Identifying Fraudulently Promoted Online Videos

Vlad Bulakh
School of Informatics and Computing
Indiana University
Bloomington, Indiana
vbulakh@cs.indiana.edu

Christopher W. Dunn
School of Informatics and Computing
Indiana University
Bloomington, Indiana
chrdunn@cs.indiana.edu

Minaxi Gupta
School of Informatics and Computing
Indiana University
Bloomington, Indiana
minaxi@cs.indiana.edu

ABSTRACT

Fraudulent product promotion online, including online videos, is on the rise. In order to understand and defend against this ill, we engage in the fraudulent video economy for a popular video sharing website, YouTube, and collect a sample of over 3,300 fraudulently promoted videos and 500 bot profiles that promote them. We then characterize fraudulent videos and profiles and train supervised machine learning classifiers that can successfully differentiate fraudulent videos and profiles from legitimate ones.

Categories and Subject Descriptors

H.3.5 [**Online Information Services**]: Web-based services; J.4 [**Computer Applications**]: Social and behavioral sciences

General Terms

Security, Videos

Keywords

Online videos; YouTube; opinion spam; classifier; SVM; decision tree; supervised machine learning

1 Introduction

Millions of Internet users visit and upload videos at online video sharing websites, such as YouTube, Vimeo, Metacafe, and Dailymotion. With such reach, various kinds of issues now plague such sites, including fraudulent promotion of videos, spam on legitimate video comments, and copyright issues. This paper focuses on the fraudulent video promotion.

Fraudulent video promotion – which is an activity that involves boosting video statistics via automated and/or illegitimate means – is gaining ground for multiple reasons. Several video sharing websites, including YouTube, Metacafe, and Videobb run advertisements on uploaded videos and share generated revenues with the uploaders. With examples of videos like "David After Dentist" fetching as much as 100K USD [1] from advertisements, it is unsurprising that many users would be motivated to boost the rankings of their videos through illegitimate means. As a concrete instance of a fraudulently promoted video from the YouTube data set we use in this paper, we found a video that increased its views by a factor of 100 over the course of a year to over a million views. This video was not uploaded by its creator, but appeared stolen and edited. We found 14 stolen copies of the video. Many copies were in the original category (people and blogs), but some were in wildly inappropriate categories, including sports, gaming, and comedy. At least one of the copies was promoting an adult website. The original video was in Russian, but the copies changed the title, description, and annotations to other languages, including English, Chinese, and Spanish. Each of the stolen copies had tens of thousands to hundreds of thousands of views and some copies were as recent as the month of writing this paper. While the obvious consequences of such theft include unfair advertisement revenue, copyright infringement and such, there could be worse consequences. In at least one of the stolen copies of the video, the edits to both the original video and the description made it look pornographic, which could not only malign the reputation of the original uploader but may also hurt them in other ways. Further, edits to videos would make it hard to automatically detect the stolen copies, and few individuals are likely to have the resources to follow the legal recourse to protect themselves or their copyright, especially if the copyright infringer resides in a different country. Due to these reasons, we believe that the problem of fraudulent video promotion deserves attention.

Unfortunately, current measures taken to find fraudulently promoted videos seem insufficient. As an example, YouTube deleted only 30% of the fraudulent videos and closed or suspended 16% of the fraudulent accounts in our data set. Moreover, YouTube purged fake reviews from videos during the time of our experimentation and only one of the fraud videos in our data set lost views, suggesting that many fraudulently promoted videos were probably not impacted.

Given this backdrop, this paper takes steps in characterizing and defending against fraudulent video promotion. In order to gather data for our investigation, we promote two YouTube videos each for a month at two popular *exchange* websites, `vagex.com` and `viewtubetrain.com`. We also bought a paid package from the most popular freelance website, `fiverr.com`, to promote a third YouTube video.

The first contribution of our work is a *characterization of fraudulently promoted videos*. Using a data set of over 3,300 fraudulent and legitimate videos each and over 500 fraudulent and legitimate profiles, we learn that fraudulent videos underdescribe themselves while extensively pointing to uploaders' Facebook and Twitter pages. An average fraud video has shorter and fewer comments but is rated higher – 4.6 on a 5-point scale when an average legitimate video is rated only at 3.6. The profiles which promote the fraudulent videos, referred to as "fraudulent profiles" subsequently in this paper, also have distinct characteristics. They are relatively new in the system but more active than legitimate profiles. They are more active in viewing and interacting with videos and rarely upload any videos.

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2578996.

The second contribution of this paper is the *development of two supervised learning-based classifiers to identify fraudulently promoted videos and profiles*. Our video classifier has an accuracy of 91.3%, and the profile classifier has an accuracy of 99.2%, suggesting that the approach is promising. With sites like YouTube hosting millions of videos, the false positives of the video classifier are higher than desired. We attribute the less-than-desired accuracy to the limited availability of data and the lack of perfect ground truth in training the classifier. We speculate that the sites themselves could adopt our approach and get significantly better results.

2 Background

2.1 Basic YouTube functionality

The YouTube website allows two basic functionalities: viewing and uploading videos. Videos can be searched by keywords or browsed by categories. Viewing a video does not require a YouTube account. However, uploading a video requires one to create an account, or a *channel* in YouTube parlance. An account without an uploaded video is more appropriately referred to as a "profile" and not a channel. But for simplicity, we use the term "profile" for all types of accounts subsequently. All profiles can comment on the videos they watch, rate them with a one- to five-star rating system by liking or disliking them, designate videos as favorites, subscribe to other members' channels, share videos, and befriend other YouTube users.

A typical profile has various pieces of information about the owner and their YouTube activity. The personal information includes interests, education, occupation, and date the user joined YouTube. Each profile also displays thumbnails of the videos uploaded by the channel owner and their other activity. The latter includes the names of the channels they have subscribed to as well as videos of other YouTube members they have commented on, picked as favorites, liked or disliked. Additionally, if a YouTube member comments on a channel, the comment is visible to everyone.

Each public video also has various pieces of information associated with it. An uploader provides a title and a category for each video they upload, along with keywords[1] and a description. The description can include links. In addition, YouTube provides the date of the upload, number of times the video has been viewed, commented on, and liked or disliked. Comments are also public.

2.2 Fraudulent video promotion

It is generally believed that the number of times a video is viewed; liked or disliked; favorited; and commented on, including responded to with videos, plays a role in determining video rankings. Moreover, subscriptions to the channel of the video uploader are also known to play a role [2]. Various online services claim to increase video rankings. The first kind offers various promotional packages that increase the views and ratings of a video, add to the number of times it is favorited and commented on, and even increase the number of subscriptions to the uploader's channel in exchange for money. The second type of video promotion services promote a member's videos in exchange for them promoting other members' videos and are referred to as *exchange* sites. When a user registers on an exchange website, she downloads a browser plugin (or equivalent standalone software). The plugin opens a new window in the browser, fetches the YouTube ID of the next video to be watched from the exchange servers it is configured with, loads `www.youtube.com`, and starts watching other members' videos on the user's behalf. In addition, standalone software such as the one offered by `www.vagex.com` can automatically like, comment, favorite, and subscribe to a channel. The user does not even have to be around while the software/plugin is working and earning points, which can later be used to promote user's video(s).

[1] As of June 2013, YouTube no longer makes the keyword information publicly available.

3 Data collection

3.1 Fraudulent videos and user profiles

We adopted two strategies for collecting data about fraudulently ranked videos. In the first, we used two popular video exchange websites, `vagex.com` and `viewtubetrain.com`, to 'promote' two short, 30-second videos showing hummingbirds drinking nectar from the bird feeder. Corresponding to each video, we created a YouTube profile and uploaded the video. In order to prevent any organic views, we made each video unlisted, meaning that it was not searchable through YouTube and could only be accessed through the URL we provided. We promoted each video at the exchange sites for one month and then recorded statistics through the YouTube API.

In exchange for this promotion, we allowed the plugins we downloaded from each exchange website to view videos of other members registered at these sites. The plugins could not take other actions, such as liking, rating, or commenting on a video.

In the second strategy, our goal was to capture paid programs that increase video rankings. Toward this goal, we used `fiverr.com`, a popular freelance website containing services offerers are willing to do for $5. There, we purchased a package that promised 1000 views, and 100 each of likes, favorites, comments, and subscriptions for our video for $5.

In addition to our own test videos, we consider all videos watched by our plugins from the two exchange websites as fraudulent since they were part of the exchange system at the time of our experiment. Similarly, all user profiles that either watched or took any action on our three videos were fraudulent as well. However, since the YouTube API does not provide any information about which profiles watched which videos, our sample of fraudulent profiles only contains those that took any action. In fact, since the `viewtubetrain.com` plugin did not possess the functionality for taking actions, this API limitation precluded adding any fraudulent profiles to our sample from this website. However, since we were able to add fraudulent videos to our sample from this site, the exercise of experimenting with this website was useful nonetheless. Note that due to the possibility of organic views or actions[2] on fraudulently ranked videos, we cannot be certain that all user profiles that view a fraudulent video are fraudulent themselves. Therefore, we do not consider profiles that have taken any action on a fraudulent video as fraudulent and exclude them from out data set. Overall, our data set contains 3,308 fraudulent videos and 502 fraudulent profiles.

Interestingly, 40% of the fraudulent profiles from the video we promoted on `fiverr.com` were the same profiles we encountered at `vagex.com`. It follows from this that the offerer of this service was using `vagex.com` and possibly other exchange websites to deliver the order. In fact, this observation is further confirmed by the ways in which the order fell short of the promised actions but overdelivered on the views, implying lack of control from the offerer. Specifically, the order added 2,679 views to our video instead of the promised 1,000, but only added 55 likes, 55 favorites, 25 comments, and 75 subscriptions when 100 of each were promised.

3.2 Legitimate videos and user profiles

In order to understand how fraudulent videos and user profiles differ from the legitimate ones, we collected a random sample of

[2] We define *organic action* as a non-automated video interaction – such as viewing a video, leaving a comment or favoriting a video – by the user who is genuinely interested in the provided content.

YouTube videos and profiles. This required special thought since we wanted truly 'random' videos and YouTube only provides popular videos in various categories. One option would be to pick random video IDs, but that is not feasible due to the size of the space and non-contiguous nature of IDs. So, we used a frequency dictionary of the top 5,000 most popular English words to collect an unbiased video sample. The process involved randomly selecting 1-3 words from the dictionary according to their frequencies, issuing a YouTube query to search for the selected words, and then randomly ordering the results by one of the 15 categories provided by YouTube plus one of the following: upload date, relevance, view count, and average rating. This process netted us 4,000 random/legitimate videos.

While the profile IDs of users who viewed, liked/disliked, or favorited a specific video are not available, those of users who commented on or uploaded a video are. We used this availability to randomly pick 10-15 user profiles from each random video. The selection process was stopped once we had gotten 529 random user profiles.

3.3 Data snapshots

We made two data snapshots in order to better understand the behavior of legitimate and fraudulent YouTube profiles and videos over time. The first snapshot was taken one month after the initial data collection efforts and the second snapshot was taken a year later.

4 Characteristics of fraudulent videos

The video characteristics described in this section ultimately become features for classification. Table 1 shows the overview of data we collected on fraudulent and legitimate videos at the first snapshot. About 12% of both fraudulent and legitimate videos were deleted by YouTube at that time. More were deleted at the time of the second snapshot, bringing the total deleted to 30%. For the rest of the analysis in this section, we focus on the 88% of both types of videos that survived (first snapshot). We learn that an average fraudulent video has been in the system for a shorter duration than an average legitimate video. In addition, a fraudulent video is more likely to have fewer views, likes/dislikes, and comments then a legitimate video, primarily because the variation of these parameters for legitimate videos is high and a few videos in our data set are wildly popular. In contrast, the average ranking of a fraudulent video (derived from weighting the number of likes and dislikes on a scale of 1 to 5) is higher than that of a legitimate video, which is unsurprising since fewer of the latter's ratings are likely to be organic, and the very purpose of fake ratings is to like the video, not dislike it.

	Fraud videos	**Legitimate videos**
Total	3,308	4,000
Deleted by YouTube or uploader	394	484
Average age of video	141	594
Average views	71,734	881,260
Average likes/dislikes	468	3913
Average rating	4.6	3.6
Average comments	651	1,571

Table 1: Overview of data collected on videos

4.1 Upload characteristics

Our data shows that most YouTube videos are short. In fact, the average length of legitimate and fraudulent videos seems to be the same, with the vast majority of videos being shorter than 500 seconds.

The description of a video plays an important role in searching of videos. When looking at the length of description for fraudulent and legitimate videos both in terms of the number of characters as well as words, we find that *fraudulent videos have shorter descriptions*, which confirms our observation. However, the differences are relatively small.

YouTube video descriptions are allowed to contain URLs, both external and internal to YouTube. When looking at the URLs contained in the videos, we find that most of them are external and many point to Facebook and Twitter pages. Upon examining the count of URLs in descriptions, we find that *61.6% of fraudulent videos contain at least one URL as opposed to 41.1% of the legitimate videos*. This is intuitive given that many fraudulent videos promote websites.

4.2 Viewership-related characteristics

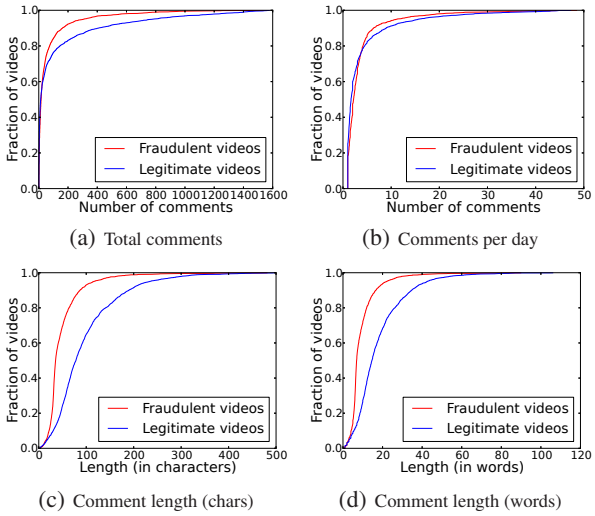
We noted earlier in this section that an average fraudulent video had fewer views than a typical legitimate video. When looking at the details, we find that average views for legitimate videos is somewhat misleading; 80% of them have very few views, but a small number have a very large number of views. In contrast, the viewership is less extreme for fraudulent videos. Given this diversity, we look at views per day instead. We found that *almost 90% of legitimate videos have very few views per day whilst fraudulent videos tend to get a relatively high number of views for each day that they are active*.

We also examined likes and dislikes, finding that hardly any fraudulent videos have any dislikes but 43% of legitimate videos have at least one dislike. Also, 3/4 of legitimate videos have very few to no likes, but less than a third of the fraudulent videos have zero likes. Given this, the fact that an average fraudulent video has a rating of 4.6 on a 5-point scale while an average legitimate one is rated at 3.6 can be explained easily.

The last action item on a video is leaving a comment, where viewers can express their opinion of the video. Figures 1(a) and 1(b) show the total comments and average comments per day for fraudulent and legitimate videos. The graphs have been truncated at 1,600 total comments and 50 comments per day to eliminate outliers. We find that 1/5 of fraudulent and legitimate videos are uncommented. On the other hand, among the videos that are commented on, legitimate videos tend to have more comments. This contrast caused average number of comments for legitimate videos to be higher in Table 1. Similarly, fraudulent videos tend to receive fewer comments per day than their legitimate counterparts. When looking at the length of comments in terms of characters as well as words, we find that *shorter comments are left on fraudulent videos*. Specifically, 60% of fraudulent videos have an average comment length of 8 words or less while only 15% of the legitimate ones have comments shorter than 8 words (see figures 1(c) and 1(d), which have both been truncated at 500 characters and 120 words, respectively). We also performed frequency analysis on the words that are most commonly used in legitimate and fraudulent videos as well as the total variation in the texts of the comments. The results suggest that fraudulent videos tend to have a lot of similar comments whilst comments left on legitimate videos are more varied.

5 Characteristics of fraudulent user profiles

Table 2 shows the overview of our data on fraudulent and legitimate user profiles. Similar to their video counterparts, the attributes discussed in this section were used as features for training user profile classifiers. The first thing we observe is that, in the short run, YouTube is as likely to shut a fraudulent profile as it is to shut a legitimate one; just over 3% each of fraudulent and legitimate profiles were closed by YouTube at the first snapshot. Even more were deleted at the time of the second snapshot, bringing the total number of deleted fraudulent and legitimate profiles to 16 and



(a) Total comments (b) Comments per day

(c) Comment length (chars) (d) Comment length (words)

Figure 1: Comments on fraudulent and legitimate videos

6 percent, respectively.[3] This points to the importance of timely identification of fraudulent profiles. We learn that a higher percentage of fraudulent profiles claim to be males. The average age of the account owner for both fraudulent and legitimate profiles is similar. Also, *fraudulent profiles have been registered more recently but are more active than legitimate profiles*. They also upload fewer videos, which is intuitive since many may not have a video to promote, and even if they do, they may choose to keep their uploader profile separate. This translates into fewer views for their channels and hence fewer subscribers.

	Fraud profiles	**Legitimate profiles**
Total	502	529
Shut by YouTube	17	17
% male (of alive profiles)	78.8%	68.3%
Days on YouTube	420	818
Days since last activity	151	200
Average age of profile owner	28	27
Average videos uploaded	16	76
Average profile views	25,824	3,765,702
Profiles without subscribers	32	129
Average subscribers to profiles	164	8,339

Table 2: Overview of data collected on profiles

Next, we explore these average statistics in detail. Looking at age, as reported by the profile owner at the time of account registration, we find that both types of profiles are close in age and that the average registrant age reported in Table 2 is a good indication of the spread of ages of the registrants. Figure 2(a) compares the number of days each type of profile has been alive on YouTube. To infer how long a profile has been active, we look at the registration date provided by the API. We find that approximately 5% of legitimate profiles have been alive for less than a day. In contrast, hardly any fraudulent profiles are new to the system. However, 3/4 of the legitimate profiles have been alive for over one year when only 1/3 of the fraudulent ones have been alive for that long. This implies that *fraudulent profiles are relatively new to the system*.

Fraudulent profiles tend not to upload any videos. This is intuitive since not all of them may be interested in promoting their videos in exchange for promoting other's videos. Even if they did have videos to promote and were engaging in fraudulently ranking other's videos through the exchange programs similar to `vagex.com`, they may have many such profiles, most or all of which would not have the video they are trying to promote. Figure 2(b) shows the number of videos uploaded by legitimate and fraudulent profiles. We note that a quarter of the legitimate profiles upload over a dozen videos while under 10% of the fraudulent profiles upload that many videos. Further, 60% of the fraudulent profiles upload no videos at all when only 40% of the legitimate profiles have an account with YouTube but upload no videos.

Next, we check how the activity of fraudulent profiles in terms of liking or disliking videos, commenting on videos, or uploading new videos compares with those of legitimate profiles. Overall, we find that *legitimate profiles take fewer actions on an average day compared to fraudulent profiles*. Specifically, 95% of legitimate profiles take 6 or fewer actions in a typical day. In contrast, only 8% of fraudulent profiles take 6 or fewer actions. Figure 2(c) shows the daily activity of both kinds of profiles.

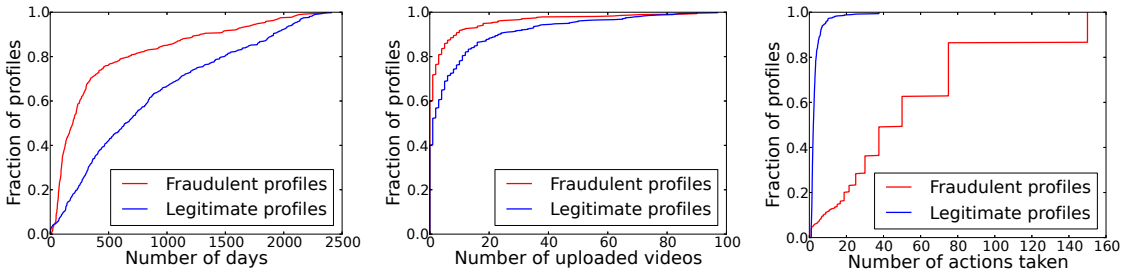
Investigating the commenting activity in detail, we find (expectedly) that *fraudulent profiles leave more total comments as well as comments per day*. Specifically, 95% of legitimate profiles leave 200 or fewer comments while only 20% of fraudulent profiles leave so few comments. The conclusion is similar for average number of comments per day, except that the difference among fraudulent and legitimate profiles is even more pronounced. In particular, 95% of legitimate profiles leave three or fewer comments per day while under 3% of fraudulent profiles have this level of commenting activity on an average day. These observations clearly indicate that *fraudulent profiles are more aggressive in interacting with videos*. Further, we examine the average length of comments left by fraudulent profiles and find that 95% of them are shorter than 32 characters, while only 10% of legitimate profiles leave such short comments. This observation is intuitive and similar to the one we made on the lengths of comments for fraudulent videos in Figures 1(c) and 1(d).

Though YouTube's formula for ranking videos is unknown, many online articles claim to have reverse engineered it. A few of them claim that a video's ranking will increase when it is shared with other people or when other profiles put in in their favorite playlists. Unfortunately, the YouTube API did not provide a way for us to infer when a profile shared a video or when a specific video was shared by some profile. It also did not provide a way for us to infer when a video was in someone's favorites playlist. However, it let us infer the favorites playlist of profiles. Figure 3(a) shows the size of the favorites playlist for fraudulent and legitimate profiles. We learn that *fraudulent profiles have more videos in their favorites playlist*. In fact, 3/4ths of legitimate profiles have fewer than 100 videos in their favorites playlist, while only 25% of the fraudulent profiles have so few videos in their favorites playlist. Since the YouTube API let us infer other playlists for profiles as well, we plot the number of videos in other, 'regular' playlists in Figure 3(b). The picture is opposite for regular playlists in that 95% of fraudulent profiles have fewer than 25 videos across all their regular playlists combined. In fact, 86% of them have zero videos in their regular playlists. In contrast, only 2/3rds of legitimate profiles have fewer than 25 videos in their playlists, and less than half have zero videos across all their regular playlists.

6 Identifying fraudulent videos and profiles

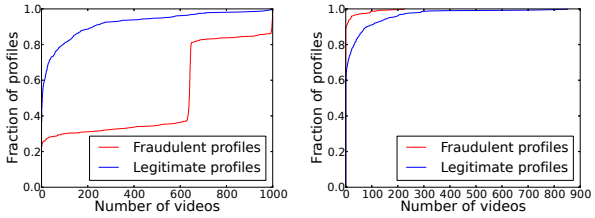
Next, we utilized supervised machine-learning-based classifiers and trained two separate classifiers, one each for videos and user profiles. Specifically, we used the RapidMiner data mining package [3], which allows for switching across classifiers and modifying their settings. For each classification experiment, we used a 10-fold cross-validation with stratified sampling. In a 10-fold cross-validation, the sample is divided into 10 parts: 9 parts are used as a training dataset, while the remaining part is used to test the clas-

[3]We use the data from the first snapshot for the rest of this section.



(a) Number of days active (b) Number of videos uploaded (c) Profile activity

Figure 2: Fraudulent vs. legitimate profiles



(a) Favorite playlists (b) Regular playlists

Figure 3: Regular and favorite playlists of profiles

sifier. This process is repeated 10 times, producing 10 results. The results reported subsequently are averages of the 10 runs.

We experimented with a number of classifiers and discovered that basic learners such as linear SVM were not powerful enough to learn the model. On the other hand, the feature weights produced by much more powerful learning algorithms (such as a neural network) were relatively large numbers with alternating signs, which is a sign of overfitting. As a result, we ended up utilizing a decision tree and a SVM-based classifier with a non-linear kernel (Anova), both of which seemed to perform well on our data sets.

6.1 Attribute usefulness

We ranked the importance of all profile and video attributes by using two feature selection methods, namely chi-squared and information gain [4]. According to them, video age and average number of views per day are more useful than other features. On the profile front, chi-squared and information gain tests both agree that average number of comments left per day, recent activity, and comment length are the most useful user profile attributes of the ones we considered. Our experiments also confirm that those features are more discriminating than others.

6.2 Video classifiers

Since we do not know the fraction of good to bad videos on YouTube, we decided to pick an equal number of good and bad videos from our data sets to train the video classifiers (6,000 videos in total). Ideally, we would use stratified sampling to come up with a representative video sample, but, unfortunately, YouTube does not release the numbers (such as percentage of videos by category) that are required to take advantage of this sampling technique. As a result, we ended up using a naive sampling approach for the good video dataset.

However, there is no guarantee that all videos in the legitimate data set are in fact legitimate. Given our estimate of the incidence rate of fraud to be 0.5%,[4] we expect about 15 of them actually to be fraudulent. To assist with this, we trained an initial classifier and then used it to find videos that might be bad and reclassify them manually. This is a well-known technique in machine learning to deal with the issue of less-than-perfectly labeled ground truth.

[4]We manually analyzed 200 random YouTube videos and only one of them appeared to be fraudulent.

The decision-tree-based ensemble of classifiers outperformed the rest with an accuracy of 91.3%. The false positive rate of our tree ensemble is 8.4%, and the false negative rate is 9.1%. A single decision tree classifier took the second place with 88.5% accuracy (10.6% false positive rate and 12.6% false negative rate). The third best was an SVM classifier with an Anova kernel with an accuracy of 86% (13.3% false positive rate and 14.9% false negative rate).

Unfortunately, the number of features available through the YouTube API has been reduced since our initial data collection efforts – features such as video keywords and number of times a particular video has been favorited are no longer available. This resulted in a lower classifier accuracy (94.7% before the API change versus 91.3% now).

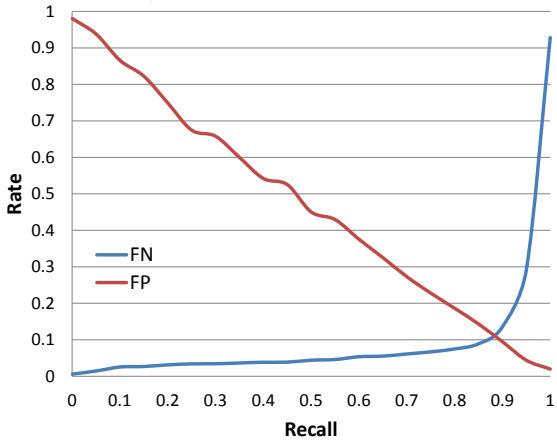


Figure 4: Relationship of recall and false positive/negative rates

Video classifier tuning: Modifying the minimum recall[5] parameter allowed us to manipulate the video classifier's false positive and negative rates. As Figure 4 shows, tuning the classifier to a low false negative rate results in a high false positive rate, which means that the algorithm would rarely miss a bad video whilst marking many legitimate videos as fraud. Although it is not ideal, it would be suitable for a first line of defense solution that uses additional mechanisms to label videos as fraudulent or not. Our decision tree video classifier is relatively fast and, if properly tuned, could be used to quickly filter the videos in two subsets – good and potentially bad. In the next step, the potentially bad subset can be further analyzed manually or via a slower, more powerful machine learning algorithm.

6.3 User profile classifiers

Similar to the video classifier training, we chose an equal number of user profiles from good and bad profile data sets (1,000 accounts in total).

[5]Recall is calculated by dividing the number of true positives by the number of all positive cases.

The best user profile classifier performed better than its video counterpart, averaging 99.2% in accuracy. Unlike with the video algorithm training, the SVM classifier with the Anova kernel came up on top, with the decision tree classifier taking the second place with an accuracy of 98.7% (1.0% false positive rate and 1.5% false negative rate). The false positive rate of our user profile SVM classifier is 0.2% and the false negative rate is 1.5%.

7 Related works

Although there have been a number of studies on user behavior and spam in online social networks, very few of them have concentrated on classifying YouTube accounts and videos that engage in fraudulent video promotion. The former topic has been discussed in [5] where the authors investigate the types and duration of activities performed by users on popular social network sites. The latter topic was partially covered by Benevuto et al. in [6], [7] as well as Hendrickson et al. in [8] where the authors looked at the video responses (i.e. videos uploaded in response to another video) on YouTube and tried to identify videos and misbehaving user profiles that pollute YouTube by uploading unrelated video responses. In addition, O'Callaghan et al. [9] and Sureka [10] investigate potential comment spammers by looking at the comments left on YouTube and their characteristics such as length, number of repetitions, and others. Similar to the above-mentioned papers, we use a number of video and user profile characteristics to train the machine learning models. We look at the videos and user profiles that were participating in the fraudulent video promotion programs while most other studies seem to concentrate on video and comment spam.

8 Discussion

Do fraudulent videos have organic views or actions? All videos we watched through the exchange programs are fraudulently ranked. While many of their views are likely to be fraudulent, we wondered if they possessed organic views from legitimate profiles. To answer this question, we gathered all user profiles that took any action on our data set of fraudulent videos and subjected them to our profile classifier. Of the 57,548 profiles that commented on the 3,308 fraudulent videos in our data set, our classifier marked 52% (30,000) as legitimate. This implies that *fraudulently promoted videos also attract organic views and actions* at least at some point in their life. While this observation is simple, it suggests that the strategy of fraudulently ranking videos is fruitful.

Do fraudulent user profiles always misbehave? All user profiles that take any action on our three test videos are fraudulent. We wanted to see if they ever view or take action on legitimate videos also. If not, this would indicate that they are dedicated profiles generated for fraudulent purposes only. To satisfy this curiosity, we took each of the 502 fraudulent profiles in our data set and subjected all the videos they had commented on to our video classifier. Of the 19,441 videos they took any action on, our classifier marked 1.93% (375) as legitimate, suggesting that *fraudulent profiles in our data set exist solely for the purpose of fraudulently ranking videos*.

Limitations of the API: The highest stable YouTube API version available at the time of our experiment allowed the collection of a wide range of video and user profile statistics, but not fine-grained details, such as profile IDs and IP addresses of users who have watched specific videos, the time that a user spent watching a video, and others. These features would have been very useful in our study, especially considering the fact that it is possible to watch a video without having a YouTube account (which suggests that most people did just that and there is no way now to get their profile information). In addition, the API limited the number of results one could get at a time to 50. This increased the time needed to gather the data. Further, not only is it not possible to retrieve any information beyond the 1,000th result through this API, but one of the most interesting features to us, user activity feed, is limited to only 150 recent entries. Moreover, the YouTube API will not retrieve the entries in the user activity feed that are older than 60 days. This made connecting videos and user profiles together more challenging since fraudulent accounts tend to be very active, and the above-mentioned limitations reduce the amount of data that can be collected. Consequently, we had to write a Google scraper in order to find videos which where commented on by user profiles from our data sets. This allowed us to have profile features such as number of comments left, comment length, number of comments left per day, and others.

9 Conclusion

In this paper, we subjected three test videos to services that fraudulently promote the rankings of YouTube videos. This allowed us to collect ground truth for fraudulent and legitimate user profiles and videos. Through a characterization of the gathered data, we were able to identify features that yield promising supervised machine learning classifiers.

10 Acknowledgements

We would like to thank Predrag Radivojac of Indiana University for his help with classifier tuning.

11 References

[1] "YouTube Videos Pull In Real Money," http://www.nytimes.com/2011/10/27/technology/personaltech/cashing-in-on-your-hit-youtube-video.html.

[2] G. Chatzopoulou, C. Sheng, and M. Faloutsos, "A first step towards understanding popularity in YouTube," in *IEEE Infocom workshop*, 2010.

[3] "RapidMiner," http://rapid-i.com/.

[4] Y. Yang and J. O. Pedersen, "A comparative study on feature selection in text categorization," in *International Conference on Machine Learning (ICML)*, 1997.

[5] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in *ACM SIGCOMM Internet Measurement Conference (IMC)*, 2009.

[6] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, C. Zhang, and K. Ross, "Identifying video spammers in online social networks," in *ACM International Workshop on Adversarial Information Retrieval on the Web (AIRWeb)*, 2008.

[7] F. Benevenuto, T. Rodrigues, V. Almeida, J. Almeida, and M. Goncalves, "Detecting spammers and content promoters in online video social networks," in *ACM International SIGIR Conference on Research & Development of Information Retrieval*, 2009.

[8] H. R. Langbehn, S. M. R. Ricci, M. A. Gonçalves, J. M. Almeida, G. L. Pappa, and F. Benevenuto, "A multi-view approach for detecting non-cooperative users in online video sharing systems." *JIDM*, vol. 1, no. 3, 2010.

[9] D. O'Callaghan, M. Harrigan, J. Carthy, and P. Cunningham, "Identifying discriminating network motifs in YouTube spam," *CoRR*, vol. abs/1202.5216, 2012.

[10] A. Sureka, "Mining user comment activity for detecting forum spammers in youtube," *CoRR*, vol. abs/1103.5044, 2011.