almeida, M. Gonalves, and K. Ross, “Video Pollution on the Web,” First Monday, vol. 15, no. 4, pp, 1-20, Apr.2010.

[4] P.Heymann, G.Koutrika, and H.Garcia-Molina, “Fighting Spam on Social Websites: A survey of approaches and Future Challenges,” Nov/Dec. 2007 .

[5] Kondra Mohan Raju, E.Madhukar, “ Classification of users in online social networks,” IJITEE, vol.3,issue-7, Dec 2013.

[6] F.Benevenuto, T.rodrigues, V.Almedia, J.Almedia, M.goncalves, “Detecting Spammers and Content Promoters in online video social networks,” International ACM SIGIR conference on Research and Development in Information Retrieval pp. 620-627, 2009.

[7] M. Cha, H. Kwak, P.Rodriguez, Y.-Y. Ahn, and S. Moon, “Analyzing the video popularity characteristics of large-scale user generated content,” IEEE/ACM Transactions on Network, 17(5):1357-1370, 2009.

[8] C. Costa, V. Soares, J.Almedia, and V.Almedia, “Fighting Pollution Dissemination in peer-peer Networks,” in ACM Symbosium on Applied Computing (SAC) pp. 1586-1590, 2007.

[9] A. Thomason, “ Blog Spam: A Review,” in Conference on Email and Anti-Spam (CEAS), 2007.

[10] Chau, Pandit, S.Wang, and Faloutsos, “Parallel Crawling for online Social Networks,” in Int’l World Wide Web Conference (WWW), 2007.