|| Volume 2 || Issue 2 || SEPTEMBER 2017 || ISSN (Online) 2456-0774 ||
INTERNATIONAL JOURNAL OF ADVANCE SCIENTIFIC RESEARCH
AND ENGINEERING TRENDS

Spam Blocker for OSN

Ms. Seema Pralhad Jaybhaye[1], Mr. K. Vishal Reddy[2]
Computer Engineering Department, Deogiri Institute of Engineering and Management Studies, Aurangabad, Maharashtra[1]
Assistant Professor, Computer Engineering Department, Deogiri Institute of Engineering & Management Studies, Aurangabad[2]
seemamail08@gmail.com[1], vishalreddy@dietms.org[2]

Abstract – Online Social Network (OSN) which plays a most important role in our day to day life, one user can communicate with one or more users by sharing several types of information. Major issue in Online Social Network is to prevent user for posting unwanted messages such as vulgar, political words etc. There is need to give the users an ability to control the messages posted on their own private/public space to avoid that unwanted contents to be displayed. In today's available OSN System unwanted post will be directly posted on the users wall; to fill this gap, in this paper, I propose a system that automatically filters the undesirable messages by allowing Online Social Network users to have a direct control on the content posted on their walls. This is done through a filtering criteria to be applied to their walls, and a Machine Learning technique based soft classifier algorithm such as Radial Basis Function Network (RBFN).To do this, Black List (BL) mechanism is proposed in my system, which avoid undesired creators messages. BL is used to determine which user should be inserted in Black List and decide when the retention of the user is finished. I have used DICOMFW as a special Facebook application due to which user can report a spam.

Keywords – *Online Social Network (OSN), Machine Learning Techniques (MLT), Black List (BL), Radial Basis Function Network (RBFN), Content-Based Messages Filtering (CBMF), Short Text Classifier (STC)*

I INTRODUCTION

Online Social Networks (OSN) is one of the most popular medium for communication, sharing and broadcasting the human life information. Now a day's Online Social Networking such as Facebook, Whatsapp etc become most popular interactive medium for communication. The advantageous feature of social networking site is the ability to create and share personal information. This information page may contain a photo, and some basic personal information such as name, age, sex, location. Most of the social networking sites on the Internet also let you post videos, photos, and personal blogs on your profile page. One of the most important features of online social networks is to find and make friends with other site members.bt some time it become problematic to avoid someone and its messages or unwanted post.

Some time a huge amount of unwanted data may be posted on user's wall. Facebook allows users to state who is allowed to insert messages in their walls (i.e. friends, friends of friends, or defined groups of friends). However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them.

In previous work the wall owner does not have the control of their own private area and this shows there is no content based preferences. Therefore it is not possible to prevent unwanted messages such as political or vulgar once. To fill this gap, I exploit new system as the spam blocker for OSN. The aim of the present work is therefore to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls. I exploit Machine Learning (ML) text categorization techniques [4] to automatically assign with each short text message a set of categories based on its content. , Black List (BL) mechanism is proposed in my system, which avoid undesired creators messages. BL is used to determine which user should be inserted in Black List and decide when the retention of the user is finished. Machine Learning Text Categorization is also used to categorize the short text messages.

The major efforts in building a robust short text classifier (STC) are concentrated in the extraction and selection of a set of characterizing and discriminate features. The system provides a powerful rule layer exploiting a flexible language to specify Filtering Rules (FRs), by which users can state what contents should not be displayed on their walls insert the neural model within a hierarchical two-level classification strategy. By using Filtering Rules OSN users have the ability to control the messages posted on their own private space to avoid unwanted content to be displayed. Filtering Rules also exploit user profiles, user relationships.

II RELATED WORK

Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno Carullo [1] provide the user to have a straight rule over their own wall to avoid the unwanted messages. Aim of this paper is, user have a direct control over messages posted on their own wall. So automated system called Filtered wall (FW), which have a capacity to filter unwanted message. This system will block the undesired message send by the user. Drawback of this paper is the user will not be blocked; only the message posted by the user will blocked. Content based message filtering and short text classifier support this system .To overcome the problem of this

paper, Blacklist rule will be implemented as future enhancement.

L. Roy and R. J. Mooney use Collaborative filtering method, but in the proposed system content based recommendation are used. It explains a content based book recommending system that develops information extraction and machine learning algorithm for text categorization. B. Carminati, M. Vanetti, E. ferrari, M. Carullo, and E. Binaghi Quality of classification is considered as the main aim. This system can usually take decision about the messages which is blocked, due to the tolerance depends on statistical data. F. Sebastian Efficiency is good, labor power will be saved is the advantage of this paper. The main approach used here is text categorization. Comparison will be performed between human expert and labor power expert. H. Schutze, D. A. Hull, and J. O. Pedersen latent semantics B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu and M. Demirbas**.** In micro-blogging services such as Twitter, the users may get overwhelmed by the raw data. One solution to this problem is the classification of Twitter messages (tweets) [27].

S. Pollock said the major issue in social network is user does not have a control over their walls because it does not support content based preferences .Therefore it is not possible to prevent undesired messages. Most of the proposal is mainly focus on providing a classification mechanism to avoid useless data [10].

III EXISTING SYSTEM

Today's OSN provide very little support to prevent unwanted messages on user wall such as facebook, Twiter allows user to state who is allowed to insert messages in their walls Content-based preferences are not supported and therefore it's not possible to prevent undesired messages such as political, vulgar. Policy-Based Personalization of Online Social Network (OSN) Contents such systems do not provide a filtering policy layer by which the user can exploit the result of the classification process to decide how and to which extent filtering out unwanted information. In OSNs, information filtering can also be used for a different, more sensitive, purpose. This is due to the fact that in OSNs there is the possibility of posting or commenting other posts on particular public/private areas, called in general walls.

Information filtering can therefore be used to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted messages. We believe that this is a key OSN service that has not been provided so far.

However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them. Providing this service is not only a matter of using previously.

Limitations:

1) Traditional filtering method such as Collaborative filtering is used.
2) Content-Based preferences are not supported.
3) Users have no control on their private wall.
4) Users are not allowed to add and define word as spam for its private wall.
5) BL is not possible.
6) There is no guaranty of complete message filtering.

IV PROPOSED SYSTEM

The aim of the present work is therefore to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls. We exploit Machine Learning (ML) text categorization techniques [1] to automatically assign with each short text message a set of categories based on its content. The major efforts in building a robust short text classifier are concentrated in the extraction and selection of a set of characterizing and discriminate features. The solutions investigated in this paper are an extension of those adopted in a previous work by us [1] from which we inherit the learning model and the elicitation procedure for generating pre-classified data.

As far as the learning model is concerned, we confirm in the current paper the use of neural learning which is today recognized as one of the most efficient solutions in text classification. In particular, we base the overall short text classification strategy on Radial Basis Function Networks (RBFN) for their proven capabilities in acting as soft classifiers.

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.

Advantages of proposed system

1) The proposed system is totally automated system.
2) By using this system User can have control on their private wall.
3) In this system Content-based preferences are supported.
4) Machine learning hierarchical classification method is used.
5) In proposed system users are allow to add and define word as spam for its private wall.
6) Process of detecting and filtering spam is transparent.
7) This system guarantees 100% filtering of messages.

V SYSTEM DESIGNING

Architecture of the proposed system includes filtering rules, classification and blacklist. The whole process will be visible clearly in Architecture.

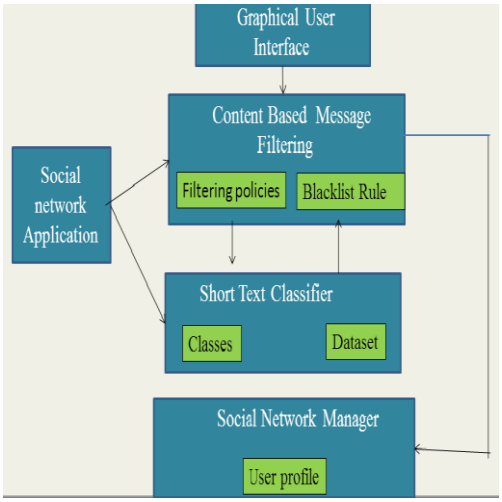

Figure 1 System Architecture

The architecture in support of OSN services is a three-tier structure:

1) The first layer, called Social Network Manager (SNM), commonly aims to provide the basic OSN functionalities (i.e. profile and relationship management).

2) The second layer provides the support for external Social Network Applications (SNA). The supported SNAs may in turn require an additional layer for their needed Graphical User Interfaces (GUIs).

3) The core components of the proposed system are the CBMF and the Short Text Classifier (STC) modules.

4) Black List (BL) can also be used to enhance the filtering process.

5) After entering the private wall of one of his/her contacts, the user tries to post a message, which is intercepted by FW.

6) ML based text classifier extracts metadata from the content of the message.

7) FW uses metadata provided by the classifier together with data extracted from the social graph and users profiles to enforce the filtering and BL rules.

8) Depending on the result of the previous step, the message will be published or filtered by FW.

VI IMPLEMENTATION

1) Short Text Classifier

Other classifier which is used in previous paper is used to classify the text which contain large amount of data, but it endure when the amount of document is little. To overcome this problem, short text classifier is used. Aim of the short text classifier is to recognize and eradicate the neutral sentences and categorize the non neutral sentences in step by step, not in single step. This classifier will be used in hierarchical strategy. The first level task will be classified with neutral and non neutral labels. The second level act as a non-neutral, it will develop gradual membership. These grades will be used as succeeding phases for filtering process. Short text classifier includes text representation, Machine learning based classification.

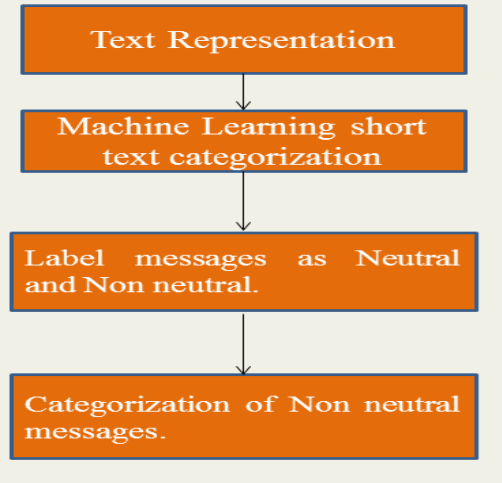

Figure 2 Short Text Classifications

Text Representation:

Representing the text of a document is critical, which will affect the classification performance. Many features are there for representation of text, but we judge three types of features. BOW, Document properties (DP) and contextual features. BOW and Document properties are already used are endogenous that is, text which is entirely derived from the information within the text message. Endogenous knowledge is well applicable in representation of text. It is genuine to use also exogenous knowledge in operational settings. Exogenous knowledge is termed as any source of information from outside the message but directly or indirectly communicate to the message itself. CF modelling is introduced; its feature is to understand the semantics of message. DP features are heuristically evaluated. Some domain specific criteria is considered, trial and error procedures are needed for some cases. Some of them are,

- Correct words: It states the amount of terms. Correct words will be calculated.
- Bad words: comparison to the correct words will evaluate. Collection of dirty words will be determined.
- Capital words: It will say about the amount of words written in message. Percentage of words in capital case will be calculated.

2) Filtering Rules

Definition 1: (Creator specification) - A Creator Specification CreaSpec, which denotes a set of OSN users. Possible combinations are 1) Set of attributes in the An OP Av form, where an is a user profile attribute name, Av is profile attribute value and OP is a comparison 2) Set of relationship of the form (n, Rt, minDepth, maxTrust) indicate OSN users participating with user in a relationship of type Rt, depth greater than or equal to minDepth, trust value greater than or equal to maxTrust.

Definition 2: (Filtering rule) - A filtering rule is a tuple (Auth, CreaSpec, ConSpec, Action) 1) Auth is the user

who states the rule. 2) CreaSpec is the Creator specification. 3) ConSpec is a boolean expression. 4) Action is the action performed by the system. Filtering rules will be applied, when a user profile does not hold value for attributes submitted by a FR. This type of situation will dealt with asking the owner to choose whether to block or notify the messages initiating from the profile which does not match with the wall owners FRs.

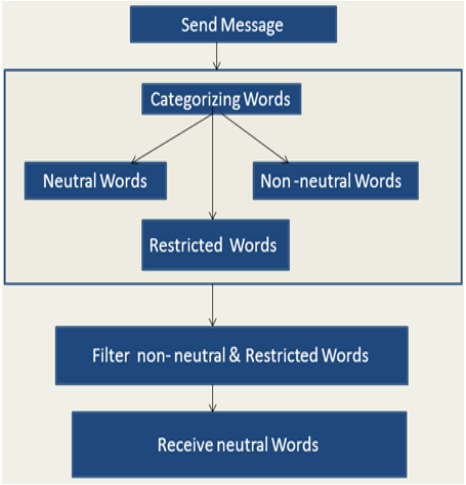

Figure 3 Filtering Policies

3) Blacklists

A further component of our system is a BL mechanism to avoid messages from undesired creators. BLs is directly managed by the system, which should be able to determine who are the users to be inserted in the BL and decide when user's retention in the BL is finished. To enhance flexibility, such information is given to the system through a set of rules, hereafter called BL rules. 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, by, at the same time, being able to post in other walls.

Similar to FRs, our BL rules make the wall owner able to identify users to be blocked according to their profiles as well as their relationships in the OSN.

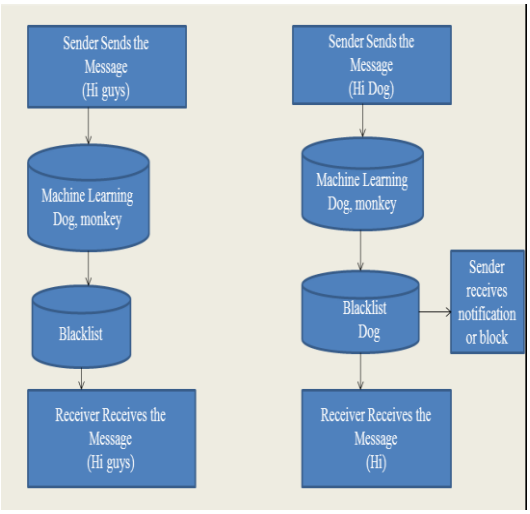

Figure 4 Blacklist Mechanism

4) Machine Learning Based Classification

Short text classifier include hierarchical two level classification process.In a machine learning classification semi supervised learning is used because some text may contain misspelling, symbols etc Short text categorization as a hierarchical two level classification process in which RBFN categories a messages as Neutral and Non- neutral.

- A first level can categories message as neutral and non-neutral.
- In second-level Non neutral messages assigning a class such as vulgar, political, other etc.
- RBFN contain a single hidden layer of processing units. Classification function is non linear, which is the advantage of RBFN.

Algorithm:

Step 1 Start
Step 2 A User tries post the message on another user wall.
Step 3 STC extract data from message content.
Step 4 Machine learning approach used for classification.
Step 5 If (Words = = neutral Words).
Step 6 Message is posted on the wall.
Step 7 Else if (Words = = non neutral Words).
Step 8 Enforce Content Based message filtering and BL Rule.
Step 9 Reject non neutral Words using Blacklist and post the filtered message on the user wall.
Step 10 Stop

VII DICOMFW

DICOMFW is a Facebook application is conceived as a wall and not as a group.

To summarize an application:

1) View the list of users FW.
2) View messages and post a new one on a FW.
3) Define Filtering Rule (FRs).

- A user tries to post a message on a wall.
- If message contain non neutral word then user who post message receives an alerting notification and it is block by FW after a particular attempts.
- User can also report a spam words.

VIII CONCLUSION

In this paper I have presented a system to filter undesired messages from OSN walls. DICOMFW is more advantageous part of my system.I would like to focus that the system proposed in this paper represents just the core set of functionalities needed to provide a sophisticated tool for OSN message filtration and classification.

REFERENCES

[1] Adoma vicious and G. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 2005.

[2] 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.
[3] 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.