Spam Detection using Approach of Data Mining for Social Networking Sites

Ritesh Kumar
Department of Computer Science,
Sinhgad Institute of Technology and Science,
Narhe, Pune-411041

Shital Ghadge
Department of Computer Science,
Sinhgad Institute of Technology and Science,
Narhe, Pune-411041

G.S. Navale
Department of Computer Science,
Sinhgad Institute of Technology and Science,
Narhe, Pune-411041

ABSTRACT
Increase in number of spam incidents is causing a very serious threat to Social Networking World which has in turn become an important means of interaction and communication between public users. It is not only dangerous to the public users, but it also covers much of the bandwidth of the Internet traffic. Most of current spam filters in use are based on the subject content of email, Facebook, twitter. Social Networking Services also provide great possibilities to take advantage of user identification and other social graph-dependent features to improve classification. In this paper, the proposed System uses machine learning [3] approach for spam detection based on features extracted from social networks constructed from social networking site message metadata and logs. Flags and scores are assigned to senders based on their possibility of being a legitimate sender or spammer. Moreover, proposed System also explores various spam filtering techniques and possibilities.

Social networking sites are vulnerable to mass spam incidence as well as users data theft such as credit card details, user activities and users taste for criminal purposes .Email subject headers are used to check spam email, spam on Social networking Sites is often accompanied by a wealth of data on the sender, metadata can be used to build more accurate detection mechanisms. System uses these terminologies to choose features that best differentiate spammers from legitimate users. On basis of this technique system flag user system or message as spam and legitimate messages.

General Terms
Data Mining, Spam Detection

Keywords
SNS (Social Network Services) [2], OSN (Online Social Network) [2], tf-idf (Term Frequency-Inverse Document Frequency).

1. INTRODUCTION
The boom in use of social media in public users has made Social Networking Services [2] most appropriate targets for spam and fraud, vulnerable to mass attacks. We can see this in recent mass attack on US celebrities code named as 'The Fappening'. Earlier checking on email subject was done to check spam messages. Many of these techniques previously used to combat conventional email and web spam. Social Networking Services [2] give facility to take advantage of user identification and other social graph-dependent features to improve classification. Majority of research has been carried out on public research available data from Social Network System, making it difficult up until now to measure the effect of private user data on algorithms for detecting site misuse.

The System proposed to reconstruct spam messages into categories for classification rather than checking them individually. Although categorization and identification has been done for offline spam analysis, we apply this technique to be used in the online spam detection problem with sufficiently low overhead. Our proposed system carries a set of parsing algorithm that effectively distinguishes spam messages. It Flags messages classified as "spam" before they reach the intended recipients, thus protecting them from various kinds of fraud.

2. BACKGROUND
2.1 Data Mining
Earlier checking for spam is done manually from data set or generic algorithms were used in case of text based spam analysis. In old times algorithm like regression analysis, Bayes' theorem and keyword matching [3] were used for uncovering patterns or spam ,this algorithm were incapable of handling large data set and can be applied on small data set.

As data sets have grown in size and complexity, direct analysis is next to impossible or is time consuming. Using Machine learning [1] data processing technique, such as big data analysis, cluster analysis, decision trees and decision rules, and support vector machines we can easily disclose the hidden pattern and spam. Data mining can be easily applied using these methods which help in discovering hidden patterns in large data sets.

3. CASE STUDY
The proposed system uses social networking sites as the example application. System can start monitoring of new posting activities in social media network. For each new instance, System first make prediction based on the algorithm, if it is uncertain, send the instance for manual labeling using human interference

Note. This demo uses Facebook as the example since it is currently the most popular social media website. However, the proposed system can be easily generalized and applied to other social sharing websites, such as Twitter and YouTube.

3.1 Requirements
- Add, manage and filter keywords and Spam/Flag System.
- View term frequency – Inverse document frequency [TF-IDF] algorithm.
- NLP Application [Natural Language Parser].

3.2 Mathematical Model

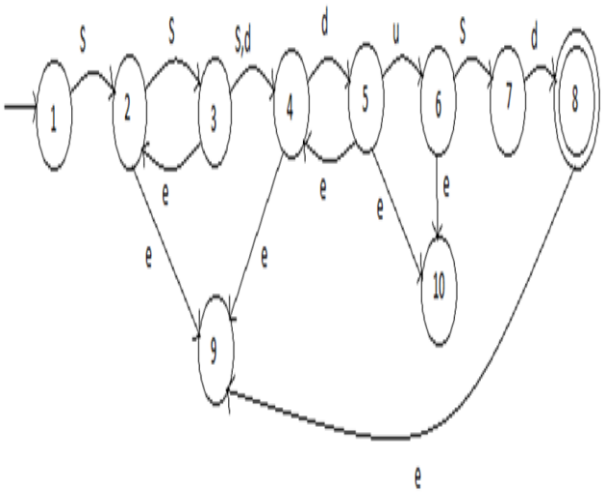

Fig 1: Finite State Machine

Where,

States:

1. Input User Messages
2. Keyword Parser
3. Tokens
4. Filtering
5. TF-IDF Algorithm
6. Message Category Identification
7. Spam Detection
8. NLP Parser(Grammar/Construct Message)
9. Connection error
10. Cannot connect to database

Transitions:

S: Success

e: Error

d: Update database

3.3 Behavioral Model

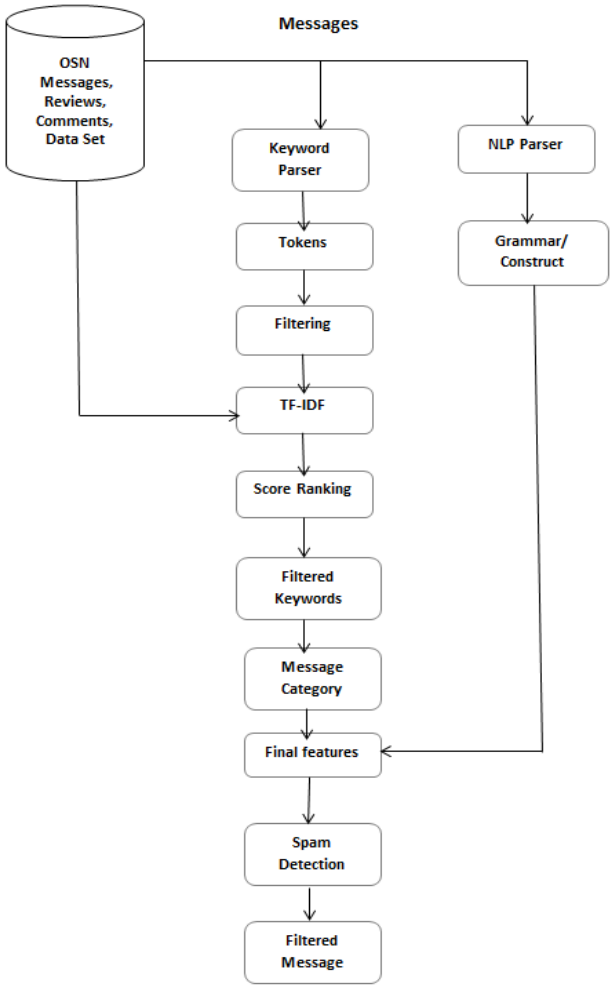

Fig 2: Flow Chart

As the computations required to be performed fall into sequential and loop construct, it remains polynomial time solvable. Therefore, the proposed system remains polynomial time solvable.

3.4 System Architecture

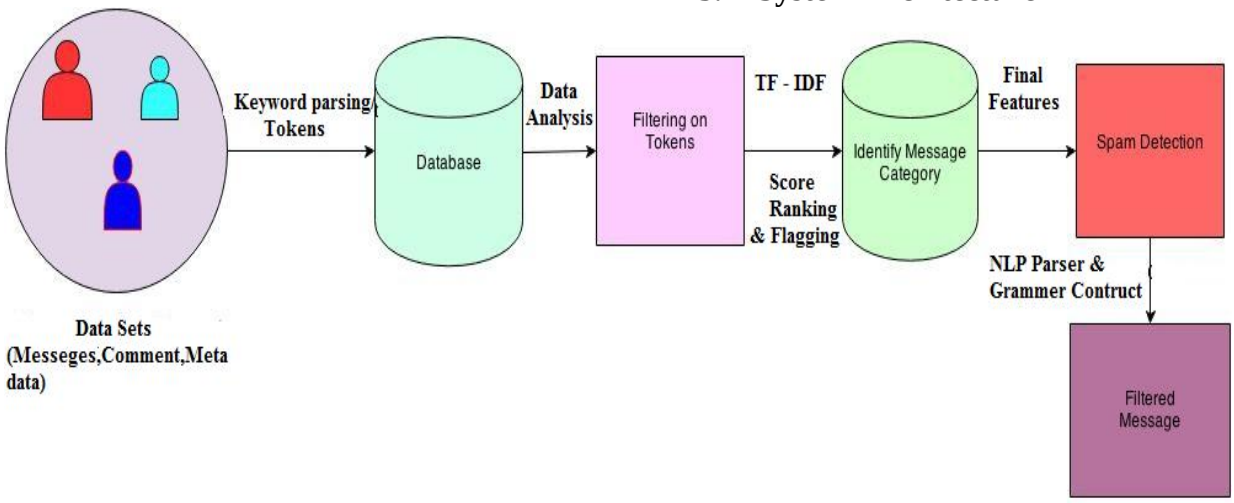

Fig 3: System Architecture

As shown in Figure, the system architecture works as follows. In the first stage, we collect social media content (including text and images) performs machine learning to build classified documents and identify spams. In the second stage, we monitor the activity of user, make prediction on basis of algorithm and send flag or spam alert to client about detected spams and update the model.

4. CONCLUSION

The paper Demonstrates the OSNs [2] (Online Social Networks) using data mining to detect the spam. For the spam detection, the blacklist, keyword blocking is applied first. Then, the data mining model or algorithm is used to detect the spam further. This proposed system can be applied over several social networking sites such as Facebook, Twitter and YouTube etc.

The future work includes practical implementation of Spam filtering in real-time environment. Also, the ability to handle more languages other than English as the proposed System can process only English as a language while filtering spam due to natural language processor restrictions are also left as a part of future work.

5. ACKNOWLEDGMENTS

This work was supported by Sinhgad Institute of Technology and Science, Narhe, Pune, India in part of Final Year Project Paper Submission. The Views and conclusions contained in this document are those of authors and should not be interpreted as representing official policies, either expressed or implied, of the sponsors.

6. REFRENCES

[1] Xin Jin, Cindy Xide Lin, Jiebo Luo, Jiawei Han A Data Mining based Spam Detection System for Social Media Networks. PVLDB , 2011

[2] Hongyu Gao, Yan Chen, Kathy Lee, Diana Palsetia, Alok Choudhary Towards Online Spam Filtering in Social Networks. In Proceedings of the 19th Annual *Network* & Distributed System Security Symposium, February 2012.

[3] Kyumin Lee, James Caverlee, Steve webb Uncovering social spammers*:* social honeypots + machine learning, Published by ACM, 2010

[4] THOMAS, K., GRIER, C., MA, J., PAXSON, V., AND SONG, D. Design and Evaluation of a Real-Time URL

Spam Filtering Service. In Proceedings of the IEEE Symposium on Security and Privacy, May 2011.

[5] F. Benevenuto, T. Rodrigues, F. Benevenuto, T. Rodrigues, V. Almeida, J. M.Almeida, C. Zhang, and K.W. Ross. Identifying video spammers in online social networks. In AIR Web, pages 45-52, 2008.

[6] Users of social networking websites face malware and phishing attacks. Symantec.com Blog.

[7] B. Byun, C.-H. Lee, S. Webb, and C. Pu. A discriminative classifier learning approach to image modeling and spam image identification. In CEAS, 2007.