A Novel Mechanism of Preventing Unparlimentary Messages in OSN Networks

Budidha Srinivasu [#1], **Bommireddy Dinesh Reddy** [*2]

M.Tech Scholar [#1], Associate Professor, M.Tech (PhD)[*2]

Department of Computer Science & Engineering,
Vignan Institute Of Information Technology,
Visakhapatnam,AP,India.

Abstract

Online Social Networks (OSNs) are becoming day by day one of the most familiar interactive medium to communicate, share, and disseminate a considerable amount of human life information. As the OSN Network is gaining its popularity ,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 novel filtering system allowing all participating OSN users to have a direct control on the messages posted on their walls. This is achieved by a Machine Learning (ML) based soft classifier algorithm which is used which for automatically labeling messages if it is recognized as block able content.

Keywords

Machine Learning, Soft Classifier, OSN

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. Daily and continuous communications [2] imply the exchange of several types of content, including free text, image, audio, and video data. For example as per the Facebook statistics [4] what we have conducted we observed average user creates 100 pieces of content each month, whereas more than 60 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 [3] strategies aimed to automatically discover useful information dormant within the data.

The main aim of the maximum number of researchers on this current problem is to provide OSN users a classification mechanism 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 as the ability to automatically control the messages written on their individual walls, by pointing out unwanted messages. We believe that this is a key OSN service that has not been provided so far. Though face book networks are capable of providing some sort of security for the stakeholders who participate in OSN, it failed in providing whole security. 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 defined web content mining techniques for a different application, rather it 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.

2. Related Work

In this section we will describe the assumptions that are used in the proposed paper.

2.1 Motivation

The main motivation of this paper is to design a system which provides customizable content-based message filtering for OSNs, based on ML 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 Mechanism (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.

2.2 A New 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 [12], 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 [12] 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.

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. 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 (SNML), 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 (ESNAs). 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.

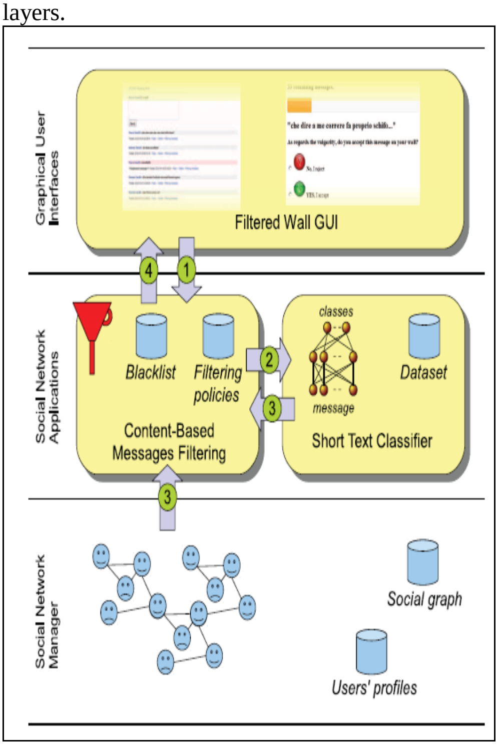
Graphical User Interfaces
"che dire a me correre fa proprio schifo..."
As regards the vulgarity, do you accept this message on your wall?
No, I reject
YES, I accept
Filtered Wall GUI
4
1
Social Network Applications
Blacklist
Filtering policies
Content-Based Messages Filtering
2
3
classes
message
Dataset
Short Text Classifier
3
Social Network Manager
Social graph
Users' profiles

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

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 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 (STCA)

STC 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.

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.

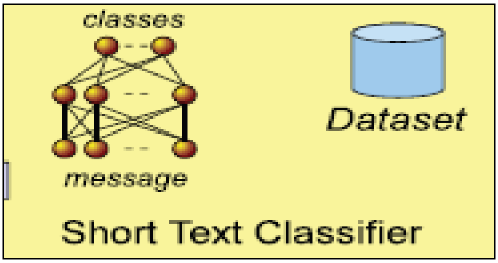

Figure. 2. Short Text Classifier Algorithm

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

5. Blacklists

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.

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. Moreover, the flexibility of the system in terms of filtering options is enhanced through the management of BLs. 8http://apps.facebook.com/dicompostfw/ This work is the first step of a wider project. The early encouraging results we have obtained on the classification procedure prompt us to continue with other work that will aim to improve the quality of classification. 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 pre-classified data may not be representative in the longer term.

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] 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.

8. About the Authors

Budidha Srinivasu is currently pursuing his 2 Years M.Tech (CSE) in Department of Computer Science and Engineering at Vignan Institute of Information Technology, Visakhapatnam. His area of interests includes Data Mining.

Bommireddy Dinesh Reddy is currently working as Associate Professor in Department of Computer Science and Engineering at Vignan Institute of Information Technology, Visakhapatnam. He is currently doing his PhD in the relevant field of Computer Science. His research interests include Networks Security, Information Security and Data Mining.