Scientific Journal Impact Factor (SJIF) : 1.711
International Journal of Modern Trends in Engineering and Research
ISSN (ONLINE) : 2349 - 9745
ISSN (PRINT) : 2393 - 8161

A Survey Filtering Unwanted Post from OSN User Wall using Automatic Blacklist Generation

Rijavan A. Shaikh [1], Ms. Rachana Kamble[2]
[1]*M.Tech. (C.T.A.), TIT Bhopal*
[2]*Department of CSE, TIT Bhopal*

Abstract — A human is social animal and social Networking plays an important role in everyday life. The best entertainment for todays human is Social Networking and becomes an important part of many peoples life today. In recent years use of online social networks increased rapidly. A User communicates with other by sharing several types of contents like text, image, smiley etc. So Online Social Networks (OSN) should be secured to protect every user's privacy. Main problem of these Online Social Network service is the deficiency of privacy for the user's own message area. Today OSN provide less support to avoid unwanted messages on private message area. For example, Facebook allows users to control who is allowed to put messages in their walls but no message based preferences are supported and therefore it is not possible to prevent undesired messages, such as vulgar and offensive messages. To overcome the limited security measures provided by OSN to filter the unwanted messages, in proposed paper an enhanced filter using machine learning technique and natural language processing based on a content filtering. Two levels of classifications are performed. Messages are categorized as normal and unwanted in first level. In Second level unwanted messages are again classified as per their category. Also blacklisting of messages and user is implemented. To avoid future posts, the user who posts the unwanted messages will be kept in black list until user removes the blacklisted user.

Keywords— Online Social Networking, Text Filtering, Text Blacklisting

I. INTRODUCTION

Social media to have a big an influence on today's people's lives. Social websites like Facebook, Google+ and Twitter etc. have millions of users and they are from all over the world. Social sites have become an important portion of the digital world. It has changed the way people live and communicate with each other. Online Social Networks (OSNs) is mainly used as sharing and communication medium of human life data. The main use of OSN is to share different types of content, including text, links, pictures, audio and video data. Online Social Network is a place where peoples create new social relationships among people with whom they share their interest, text, picture and real time conversation. A social network service manages profiles of each user, his interest, social links and wide range of additional services like finding new peoples with same interest, hobbies and location. OSN is a web based service which allows individual users to create their profile, to create a list of users as friend, family or colleague with whom to connect and whom to allow connection within system. The well-known social networking websites which are mainly used to connect with peoples are: Facebook, BlogSpot, Google+, Twitter, YouTube widely used worldwide [9].

Web content Mining is about searching the useful and related information from a huge amount of available Data. Information filtering using web content mining can be used for a many other purposes as per the requirements in OSN. This is because there is possibility of posting a message or commenting on other post on user's message area called *Wall*. Unwanted messages are filtered using information filtering which is mainly used to give user the ability to control the message written on their own walls [10].

II. FILTERED WALL ARCHITECTURE

The figure 1 shows modified filtered wall architecture used in current system.

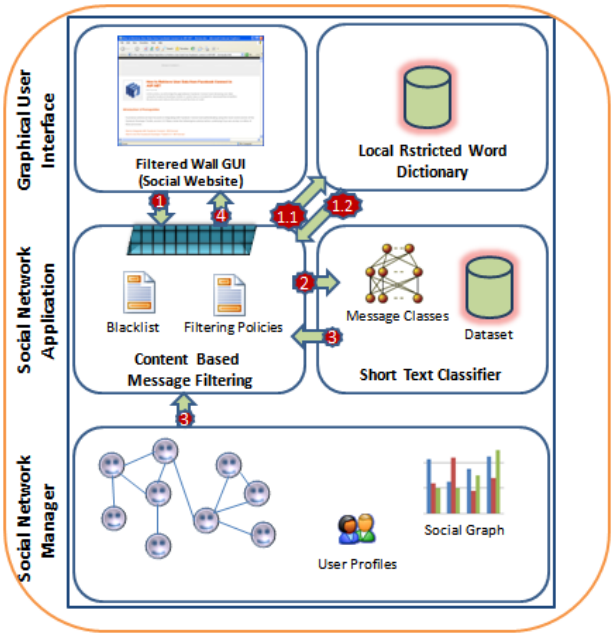

Fig 1 Filtered Wall Architecture

A. Social Network Manager (SNM)

Social Network Manager [1] is first level that provides the important OSN functionalities (i.e., profile and relationship). It also manages data related to the user profile and user relationships. All user's data will be provided to second layer for applying Filtering Rules and Black lists (BL).

B. Social Network Application (SNA)

Content Based Message Filtering (CMBF) and Short Text Classifier are composed in second layer. This level plays main role in message categorization. Also Black list is maintained for bad words and the user who frequently sends bad words.

C. Graphical User Interface (GUI)

Third layer is a Graphical User Interface application for the user who wants to post his messages as an input. Another main function of this layer is to filter the unwanted messages using Filtering Rules (FR) and user who post the unwanted messages will be kept in Black List until user removes the blacklisted user. The GUI also consists of Filtered Wall (FW) where the user is able to post and see his desirable messages [1].

Fig. 1 points can be summarized as follows:
1. The FW captures a message that the user and his/her friends tries to post.
2. A Machine Learning based text classification technique tracks metadata from the content of the posted message.
3. FW uses this metadata extracted by the classifier, parallel with data extracted from the social graph, LRWD and users' profiles, in intend to implement the filtering Rules and Black List.
4. The result produced by step 1 through 3, the message will be displayed or blocked by FW [1].

III. RELATED WORK

In the related work, the recent techniques for the content-based filtering used in Online Social Networking (OSN) are deliberated.

Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno [1], has discussed that OSN's user have a direct control over their own filtered wall to avoid the unwanted messages. In this paper is, a system is designed that gives users a better control over their posted and received messages. To achieve this automated system called FW is used, which have a capacity to detect and filter unwanted messages as per user's requirement. The unwanted messages sent by the other users an blocked by system. In this paper only the content posted by the user will block; but user who is posting that message will not be blocked. The system achieves this all by using content based message filtering technique and short text classification.

L. Roy and R. J. Mooney [2] use mutual filtering method. It explains the content based book proposal system that develops information pulling out and machine learning technique for text categorization. In our proposed system content based filtering is used.

B. Carminative, M.vanetti, E.ferrari, M.Craullo [3]. In this paper using statistical information, the system can generally take decision about the message which is blocked.

Bodicev and M.Sokolova [4] classification of text is done in very complicated and specific terminology. The proposed solution needs the application of learning process. For confining the text characteristic, the text is compressed by using Fractional Matching method and then a language model is developed. The output of fractional matching compression provides consistent care of text classification.

M.Carullo, E.Binaghi, and I. Gallo [5] suggest clustering of document is helpful in many areas. Two categories of clustering general purpose and text tilting are discussed; these both will be used for clustering of information. Novel heuristic online document clustering is predictable, which is proficient in clustering of text tilting parallel measures. Presentation measure is done in F-measure, and then it will be counterpart up with other methods. The result will indicate the power of proposed system.

K.Babu, P.Charles [6] exploit Machine learning text categorization techniques to automatically assigns each short text messages a set of categories depend on its content. The authors worked mainly for preparing robust short text classifier (STC). Still the system does not deal with any kind of preprocessing of posted message to extract metadata. The metadata from posted content helps in learning more accurately. The proposed system takes this into consideration.

IV. EXISTING SYSTEM

Today most of the OSNs provide very little support to block unwanted messages on user walls. For example, in Facebook a user can set who is allowed to post messages in their wall, but there is no provision for content- based preferences and therefore it is not possible to prevent unwanted messages, such as hate, violence and vulgar ones independent of the user who posts them.

A. Need of System

1. Some people will use the indecent and vulgar words in commenting on the public posts. Even though the Social Networks have the restrictions on the users who can post, view or comment on messaged at any user's wall, but they do not have any restrictions on content of post.
2. Providing this service is not only a matter of using previously defined web content mining techniques for a different application, rather it requires to design ad hoc classification strategies.

B. Advantages of Proposed System

1. Proposed system is able to detect some words of same meaning i.e. synonyms.
2. No natural language processing is done prior to add category in database of restricted words.
3. System automatically generates blacklist for restricted words.

V. PROPOSED SYSTEM

In this paper, Blacklist mechanism is used, where the user's list will be avoided for the moment to post on user wall. This paper is the extension of previous paper, all classification and filtering rules will be included, additionally BL rule is used. Based on the user wall and relationship, the owner of the wall can block the user. This prohibition can be approved for an uncertain period of time.
The technique which is used in previous paper will be explained shortly, they are:
A .Short text classifiers
B. Filteration
C. User Black Listing

A. Short Text Classification:

The technique for classification of text which contains large amount of data is available and easy. but it creates problem when the amount of document is short. Due to this problem, short text classification is used. The main aim of the short text classifier is to recognize and separate the normal post and categorize the unwanted post in step by step, not in single step.
Step I: In first level, post is classified as normal and unwanted type of post.
Step II: The second level act as a unwanted post categorizer and decides to which class given post belongs. This class information will be used as part of filtering process. Short text classifier includes machine learning ML based classification.

B. Filtration

In OSNs like in everyday life, the same message may have different meanings and relevance based on who writes it. As a consequence, FRs should allow users to state constraints on message creators. Creators on which a FR applies can be selected on the basis of several different criteria; one of the most relevant is by imposing conditions on their profile's attributes. In such a way it is, for instance, possible to define rules applying only to young creators or to creators with a given religious/political view. Given the social network scenario, creators may also be identified by exploiting information on their social graph. This implies to state conditions on type, depth and trust values of the relationship(s) creators should be involved in order to apply them the specified rules. Fig.5.3. shows the filtering process.

C. User Blacklisting Process

A further component of the system is a Blacklist (BL) mechanism to avoid messages from undesired creators. BL is directly managed by the system, which should be able to determine who are the users to be inserted in the BL and for how much time and decide when user's should be removed from BL is finished. To enhance flexibility, such information is given to the system through a set of rules, hereafter called BL rules. Such rules are not defined by the Social Network Management, therefore they are not meant as general high level directives to be applied to the whole community. Rather, the prosed system decide to let the users themselves, i.e., the wall's owners to specify BL rules regulating who has to be banned from their walls and for how long. Therefore, a user might be banned from a wall, and at the same time, he will not be able to post in the wall.

The added new technique which is used in this paper are:

D. Local Restricted Word Dictionary

Sometimes the message contains the words which are synonyms of unwanted words. These words are remains unfiltered in filtration process. For example, "I will **kill** you" is the unwanted message containing restricted word KILL so it will be blocked. But if the message posted is "You will be **dead**" then it will not be blocked and got posted. Therefore before starting the filtration process a local dictionary containing every possible combination of restricted words is prepared. The dictionary will also contain opposite words of restricted words. The dictionary will be populated as the user adds a new restricted word.

E. Negation Prefixes

The restricted word (RW) is also detected by identifying negation prefix words in combination with opposite words. The negation prefix include NOT, NON, NO, UN words. If an opposite word of given RW is found with any of the negation prefix, then the message is decided as unwanted message.

VI. METHODOLOGY

In proposed system, a user can register himself as a New User and creates a profile. The user can login in order to start begin his work on OSN. The user can create and manage their own groups. The additional features are like user can find his friends based on preferences set in profile. User can send request to his friend or accepts a request send by other. User can Post or receive messages.

Steps in Preparing Local Restricted Word Dictionary (LRWD):

1. OSN user decides restricted words (RW) by categories.
2. Search synonyms of restricted words (RW) and store them to LRWD.
3. Search opposite of restricted words (RW) and store them to LRWD.

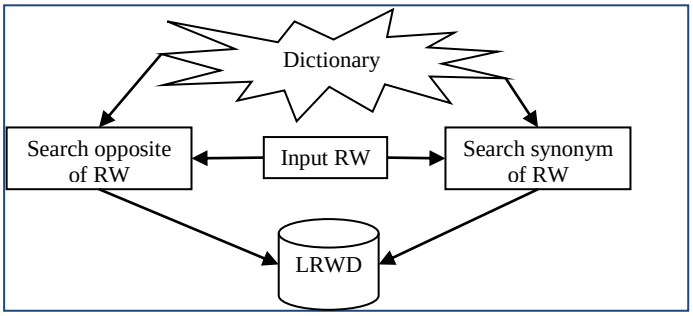

Fig. 2: Generating LRWD

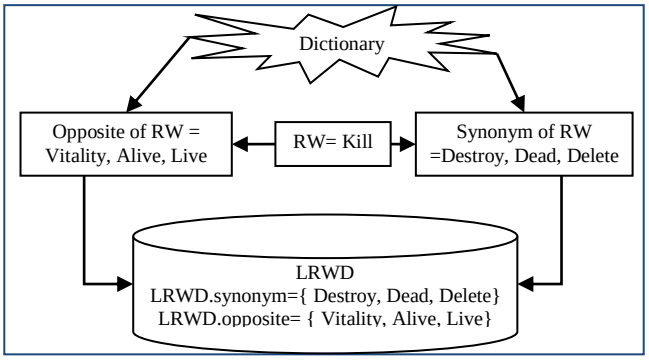

Fig. 3: Example: Generating LRWD for word Kill

A. Filtering Process With NLP:

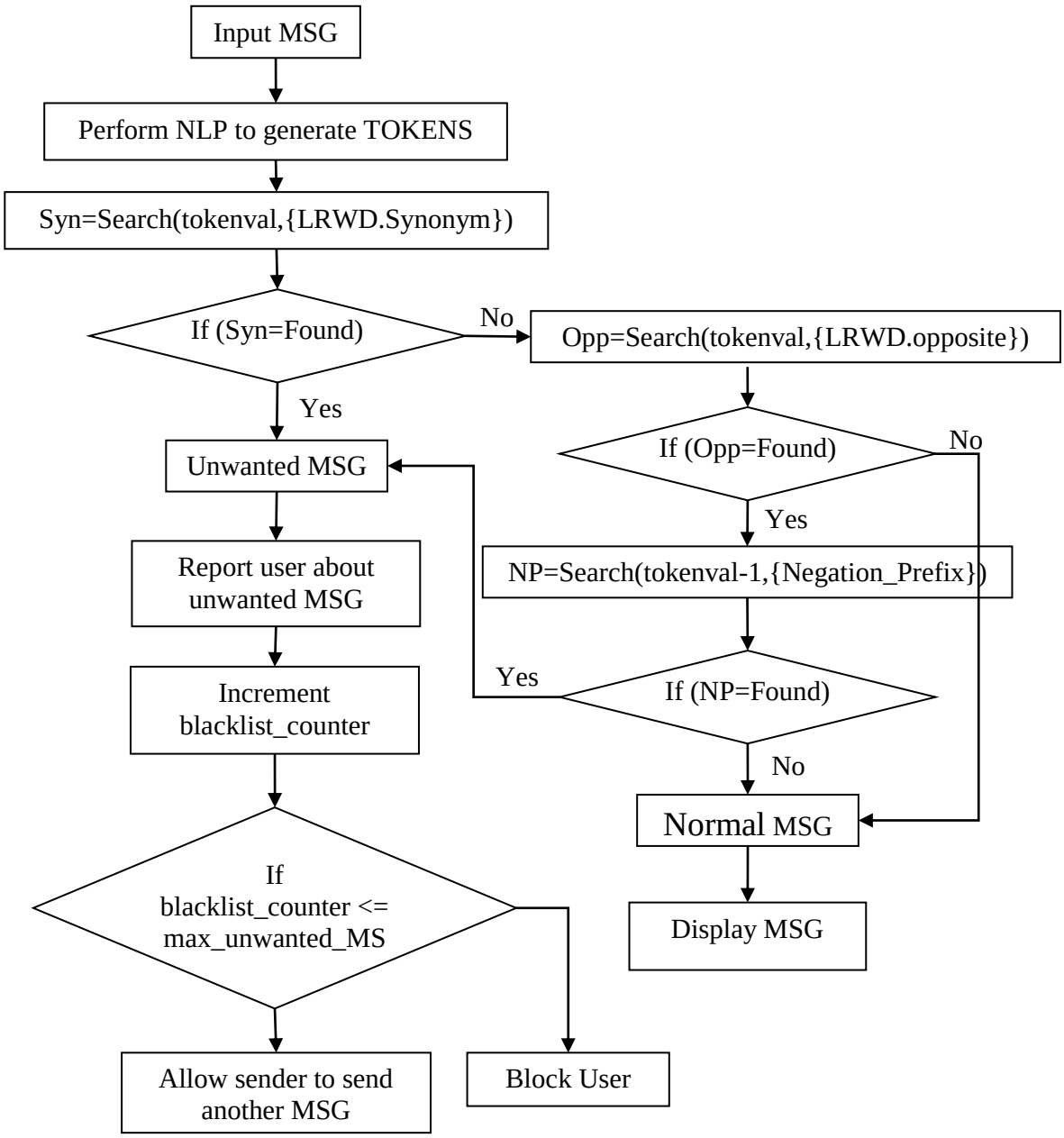

Fig. 4: Message Filtering and Blacklisting Process

CONCLUSION

The survey paper Filtering Unwanted Post from OSN User Wall using Automatic Blacklist Generation presents a filter on OSN that can minimize the unwanted messages posted on OSN. The first step of is to classify the post as normal and unwanted using several techniques. Finally Blacklist of words and user is also implemented so that owner of the user can insert the user who post unwanted message posts. Also a new database of blacklisted words with categories is prepared that can used for future classification work using machine learning.

REFERENCES

[1] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, an Moreno Carullo, " *A System to Filter*

Unwanted Messages from OSN User Walls",2013.

[2] R.J. Mooney and L. Roy, *"Content-Based Book Recommending Using Learning for Text Categorization,"* Proc. Fifth ACM Conf.Digital Libraries, pp. 195-204, 2000.

[3] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari,*"Content-Based Filtering in On-Line Social Networks,"* Proc. ECML/PKDD Workshop Privacy and Security Issues in Data Mining and Machine Learning (PSDML '10), 2010.

[4] V. Bobicev and M. Sokolova, *"An Effective and Robust Method for Short Text Classification,"* Proc. 23rd Nat'l Conf. Artificial Intelligence (AAAI), D. Fox and C.P. Gomes, eds., pp. 1444-1445, 2008.

[5] M. Carullo, E. Binaghi, and I. Gallo, *"An Online Document Clustering Technique for Short Web Contents,"* Pattern Recognition Letters, vol. 30, pp. 870-876, July 2009.

[6] K. Babu, P. Charles, *"Machine Learning Text Categorization Techniques",* International Journal of Computer Science and Information Technologies, Vol. 5 (2) , 2014, 1748-1753

[7] N.J. Belkin and W.B. Croft, *"Information Filtering and Information Retrieval: Two Sides of the Same Coin?"* Comm. ACM, vol. 35, no. 12, pp. 29-38, 1992.

[8] M. Chau and H. Chen, *"A Machine Learning Approach to Web Page Filtering Using Content and Structure Analysis,"* Decision Support Systems, vol. 44, no. 2, pp. 482-494, 2008.

[9] S. Venkata Lakshmi, K. Hema, *"Filtering Information for Short Text Using OSN",* International Journal of Advanced Research in Computer Science & Technology, pp. 317-319, 2014.

[10] Kanika Sharma, Manavjeet Kaur, *"A Review on Unwanted Message Filtering System from OSNs User Wall"*, International Journal of Advances in Science and Technology, pp. 84-86, 2014.

