# Spam Identification on Facebook During Events

Prateek Dewan, Ponnurangam Kumaraguru
Indraprastha Institute of Information Technology - Delhi, India
{prateek,pk}@iiitd.ac.in

## 1. ABSTRACT

Over 800 million users log into the Facebook social network every day. [1] In addition to socializing, people also use Facebook to stay updated about latest events and happenings around the world [2]. This aspect of information consumption from Facebook during events is heavily exploited by spammers. In a recent example, the viral nature of the famous biting incident during the 2014 FIFA World Cup was exploited by spammers to spread phishing links. Multiple Facebook posts prompted users to sign a fake petition in defence of the Uruguayan player, who was banned for biting an opponent. [2] OSM activity rises manifold during events like sports, natural calamities etc. [4], thus making Facebook an even more lucrative breeding ground for spammers. The aforementioned spam activity indicates that existing approaches being used by Facebook [3] do not suffice in countering spam, especially during events, when their spread is maximum. This behavior not only violates Facebook's terms of service, but also degrades user experience. Facebook has confirmed spam as a serious issue and taken steps to reduce spam content in users' newsfeed recently. [3] Identifying Facebook spam, however, remains a challenge because of lack of profile and network features, which have previously been heavily used for spam detection on other OSMs.

The aim of this work is to develop automated techniques to identify and combat spam on Facebook during events. We define spam as content which is irrelevant / unrelated to the event under consideration, and / or aimed at spreading phishing, malware, advertisements, self promotion etc., including bulk messages, profanity, insults, malicious links, scams, fake information etc. To this end, we collected over 6.4 million public posts on Facebook during 16 events (from April 2013 to Dec. 2013) from Facebook's Graph API Search. We then identified and manually annotated the top 220 most active users / pages in our dataset (about 15 top users / pages in each event) as spammers and non spammers (Table 1), by looking at the content they posted.

Using a set of 21 features (drawn from previous literature on spam) extracted from these posts, we applied machine learning algorithms to identify spam, and achieved an accuracy of over 99% using the Random Forest classifier (Table 2). Further analysis revealed that Facebook pages posted 4 times more spam than users. We also discovered that spam posts had more repetition as compared to legitimate posts.

| | Spam | Non spam | Total |
|---|---|---|---|
| Users | 25 | 195 | 220 |
| Posts | 7,882 | 58,806 | 66,688 |

**Table 1: Annotated sample dataset of most active users during the 16 events.**

| Classifier | Accuracy | F-Score |
|---|---|---|
| Naive Bayes | 93.00% | 0.927 |
| J48 DecisionTree | 99.82% | 0.998 |
| Random Forest | **99.90**% | 0.999 |

**Table 2: Classification results for Naive Bayes, J48 Decision Tree, and Random Forest classifiers.**

Existing approaches to combat spam on Facebook, have concentrated on identifying spam users and campaigns by making use of the URLs posted on the network. However, multiple advertising and self-promotional campaigns on Facebook don't make use of URLs at all to spread spam. It is therefore, impossible for existing techniques to be able to identify and combat such categories of spam. Spam detection approaches used in other OSM services like Twitter [1] cannot be directly ported to Facebook due to the public unavailability of critical pieces of information like profile, and network information, no limit on post length etc.

We are in process of building a bigger labeled dataset for spam using human annotators. In future, we intend to combine content and platform specific features extracted from this labeled dataset with those previously used for detecting spam with URLs, to generate an exhaustive feature set to identify spam. This exhaustive feature set will be fed to multiple machine learning classification algorithms to create an efficient, real time spam detection framework to combat spam on Facebook during events. This approach may also cater to spam which does not contain URLs. To the best of our knowledge, this is the first attempt at identifying spam on Facebook during events.

[1] http://newsroom.fb.com/company-info/

[2] http://www.marca.com/2014/07/18/en/football/barcelona/1405709402.html

[3] http://newsroom.fb.com/news/2014/04/news-feed-fyi-cleaning-up-news-feed-spam


*ACM Conference on Online Social Networks (COSN)* Dublin, Ireland


## 2. REFERENCES

[1] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting spammers on twitter. *CEAS*, 6(12), 2010.

[2] J. Holcomb, J. Gottfried, and A. Mitchell. News use across social media platforms. *Technical report, Pew Research Center.*, 2013.

[3] T. Stein, E. Chen, and K. Mangla. Facebook immune system. In *Proceedings of the 4th Workshop on Social Network Systems*, page 8. ACM, 2011.

[4] M. Szell, S. Grauwin, and C. Ratti. Contraction of online response to major events. *PLoS ONE 9(2): e89052, MIT*, 2014.