A Novel Mechanism of Preventing Unparlimentary Messages in OSN Networks

Mulagapati DurgaDevi [#1], J.Peter Praveen [*2]

Department of Computer Science & Engineering,
Visakha Institute of Engineering & Technology,
57th Division, Narava, Visakhapatnam, AP (INDIA) - 530027.

Abstract

Online Social Networks (OSNs) plays a very important role in today's day to day communication. By using OSN network a lot of billions of users share their recent updates and personal feelings all around the world through public and private access policy pattern. As the OSN Network is increasing its popularity by gaining a lot of user's attention ,the major important issue that was faced by OSN user's is the ability to control the messages posted on their own private space to avoid that unwanted content is displayed. To solve this problem, in this paper, we proposed a new novel filtering system which allows all the participating OSN users to have a direct control on the messages posted on their walls. This new mechanism allows users to post the wall message with a free of abused words. This is achieved by a Machine Learning (ML) based soft classifier algorithm which is used for automatically labeling messages if it is recognized as black list word.

Keywords

Machine Learning, Soft Classifier, OSN's, Black Lists, and Abused Words.

1. Introduction

Online Social Networks (OSNs) [1] plays a very important role in today's day to day communication. By using OSN network a lot of billions of users share their recent updates and personal feelings all around the world through public and private access policy pattern. As the OSN Network is increasing its popularity by gaining a lot of user's attention, the major important issue that was faced by OSN user's is the ability to control the messages posted on their own private space to avoid that unwanted content is displayed. With the increase in user registrations in OSN Networks a lot of data like text wishes,audio's,video's and picture messages is been communicating through various OSN users daily through web[2]. For example as per the recent facebook statistics [4] which was conducted by a team of reviewers specially to obtain the traffic of facebook usage count, we got an report like on an average each user in any OSN network creates 150 pieces of content every month, whereas more than 65 billion pieces of content are shared each and every month. The huge and dynamic character of these data creates the premise for the employment of web content mining [3] strategies aimed to automatically discover useful information dormant within the data.

Many researches have been keep on conducting several research work in order to avoid the unwanted content to be published in OSN networks. For this reason they have been extended the current OSN network with a new facility like information filtering mechanism. This new filtering mechanism classifies the original message into a group of words in order to avoid the unwanted messages to be passed through the network. In OSNs, information filtering can also be used for

different instances I.e. Sensitive information sharing over private walls. 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.

We consider information filtering mechanism as the new mechanism which has the ability to automatically control the messages written on their individual private/public walls, by pointing out unwanted messages. We solely believe that this is one of the major key OSN services that have not been provided so far. Although face book networks are capable of providing some sort of security for the stakeholders who participate in OSN, it failed in providing complete security in all levels. However, there was no content-based preferences available in the current OSN networks; therefore it is not possible to prevent undesired messages, such as political messages or vulgar ones, no matter of the user who posts them. To Provide this service we not only use previously defined web content mining techniques for a different application, rather it also requires to design ad hoc classification strategies. This is because wall messages are constituted by short text for which traditional classification methods have serious limitations since short texts do not provide sufficient word occurrences.

The main aim of our present work is therefore to propose and experimentally evaluate an automated message filtering system, called as Filtered Wall (FW), which is able to filter unwanted messages from OSN user walls. We exploit a new Machine Learning (ML) text categorization techniques **[5]** to automatically assign with each short text message a set of categories based on its posted content.

The experiments what we have carried out on the OSN users show the effectiveness of the developed filtering techniques. In particular, the overall technique of filtering mechanism was experimentally evaluated by taking several OSN users into account with the help of ML short classification method and we have successfully proved the effectiveness of the system in applying FRs. Finally, we have provided a new protocol implementation of our system having Facebook as target OSN, even if our proposed system can be easily applied to other OSNs which are readily available in the real time environment.

2. Background Work

In this section we will describe the assumptions that are used in the proposed paper. We also discuss the background work that was carried out for performing the filtering techniques.

2.1 Motivation Work

The main motivation work of this paper is to design a system with the help of ML Approach for filtering the unwanted messages in any OSN networks based on content based filtering. As we have already discussed about these concept in the introduction section, to the best of our knowledge, we are the first person to propose such kind of novel application for OSN networks. However, our proposed work has jointly relationships both with the state of the art in content-based filtering, as well as with the field of personalization based on access policies for OSNs and, more in general web contents.

2.2 Content-Based Filtering Technique (CBFT)

IF systems are the systems which are designed to classify a series of dynamically generated information, which is dispatched in an asynchronous manner by a two or more than different users likely to satisfy their requirements **[6].**In CBFT, each user is assumed to operate always independently. As a result, a CBFT system always selects information items based on the comparison between the item content and the user preferences as opposed to a collaborative filtering system that chooses items based on the comparison between people with similar preferences **[7], [8].** As the information filtering technique which comes under the branch of Artificial Intelligence, it is always done on text type of data, so the information filtering technique can also come under text

categorization mechanism, where the text can be categorized into various stems and each stem can be identified as blocking word or non-blocking word.

2.3 A Novel Policy-Based Personalization of OSN Contents

In OSNs there have been some proposals exploited for classification mechanism for access personalization. For example, in paper no [9], a classification method has been proposed to categorize short text messages in order to avoid overwhelming users of micro blogging services by raw data.The system described in paper no [9] focuses on Twitter and associates a set of categories with each tweet describing its content. The user has the privilege of choosing some specific type of tweets not all based on his privilege.

3. Proposed Methodology and its Filtered Architecture

In this paper we are going to implement filtered wall architecture in any OSN. The architecture in support of OSN services is a three-tier structure (as shown in Figure. 1). The first layer or primary layer , called Social Network Manager Layer (SNM), which is used to provide the basic OSN functionalities (i.e., profile management and relationship management), whereas the second layer provides the support for External Social Network Applications (SNAs). The third layer will be used in turn to provide Graphical User Interfaces (GUIs) support. According to this reference architecture, the proposed system is placed in the second and third layers.

In particular, users interact with the system by means of a GUI to set up and manage their FRs/BLs. Moreover, the GUI provides users with a FW, that is, a wall where only messages that are authorized according to their FRs/BLs are published.

From the Figure .1 we can clearly get any idea pictorially about the Filtered wall architecture of any OSN, the path followed by a message, from its writing to the possible final publication can be summarized as follows:

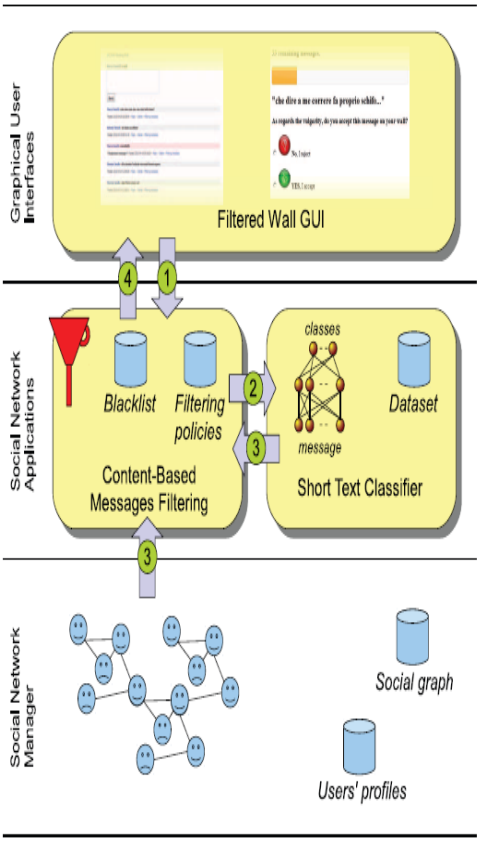

Figure. 1. Filtered wall conceptual architecture and the flow messages

Algorithm Steps

1. Firstly in order to login in OSN network, each and every osn user should register in this network with a valid user details.

2. After OSN user Successful registration he/she enters his login id and password in order to enter into their own private wall.

3. After he enters into his private wall of one of his/her contacts, the login user tries to post a message, which is intercepted by a FW.

4. A Machine Learning based text classifier method extracts data about data from the content of the message.

5. FW uses data about data provided by the classifier, together with data extracted from the social graph and users 'profiles, to enforce the filtering and BL rules.

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

4. Short Text Classifier Methodology (STCM)

STCM algorithm is mainly used for text categorization process, which is a methodology of Machine learning models. This was used in our proposed application in order to categorize the user message into a group of hierarchical stems and identify if there are any blocked content available in the transmitting message that was entered by the OSN user which is clearly shown in figure 2.

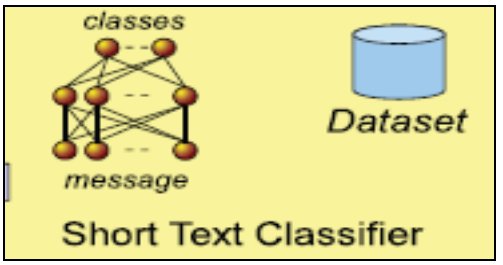

Figure. 2. Short Text Classifier Algorithm

This choice is mainly motivated by related proposed work showing advantages in classifying text and/or short texts using a hierarchical strategy [10]. The first-level problem is named as a hard classification problem in which short texts are labeled with crisp of two names like Neutral words and Nonneutral words. The second-level soft classifier acts on the crisp set of nonneutral short texts and, for each of them, it "simply" produces estimated appropriateness or "gradual membership" for each of the conceived classes, without taking any "hard "decision on any of them. Such a list of grades is then used by the subsequent phases of the filtering process.

5. Implementation Modules

Implementation is the stage where the theoretical design is automatically converted into practically by dividing this into various modules. Our proposed application is divided into following 3 modules. They are as follows:

Filtering rules Module

In OSNs 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.

Online setup assistant for FRs thresholds Module

We address the problem of setting thresholds to filter rules, by conceiving and implementing within FW, an Online Setup Assistant procedure. OSA presents the user with a set of messages selected from the dataset. The collection and processing of user decisions on an adequate set of messages distributed over all the classes allows computing customized thresholds representing the user attitude in accepting or rejecting certain

contents. A certain amount of non-neutral messages taken from a fraction of the dataset and not belonging to the training/test sets, are classified by the ML in order to have the second level class membership values.

Blacklists Module:

A Blacklist mechanism avoids messages from undesired creator's independent from their contents. We decide to let the users themselves to specify BL rules regulating who has to be banned from their walls and for how long according to their profiles as well as their relationships in the OSN. More precisely, among possible information denoting users' bad behavior we have focused on two main measures. The first is related to the principle that if within a given time interval a user has been inserted into a BL for several times. In contrast, to catch new bad behaviors, we use the Relative Frequency that let the system be able to detect those users whose messages continue to fail the FRs. The two measures can be computed either locally or globally.

6. Conclusion

In this paper, we have presented a system to filter undesired messages from OSN walls. The system exploits a ML soft classifier to enforce customizable content-dependent FRs. In particular, future plans contemplate a deeper investigation on two interdependent tasks. The first concerns the extraction and/ or selection of contextual features that have been shown to have a high discriminative power. The second task involves the learning phase. Since the underlying domain is dynamically changing, the collection of preclassified data may not be representative in the longer term. The present batch learning strategy based on the preliminary collection of the entire set of labeled data from experts, allowed an accurate experimental evaluation but needs to be evolved to include new operational requirements. In future work, this problem may be addressed by investigating the use of online learning paradigms able to include label feedbacks from users.

7. 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]http://www.facebook.com/press/info.php?statistics.

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

[6] N.J. Belkin and W.B. Croft, "Information Filtering and Informationn 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] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M.Demirbas, "Short

Text Classification in Twitter to ImproveInformation Filtering," Proc. 33rd Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '10), pp. 841-842,2010.

[10] 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.

8. About the Authors

Mulagapati DurgaDevi is currently pursuing her 2 Years M.Tech (CSE) in Computer Science and Engineering at Visakha Institute of Engineering & Technology, Narava, Visakhapatnam. Her area of interests includes Secure Computing. Her mail id is m.devi.mca.06@gmail.com

J.Peter Praveen is currently working as Assistant Professor, in Computer Science and Engineering at Visakha Institute of Engineering & Technology, Narava, Visakhapatnam. His research interests include Data Mining and Secure Computing. His mail id is jppraveen17@gmail.com

