Blocked Protocol: A New Filter to Block Un-Deliberative Content in Social Networks

J.Santoshi Kumari [#1], Chinna Babu Galinki [*2]

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

Department of Computer Science & Engineering,
Avanthi's St.Theressa Institute of Engineering & Technology,
Garividi, Vizianagaram Dist, AP, India.

Abstract

Online Social Networks (OSNs) are today one of the most popular interactive medium to communicate, share, and disseminate a considerable amount of human life information. With this OSN each and every user communicates with one other who was located in different regions around the world.Eventhough 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 rude messages on their private walls and even post the same on others wall. To solve this problem, in this paper, we have proposed a very new 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

Machine Learning, Soft Classifier, Filtering Protocol, Private OSN Wall

1. Introduction

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.

For example as per the Facebook statistics [4] what we have conducted we observed average user creates 150 pieces of content each month, whereas more than 70 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.

The major efforts in building a robust short text classifier (STC) are concentrated in the extraction and selection of a set of characterizing and discriminant features. The solutions investigated in this paper are an extension of those adopted in a previous work by us **[5]** from which we inherit the learning model and the elicitation procedure for generating pre classified data. The original set of features, derived from endogenous properties of short texts, is enlarged here including exogenous knowledge related to the context from which the messages originate. 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 **[4].**

To the best of our knowledge, 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. Background Knowledge

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

2.1 Main Motivation

The main motivation of this current paper is to design a system which provides customizable content-based message filtering for OSNs, based on Machine Learned Techniques. As we have already discussed out in the introduction, 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 Content-Based Filtering Model (CBFM)

Information filtering systems are the systems which are designed to classify a stream of dynamically generated information dispatched asynchronously by a two or more than different users likely to satisfy their requirements [6].In CBFM, each and every user is assumed to operate independently. As a result, a CBFM system always selects information items based on the correlation 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.

In contrast to the above statement, authors like **[13]** Golbeck and Kuter proposed an application, called A Film Trust that exploits OSN trust relationships and provenance information to personalize access to the website, but failed to provide a filtering mechanism separately to decide or restrict some un authorized messages. In contrast, our filtering policy language allows the setting of FRs according to a variety of criteria, which do not consider only the results of the classification process but also the relationships of the wall owner with other OSN users as well as information on the user profile. Moreover, our system is complemented by a flexible mechanism for BL management that provides a further opportunity of customization to the filtering procedure.

3. Message Filtering Protocol 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).

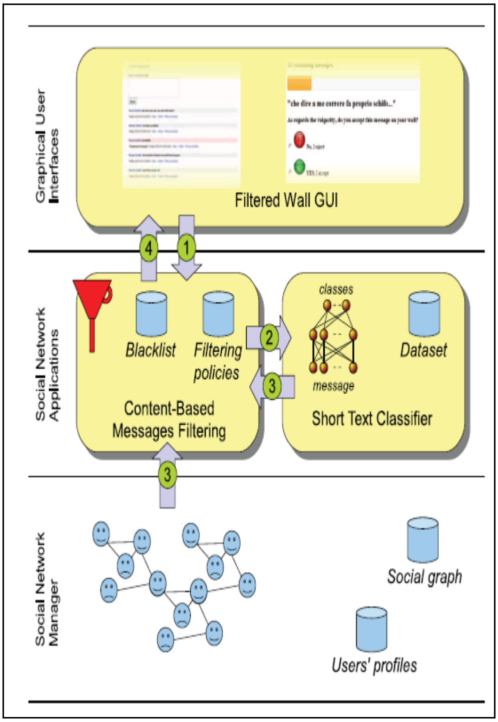

Figure. 1. Filtered wall conceptual architecture and the flow messages

1. The first layer or top 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 last 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. System Architecture Diagram

A further component of our system is a BL mechanism to avoid messages from undesired creators, independent from their contents. BLs are directly managed by the system, which should be able to determine who are the users to be inserted in the BL and decide when users retention in the BL is finished. To enhance flexibility, such information is given to the system through a set of rules, hereafter called Black List rules. This architecture flow was clearly shown in figure 2.

4

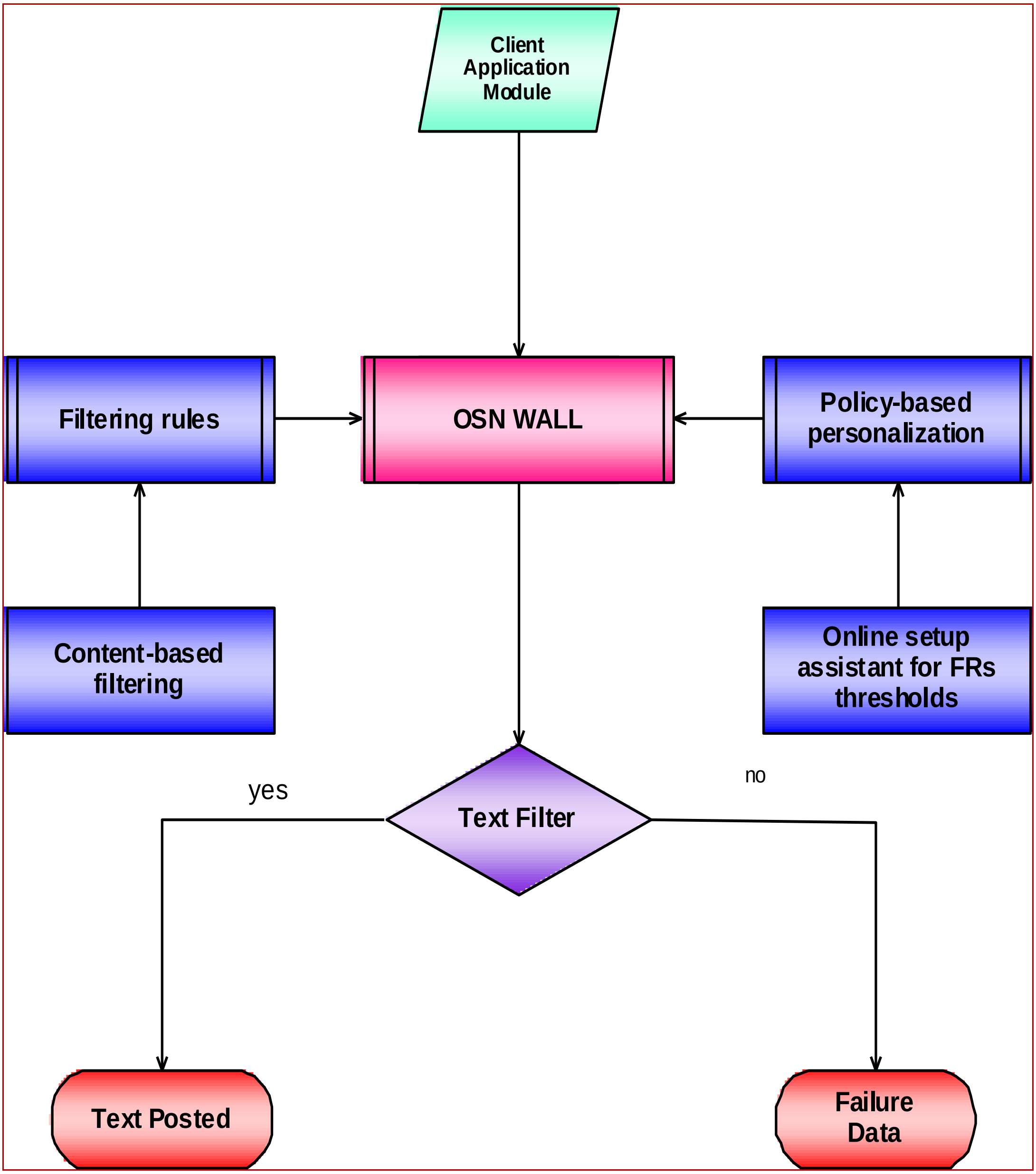

Figure. 2. System Flow Diagram

5. 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 3.

Our research study is aimed at designing and for evaluating various representation techniques in combination with a neural learning strategy to semantically categorize short texts. From a Machine Learning based mechanism, we classify the task by defining a hybrid two-level strategy assuming that it is better to identify and eliminate "neutral" sentences, then classify "non-neutral" sentences by the class of interest instead of doing everything in one step.

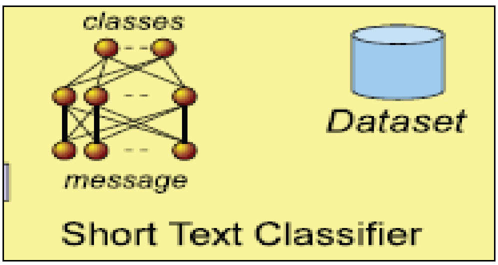

Figure. 3. Short Text Classifier Algorithm

This choice is mainly motivated by related proposed work showing advantages in classifying text and/or short texts using a hierarchical strategy **[14].** 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.

6. Implementation Modules

Implementation is the stage where the theoretical design is automatically converted into practically by dividing this into various modules. We have implemented the current application in Java Programming language with Front End as JSP,HTML and Back End as MYSQL data base. Our proposed application is divided into following 5 modules. They are as follows:

1. OSN User Registration Module
2. OSN User Authentication Module
3. Admin adds Filtering Words Module
4. Online Setup Assistant for FRs Threshold Module
5. Blacklist Module

1. OSN 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.

2. OSN 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.

3. Admin Adds Filtering Words Module

In this module the admin for this new OSN message filtering Protocol adds a several un-parlimentary words into the database table which holds the words based on category we choose. For example if we place the word like “Stupid” in the category to Vulgar.If the user who used the same word while communication, he gets reply as Stupid comes under communication layer.

4. 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.

5. 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.Another new component of our proposed system is a BL mechanism to avoid messages from undesired creators, independent from their contents. 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. Such rules are not defined by the SNMP; therefore, they are not meant as general high-level directives to be applied to the whole community. Rather, we 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, by, at the same time, being able to post in other walls.

7. 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. To the best of our knowledge, this is the first proposal of a system to automatically filter unwanted messages from OSN user walls on the basis of both message content and the message creator relationships and characteristics. The current paper substantially extends for what concerns both the rule layer and the classification module. The experiments we have carried out show the effectiveness of the developed filtering techniques. In particular, the overall strategy was experimentally evaluated numerically assessing the performances of the ML short classification stage and subsequently proving the effectiveness of the system in applying FRs. Finally, we have provided a prototype implementation of our system having Facebook as target OSN, even if our system can be easily applied to other OSNs as well.

8. 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] 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] 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.

[13] J. Golbeck, "Combining Provenance with Trust in Social Networks for Semantic Web Content Filtering," Proc. Int'l Conf. Provenance and Annotation of Data, L. Moreau and I. Foster, eds., pp. 101-108,2006.

[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.

9 About the Authors

J.Santoshi Kumari is currently pursuing her 2 Years M.Tech (CSE) in Department of Computer Science and Engineering at Avanthi's St Theressa Institute of Engineering and Technology, Garividi, Vizianagaram Dist. Her area of interests includes Networks, Data Mining and Data Ware Housing.

Mr.Chinna Babu Galinki ,well known excellent teacher Received M.Tech (CSE) from Andhra university is working as Associate Professor and HOD, Department of Computer Science & Engineering , Avanthi's St Theressa Institute of Engineering and Technology. He has 4 years of teaching experience. To his credit couple of publications both national and international conferences /journals. His area of Interest includes Data Warehouse and Data Mining, Embedded Systems and other advances in computer Applications.

