Unwanted Text Filter for OSN

Sampada More, Amol Dhanure, Alpana Kulkarni, Prakash B. Nimbalkar
SCOE, Sudumbare, Pune University, Pune,
Maharashtra, India

Abstract— In today's Online Social Networking, the basic issue needed to be handled is filtering of messages and hence avoid unwanted post posted on the private wall or timeline. The OSN today provides a very limited support to this user requirement. To overcome these limitations, in this paper we propose a system "Unwanted Message Filter for OSN User Walls" which will allow the OSN user to have direct control on the messages which are supposed to be posted on his wall. We can achieve this through a adaptable protocol based system that allows the users to customize the criteria of filtering applied on their OSN walls to achieve a good sense of security. We also use a machine learning based on soft classifier which automatically sorts and labels messages so that it can support Content Based Filtering.

Keywords— Blacklist Management, Content Based Filtering, Policy Based Personalization, Short Text Classification

I. INTRODUCTION

Online Social Networks (OSNs) are today very popular medium to share the data and communicate. A large number of people share a considerable amount of human life information through such OSNs. OSN generates a persistent and steady flow of contents which involves text messages, audio data and video data. According to Facebook statistics average user creates 90 pieces of content each month, whereas more than 30 billion pieces of content which include web links, new stories, blogs posts, notes, photo albums, etc are shared each month. Thus, a large source of such data can create the area for the employment of Web Data Mining plans and strategies made for automatic discovery of useful information hidden within the published data.

Information filtering has been explored mainly for text messages and textual web contents with the aim to provide classification techniques for avoiding overwhelming of unimportant data. The same techniques can be used in OSNs for more sensitive purpose where data is posted on private walls or timelines. Information Filtering can here give user the ability to automatically control the text which would be supposed to be posted on their own wall or timelines by giving the random friends of the user the right to approve the text message which is suspected as unwanted by the filtering policies. We believe that we are the first to provide this service for OSNs as some published messages could still unwanted or sometimes even acceptable content can remain unpublished even after undergoing Content Based Filtering. For example, Facebook allows user to choose who is allowed to post on their walls (i.e. no one, friends, a group of friends, friends of friends or everyone) but it does not provide the ability to scan or check the messages before posting. So this may contain political or vulgar information which is not desired to be posted. Providing this technique is not only matter of using previously established techniques but also we require to design and make such system which could provide the approval security which the users wish. This is because wall messages which are constituted and published by traditional classification strategies have serious limitations regarding sensitive subject.

The aim of our project is therefore to propose and evaluate experimentally an automated system called Refined and Secured Wall which will be able to filter unwanted messages from the OSN walls and approve those messages from the user friends. Here we exploit the Machine Learning (ML) categorization technique to divide the short text messages into different categories. Thus, we can concentrate to extract and select the different data for friend's approval.

Besides classification facilities, our system will provide a strong Rules Layer which will use a flexible language to specify Filtering Rules (FRs) for the system according to one's need which will state which content should not be posted on their wall. Also, the system will provide support for Blacklist which can prevent posting of messages from the blocked people. Although we have implemented our system on a self created Social Network still we can apply this system to other OSNs as well.

To the best of our knowledge we are first proposing a service to a system to automatically send request to friends to approve the messages declared as 'unwanted' by the filtering policies including message content and message creator relationship with the user. The major difference includes the features added to previously used classification techniques.

II. RELATED WORK

Our main contribution to this paper is providing a system which will provide enhanced features of providing customized content based message filtering for OSNs based on the ML techniques. We have mentioned in our introduction that, to the best of our knowledge we think, we are the first proposing the enhanced features of "message approval by friends" to the system application of OSNs. However, our work is related with the skills used in Content Based Filtering and Policy Based Personalization techniques used by OSNs. Therefore, we need to survey the literature in both the fields.

III.   CONTENT BASED FILTERING

Each user operates as an individual in the Content Based Filtering. Here, the system selects the information data which is co-related with the user provided content and sends the data for further processing. Content Based Filtering mainly comprises of the use of Machine Learning techniques due to which a classifier is automatically induced by scanning the pre defined examples. A variety of related work is recently appeared which provide exaction features, collection of various samples and machine learning.

Several approaches prove that the Bag of Words (BoW) method gives good performance and is more powerful than sophisticated text representation that have superior quality but lower statistical quality. As long as learning model is considered there is a number of major approaches used in text classification and content based filtering technique. They have some advantages and disadvantages in function of application dependent issues. The analysis has been conducted for confirming superiority of those classifiers.

The application which uses Content Based Filtering on messages posted on OSN walls or timelines has additional challenges. The length of these messages is short and it should be compared with large topics that can be discussed. Therefore, in our project we consider gradual membership to class which is a key feature for defining flexible Policy Based Personalization technique.

IV.   POLICY BASED PERSONALIZATION

We have listened to some techniques that show mechanism for classification which can be used to personalize the access of users in OSNs. This can prevent flooding of users who only access the data by micro-blogging services used for raw data. The system here focuses on the mechanism used in networks such as Twitter and therefore combine a set of categories with each tweet which describes its own content making the user access simpler to find and view the tweets related to his area of interest. However, such system are not capable of providing a filtering policy layer through which the user could find the result to classify the process and cannot how and to what extent one should filter out unwanted information.
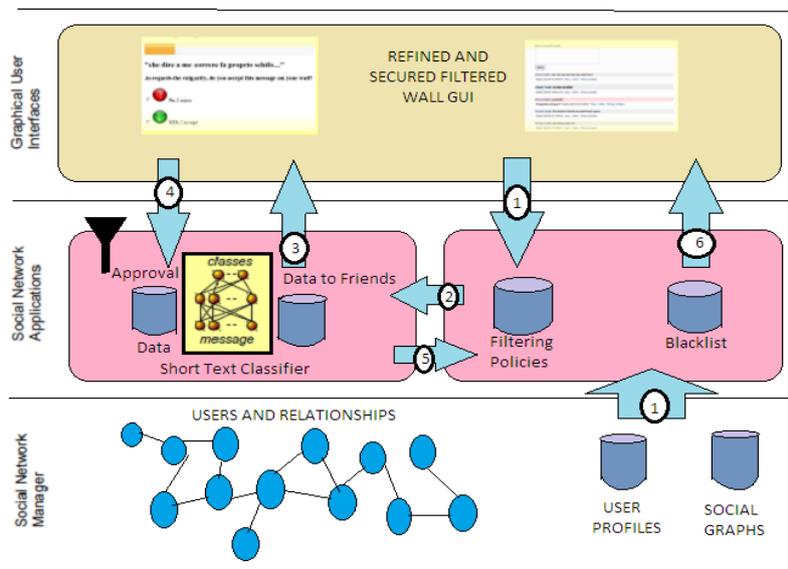

Fig. 1.Refined and Secured Wall Conceptual Architecture and flow messages from writing to publication of the messages on user walls.

V.   REFINED AND SECURED WALL ARCHITECTURE

Refined and Secured Wall Architecture is a three layered architecture design that we use in our proposed system. The first layer is Social Network Manager (SNM) layer. This layer provides basic OSN functions of profile and relationship management. The second layer is a Social Network Application layer. It uses all the filtering techniques to clean the unwanted messages. It includes the Filtering Policies, the Short Text Classifier, the data sending for approval to user friends, the approval of the text from those friends and the blacklist management. The actual process of filtering is carried out here in this layer. The third layer is the Graphical User Interface (GUI) layer which provides the user to interact with each other via the OSN.

With respect to the above mentioned architecture our proposed system is placed mainly in the second and third layers. The user will interact with the help of GUI with the proposed system and setup and will thus be able to extract the purified messages without unwanted matter with the help of filtering rules and filtering policies and Filtering Rules.

The Blacklist management is also done in the SNA layer and it provide user with a firewall or security from the blocked people for publishing of messages on their walls. The core components of the proposed system are Content Based Filtering module, Short Text Classifier module and Approval from Friends module. For graphically depicted Figure 1, the path of message from its writing to publication is summarized as follows:

1) User tries to post the message on his contacts after entering to his private wall which is intercepted by Refined and Secured Wall.

2) The ML based classifier will extract metadata from the content.
3) The data will undergo Blacklist rules and will be send to filtering policies.
4) Suspected data will be sent to three random for approval.
5) The data will be posted on the wall only if it is approved by those three people with the Refined and Secured Wall.

VI. SHORT TEXT CLASSIFIER

When we need to classify the text in large documents such as newswires or large information data wires at that time the established techniques may work well But when it comes to Social Networking, we rarely use such large data sets to share the information or messages on user walls. Here major information transfer or major communication messages are in a short text format. This is when the established Techniques suffers. So to overcome this we define some set of characterizing rules and different features which allow the representation of underlying concepts to collect the consistent and complete set of supervised examples. Our study is aimed to design and evaluate various techniques of representation to combine a neutral strategy of learning to categorize short texts. From Machine learning point of view, we get the task of classifying the sentences into two categories. Those are "Neutral Words" category and "Non- Neutral Words" category. So a sentence will basically undergo two layers of classification. First layer divides the sentence by labeling it with crisp tags of "Neutral" and "Non neutral" words by tagging it with a small labeled crisps and then the second layer works only on the set of 'Non neutral" elements.

For each such elements it selects the random three people in the Friend List of the user profile and send them a notification to check and approve the suspected data .Short text classifier itself does not take any hard decisions about the publishing of data.

VII. BLACKLIST MANAGEMENT AND FILTERING RULES

This section describes the rules adopted for filtering of the messages. We start with describing the Filtering Rules and then illustrate the use of Blacklists. Let us consider a Social Network which corresponds to directed graph, where each node corresponds to a user and edge corresponds to the relationship between two users. So our filtering rules and blacklist management is done in such a way that we could consider the trust value between the relationships and could manage the publishing of messages.

VIII. FILTERING RULES

For defining a Filtering Rules Specification Language we consider three main issues which affect message filtering decisions. Firstly, in OSNs the same sentence can have different meaning depending on who writes it. So first we should put conditions on creators of messages. Secondly, we should differentiate the messages into two categories that are Neutral and Non-neutral elements so that Non neutral elements can be sent to friends for approval or blocking. And lastly, the action the system has to perform after the friend's instruction (i.e. to post or block the message) arrives. So, all of these three factors are considered while defining the filtering rules for the system.

IX. BLACKLIST MANAGEMENT

The Blacklist management is used to avoid and block messages from the unwanted creators or the blocked people. Blacklists are directly managed by the system and users themselves decide who should be inserted in the blacklist and up to what period of time. After that the user's interference in the decision making of publishing of messages is finished and the system itself takes the control of the blocking of the message from the user walls. Similar to Filtering Rules, the Blacklist rules give the ability to user to identify the users to be blocked according to their profiles or some indirect relationships between them or even if there is direct relationship but still a bad opinion about some friend. This banning can be adopted for a undetermined time or a specific time period. Moreover, the banning criteria can be continued further as per User's wish. That means if the person remains in blacklist for a particular specified time then it may deserve to stay there for another while till it is been removed by the user.

X. CONCLUSIONS

In this paper, we have presented a customized message filtering system for OSNs. Our system mainly uses machine learning classifier to customize content dependent filtering rules. The system becomes more flexible through blacklist management rules.Our project is the first step of a wider project. The encouraging result of this project is motivating us to continue with enhancement of other features which will provide users with even more secure and better system. In particular our future plan involves deeper investigation two interdependent tasks. First will be learning stage. Since the underlying domain is changing dynamically, the collection of pre-classified data may have limitations in longer terms.

The second stage will provide user with high discriminative power and even more faster friend's acknowledgement for suspected message. Additionally, we also are planning to enhance our system more sophistically when system itself will decide when user should be inserted in blacklist. Additionally, we also want to study strategies and techniques which limit the interference of user in bypassing of message in filtering system.

REFERENCES
[1]   A. Adomavicius 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. [4] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002.

[5]   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.

[6]   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.

[7]   P.J. Denning, "Electronic Junk," Comm. ACM, vol. 25, no. 3, pp. 163-165, 1982.

[8]   P.W. Foltz and S.T. Dumais, "Personalized Information Delivery: An Analysis of Information Filtering Methods," Comm. ACM, vol. 35, no. 12, pp. 51-60, 1992.

[9]   P.S. Jacobs and L.F. Rau, "Scisor: Extracting Information from On- Line News," Comm. ACM, vol. 33, no. 11, pp. 88-97, 1990.

[10]   S. Pollock, "A Rule-Based Message Filtering System," ACM Trans. Office Information Systems, vol. 6, no. 3, pp. 232-254, 1988.

[11]   P.E. Baclace, "Competitive Agents for Information Filtering," Comm. ACM, vol. 35, no. 12, p. 50, 1992.

[12]   P.J. Hayes, P.M. Andersen, I.B. Nirenburg, and L.M. Schmandt, "Tcs: A Shell for Content-Based Text Categorization," Proc. Sixth IEEE Conf. Artificial Intelligence Applications (CAIA '90), pp. 320-326, 1990.

[13]   G. Amati and F. Crestani, "Probabilistic Learning for Selective Dissemination of Information," Information Processing and Management, vol. 35, no. 5, pp. 633-654, 1999.

[14]   M.J. Pazzani and D. Billsus, "Learning and Revising User Profiles: The Identification of Interesting Web Sites," Machine Learning, vol. 27, no. 3, pp. 313-331, 1997.