A Novel Protocol for Preventing Posting of Abused Messages in any OSN Networks

Gopika Chaganti [#1], Prof. Peri Srinivasa Rao [*2]

M.Tech Scholar [#1], Professor & HOD [*2]

Department of Computer Science & System Engineering,
Andhra University College of Engineering,
Visakhapatnam, AP (INDIA)

Abstract

A **social network** is a social structure made up of a set of social actors (such as individuals or organizations) and a set of the dyadic ties between these actors. These social networks are generally spread all around the world through internet. Online Social Networks (OSNs) are today one of the most prominent interactive medium to communicate, share, and disseminate a considerable amount of human life information. As the social network is gaining its popularity in usage by various OSN users, the major problem that was faced by OSN user is the ability to control the message content posted on their own private space to avoid that unwanted content is displayed. Now a days a lot of users are posting very un-parliamentary or rubbish messages on their private walls and even post the same on others wall. To solve this problem, in this paper, we have proposed a novel filtering protocol allowing all participating OSN users to have a direct control on the messages posted on their walls. This protocol was implemented by using automatic identification of un-parliamentary words from the total message by using a Machine Learning (ML) based soft classifier algorithm which labels the messages into blocked content based on the category.

Keywords

ML Approach, Soft Classifier, Filtering Protocol, Private OSN Wall, Blocked Content

.

1. Introduction

In general, online social networks are self-organizing, emergent, and complex, such that a globally coherent pattern appears from the local interaction of the elements that make up the system. These patterns become more apparent as network size increases. Online Social Networks (OSNs) [1] are becoming day by day one of the most familiar interactive medium to communicate, share, and disseminate a considerable amount of human life information. In Online Social networks, 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.

In a recent Facebook survey what have been conducted by a team of experts, we came to know that on a Facebook on an average user creates 160 pieces of content each month, whereas more than 80 billion pieces of content are shared each month. The huge and dynamic character of these data creates the premise for the employment of web content mining [2], [3] strategies aimed to automatically discover useful information dormant within the data.

In this paper, our main 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.

Till now there was no such a type of mechanism in any OSN networks, this is the first proposal of a system, which automatically filter unwanted message content from OSN user walls on the basis of both message content and the message creator relationships and characteristics. The current paper substantially extends [5] for what concerns both the rule layer and the classification module. Major differences include, a different semantics for filtering rules to better fit the considered domain, an online setup assistant (OSA) to help users in FR specification, the extension of the set of features considered in the classification process, a more deep performance evaluation study and an update of the prototype implementation to reflect the changes made to the classification techniques.

2. Related Work

In this section we will describe the assumptions and background knowledge that is used for developing this novel message filtering protocol.

2.1 Main Motivation

The main motivation of this proposed paper is to design a novel system which provides customizable content-based message filtering for OSNs, based on Machine Learned (ML) Techniques. As we have already discussed out the same in the introduction column, to the best of our knowledge, we are the first proposing such kind of novel application for OSN networks. However, our proposed work has collaborative relationships both with the state of the art in content-based filtering, as well as with the field of policy-based personalization for OSNs and, more in general, web contents.

2.2 Novel Content-Based Filtering Model (NCBFM)

In NCBFM, each and every user is assumed to operate independently I.e. he will post the content on his own. As a result, a NCBFM system always selects information items based on the cor-relation between the content of the items and the user preferences as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences [7], [8]. While EMAIL service was the original domain of previous work on information filtering, several papers have addressed diversified domains including newswire articles, Internet "news" articles, and broader network resources [9], [10], [11]. As the information filtering is always done on text type of data, it may also come under text classification mechanism.

3. Novel Message Filtering Protocol (NMFP)

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).

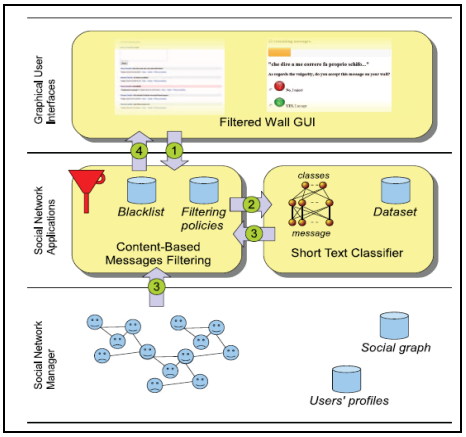

Figure. 1. Filtered wall conceptual architecture and the flow messages

1. The first layer or top most layer is called as Social Network Manager Layer (SNML), which is used to provide the basic OSN functionalities (i.e., It maintains Profile Management Details as Well as Relationship Details)

2. The second layer or middle layer provides the support for External Social Network Applications (ESNAs).

3. The third layer or lower 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

1. After an OSN user Successful login he/she enters the private wall of one of his/her contacts, the user tries to post a message, which is intercepted by a FW.

2. A ML-based text classifier method extracts data about data from the content of the message.

3. 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.

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

4. Short Text Classifier Algorithm

Short Text Classifier algorithm is mainly used for text categorization, which is a methodology of Machine learning models. This was used in our proposed application in order to categorize the user message into stems and identify if there are any filtered content available in the message that was passed by the OSN user which is clearly shown in figure 2.

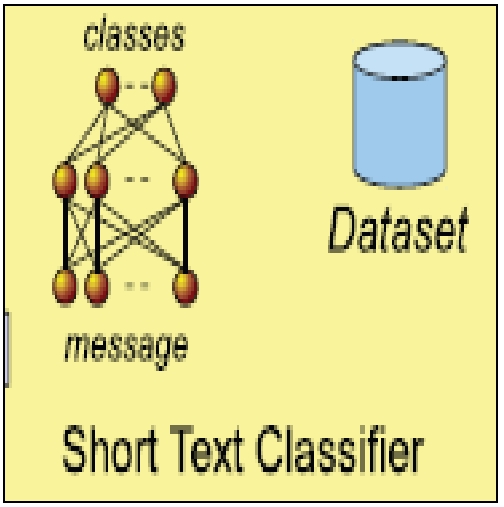

Figure. 2. Short Text Classifier Algorithm

This choice is mainly motivated by related proposed work showing advantages in classifying text and/or short texts using a hierarchical strategy [12]. 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 Non-neutral words. The second-level soft classifier acts on the crisp set of non-neutral 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 5 modules. They are as follows:

A. User Registration Module

In order to participate in OSN network communication each and every OSN user should register first for getting a valid username and password. So in this module all the OSN users register for getting a valid username and password for login into their walls.

B. User Authentication Module

In this module all the registered OSN users will try to enter into their individual private walls by giving their valid username and password what they have entered while registration process. If the user enters valid username and password, he/she can enter into their account and validation and authentication is success at that time. If the details are wrong then user can\t participate in communication.

C. 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.

D. 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.

E. 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 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] 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

Gopika Chaganti is currently pursuing her 2 Years M.Tech (CSSE) in Computer Science and System Engineering at Andhra University College of Engineering, Visakhapatnam. She completed her B.Tech from Gayatri Vidya Parishad. Her area of interests includes Data Mining.

Prof. Peri Srinivasa Rao is currently working as Professor & HOD, in Computer Science and System Engineering at Andhra University College of Engineering, Visakhapatnam. He completed Ph.D (IIT – Kharagpur). His research interests include Image Processing, Queuing Applications.

