

Research Article

Filtration of Unwanted Text Messages From Online Social Networks Using Probabilistic Classifier

Garima Singh¹, Prof. Amit Yerpude², Prof. Toran Verma³¹⁻³ Dept. of Computer Science and Engineering, Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India.***Corresponding author**

Garima Singh

Email: garima93singh@gmail.com

Abstract: Now-a-days, Online Social Networking sites are important part of human society. In other way a person maintains a social networking site account which is then used for the following purposes like building relations, for business, exchange of data etc. The use of OSN's is therefore tremendously increasing day by day and its demanding for concentration over protection of private information of user's, their likes and dislikes etc. There are different filtering criteria's provided by various OSN's but the role of user to prevent his/her privacy is very less with the help of those criteria's. This paper represents a model that helps to allow the user to take over the control of their profile's and walls. The key parameters of the proposal illustrate are that the message posted on user wall's will be filtered for unwanted content in any form and type and will be posted with the users consent only. The feasibility of this model is presented considering OSN's scenarios.

Keywords: Filtering System, Online Social Networks, Information Filtering, Clustering

INTRODUCTION

The internet is filled with almost every types and kind of information. To achieve or say to use and gain knowledge from this data there are equal numbers of brains i.e. users exist on internet. The same users can share, disseminate as well as communicate multiple variety of information among them. The information is in different types say texts, audio, video, images etc. The medium for this type of data exchange can be in the form of mails, messages and social networks on the internet.

Online Social Networks: OSN's is an internet service that helps the user to increase social networks and relations with other users for sharing of backgrounds, interests, discover real life connections and participate in multiple activities. OSN services basically are web based group centered services in which users can share almost every type of information.

According to the statistics there are 1.28 billion users of Facebook and on an average more than 30 billion pieces of content are shared each month [1]. Similar is the case with Twitter comprising of 255 million users

and sending of 500 million tweets every day [8]. The same trend is followed by other OSN's like Instagram, Google plus, Linked In, etc. this type of huge and with dynamic characteristics the data that concludes for a strategy with the help of web content mining to establish and extract useful information from the data.

Therefore in a way OSN's has transformed the way of interaction and exchange of information among users over the internet. The aim at which OSN's lack its concern is filtering the Information. There is less support for prevention of unwanted messages on users profile and walls. A small support system is arranged by some of the OSN's.

For an illustrate consider Facebook but it also allows filtering up to a certain level i.e. friends, this may happen sometimes that friends of friends or some groups may posts message on user's profile which is also termed as a wall. But the variety of types and contents of message is not explained and is not taken into consideration by support system. User who is permitted to post a message can post any type of content in the message regardless of the category

associated with the message. The message content may be vulgar, political or personal and thus the data contained by the message no longer left private on public platform.

Filtered Wall Architecture

The architecture of OSN services is a three-tier structure. It mainly consists of three layers

1. Social Network Manager (SNM)
2. Social Network Application (SNA)
3. Graphical User Interface (GUI)

Social Network Manager:

It persists the basic OSN related functionalities i.e., the user profiles and relationship management and also provide the information to the social network application layer.

Social Network Application:

To provide external social network application, the layer applies filtering rules (FR) and blacklists (BL). SNA layer composed of short text classifier and Content Base Message Filtering (CBMF). It is very important layer because the classifier distinguishes each message according to its content and CBMF filters the message according to filtering criteria and blacklist provided by the user. CBMF is used to select data item based on the correlation between the content of item and the user preferences.

Graphical User Interface:

It provides the interaction between the users and the system. Which user gives her input and is able to see published wall messages on social media. Moreover GUI also provides user the facility to apply filtering rules for his/her wall messages and helps to give list of BL user who is temporally prevented to publish messages on user's wall. The GUI also include of Filtered Wall (FW) where the user is able to see her desirable messages.

- After entering one of the user to the private wall, the user tries to post a message which has been captured by Filtered wall.
- A ML-based text classifier obtain data from the message content.
- Filtered wall uses data given by the classifier, along with data obtained from the user's profiles, to implement the filtering rules and blacklists techniques.
- Considering the result of the previous step, the result message will be filtered.

LITERATURE SURVEY

Priyanka Salunkhe [1], focused on today various social networking sites are available which make people remain in constant touch with each other. Sharing any type of data has become easy. There are

great advantages of such social networking sites excepting a few minor drawbacks like poor security which create huge problems to people when they were active on such sites. As we have seen Facebook allows users to post comment on another user's wall even when they were unknown to each other. But if that comment is a vulgar one then it may cause serious problem to user reputation. To avoid such a problem Information filtering is used to filter the content of the message.

R.Yasotha [2], focused on a new approach to automatically categorize text documents. Instead of the common practice of string matching approach used for automatic text document categorization this paper proposes a Latent Dirichlet allocation (LDA) based approach. Here the clusters identified by LDA are labeled based on the underlying natural clusters on the domain in concern. Proposed model was able to categorize unseen documents with an accuracy of 66.66%. The lack of accuracy is due to the limitation of classification on one level. The proposed approach can be extended to further lower levels and would be able to achieve greater accuracy.

Sunil Yadav [3], the goal of this paper was to provide an initial base line to build a filtered discussion forum on the most critical issues related to the online social networks. The development of the sample learning websites and blacklisted words table made the filtering easier. As blogs comment and feedback comments are concerned, our proposed work has worked successfully.

Swapnali V. Jadhav [4], have discussed the literature survey of the filtering system. We are developing a system to filter undesired messages from OSN walls. The wall that restricts the unwanted message called as the Filtered Wall (FW). In this report we discussed the idea about the system. Additionally, we studied strategies and techniques limiting the inferences that a user can do on the enforced filtering rules with the aim of bypassing the filtering system, such as for instance randomly notifying a message that should instead be blocked.

Marco Vanetti [5], have presented a system to filter undesired messages from OSN walls. The system exploits a ML soft classifier to enforce customizable content-dependent FR's. Moreover, the flexibility of the system in terms of filtering options is enhanced through the management of BLs. This work is the first step of a wider project. The early encouraging results we have obtained on the classification procedure prompt us to continue with other work that will aim to improve the quality of classification.

METHODOLOGY

The system architecture work flow is presented in fig 1. We have created a Facebook like website where user can sign-in and posts comments and send friend request to each other. Proposed framework is divided into various modules:

1. Extremely Short Text Classifier
2. Filtering Rules
3. Filtered Wall

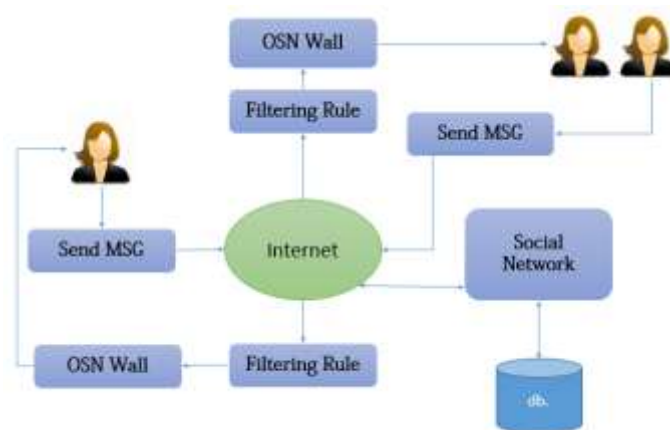


Fig-1: Shows the system work flow

Extremely Short Text Classifier

Filtering of posts and comments are important in any social networking websites. The filtering rule must be powerful to recognize the spammer and non-spammer effectively. We have used Naïve Bayes algorithm for short text classification into spam category using probability distribution. The category are presented below:

- i. Vulgar Words
- ii. Violence Words
- iii. Offensive Words
- iv. Hate Words
- v. Sexual Words

Our framework effectively classifies the comments and posts in the above categories.

Filtering Rules

It is the collection of words which are collected from online in various categories mention above. The framework looks into the database for these words and apply probability distribution algorithm.

Filtered Wall

Filtered call contains the comments and posts filtered by applying probability distribution method. It contains text in only two cases:

- Case: 01: The posts do not contain any spam messages.
- Case 02: The user has approved the spam messages by accepting that it is not a spam.

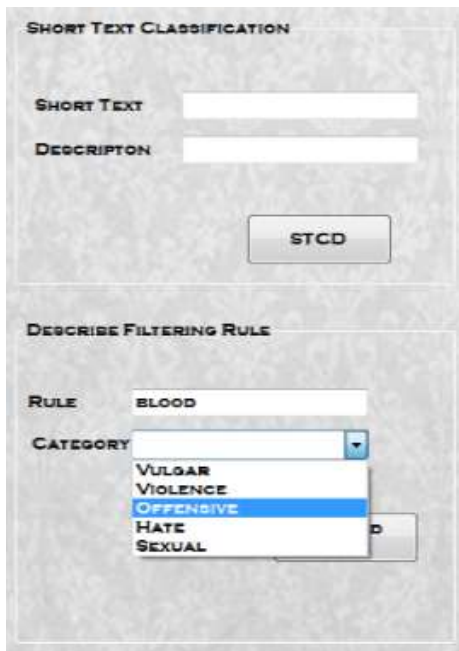


Fig-2: Snapshot of storage of filtering rules

RESULT

We have utilized Visual Studio for modeling our experiment. The replica of Facebook alike website is designed and algorithm is executed.

Firstly, the user A sends message to the B. Before it appears on the user B walls, the proposed framework firsts check with algorithm that the message is spam or not. If it contains some of the spam messages

than it will be blocked otherwise it will be displayed onto the wall.

Secondly, when user B logs in, the use can see the blocked messages from the blocked section. From there, the user can accept the post or can also deny the posts to be posted on wall. Fig. 2 shows the snapshot of detected spam probability.

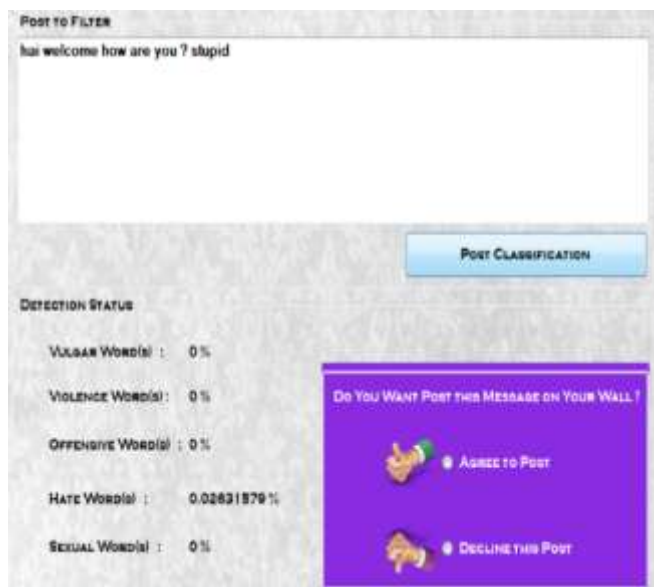


Fig-2: shows the snapshot of filtration process

CONCLUSION

This paper presents approach for filtration of the online social network posts and comments. Our

framework effectively utilizes the probability distribution algorithm and filtering rules to detect spammers or to prevent user wall from being abused.

Our proposed mechanism uses probabilistic model for calculation of term frequency. It is better than that previously used TF based methods.

REFERENCES

1. Salunkhe P, Bharne S, Padiya P. Filtering unwanted messages from OSN walls. In *Innovation and Challenges in Cyber Security (ICICCS-INBUSH)*, 2016 International Conference on 2016 Feb 3 (pp. 261-264). IEEE.
2. Yasotha R, Charles EY. Automated text document categorization. In *Intelligent Computing and Information Systems (ICICIS)*, 2015 IEEE Seventh International Conference on 2015 Dec 12 (pp. 522-528). IEEE.
3. Yadav S, Das S, Rudrapal D. A system to filter unsolicited texts from social learning networks. In *Computing, Communications and Networking Technologies (ICCCNT)*, 2013 Fourth International Conference on 2013 Jul 4 (pp. 1-5). IEEE.
4. Vanetti M, Binaghi E, Ferrari E, Carminati B, Carullo M. A system to filter unwanted messages from OSN user walls. *IEEE Transactions on Knowledge and data Engineering*. 2013 Feb;25(2):285-97.
5. Jagtap B, Rathod R, Vitthal Anpat U, Jajoo MA. A System to Filter Unwanted Messages from OSN User Walls. *International Journal of Advance Foundation and Research in Computer (IJAFRC)*.;2.