Mining Suspicious Messages in Online Forums

Priyanka[#1], Poorani[#1], Bhavani[#1],Kamalameena[#1] & Hariharan[#2]

Student[#1],Associate Professor[#2]
Department of computer Science and Engineering
TRP Engineering College, Irungalur, Trichy, Tamil Nadu, India
[1]priyankavenu94@gmail.com
[1]poorani.k17@gmail.com
[1]bhavanirajendran08@gmail.com
[1]kamala1695@gmail.com
[2]mailtos.hariharan@gmail.com

Abstract --- **Now-a-days there is an increasing dependency on web applications, ranging from individuals to large organizations. The security motivation of web application especially online social network is a great concern. Therefore in this paper a system is proposed which is an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls. We exploit Machine Learning (ML) text categorization techniques to automatically assign with each short text message a set of categories based on its 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 discriminate features. The solutions investigated in this paper are an extension of those adopted in a previous work by us from whom we inherit the learning model and the elicitation procedure for generating reclassified 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. In addition, the system provides the support for user-defined Blacklists (BLs), that is, lists of users that are temporarily prevented to post any kind of messages on a user wall.**

Keywords --- **Machine Learning, Filtered wall, short text classifier, Blacklists**

I. INTRODUCTION

Online Social Networks (OSNs) which are today one of the most popular interactive medium to share, communicate, and distribute an important amount of human living information. Along with Face book information average user creates 90 pieces of substance every month, while more than 30 billion quantity of substance (web links, news stories, notes, blog posts, photo albums, etc.) are distributed every month. The vast and dynamic character of this information produces the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant contained by the information[8]. Information filtering has been significantly searched for what concerns textual documents and, more recently, web content. However, the aim of the majority of these proposals is mainly to provide users a classification mechanism to avoid they are overwhelmed by unsuccessful information [7].

In OSNs, information filtering can also be exploited for a dissimilar, more responsive, purpose. This is due to the fact that in OSNs there is the possibility of posting or commenting other posts on exacting public/private regions, called in common walls. Information filtering can therefore be used to provide users the capability to automatically control the messages written on their individual walls, by filtering out surplus communication. We believe that this is a key OSN service that has not been offered so far [8]. Certainly, in the present day OSNs provide very tiny maintain to prevent unwanted messages on user walls. For instance, Face book permits users to status who is allowed to insert messages in their walls [10].Though, no content-based preferences are maintained and therefore it is not possible to prevent undesired messages, for instance political or offensive ones, no matter of the user who posts them. Providing this service is not only a topic of using previously defined web content mining methods for a different purposes, rather it entails to propose ad-hoc categorization strategies [12].This is because wall messages are represented by tiny text for which traditional classification methods have serious limitations since short texts do not provide sufficient word occurrences.

This paper includes the proposed system which filters unwanted messages from Online Social Networks user walls on the basis of both message content and the message creator relationships and characteristics [13]. It acts as a security for online chatting and act as evidence for investigation. It provides the support for user-defined Blacklists (BLs).Blacklists is lists of users that are temporarily prevented to post any kind of messages on a user wall [18].

The paper is organized as follows. Section 1 presents the introduction. Section 2 highlights some of the related work and section presents the proposed work. Section 4 gives the results and the outcome of the proposed system which includes the output image filtering the vulgar message. Section 5 concludes and describes the future work which could be enforced on this system.

II. RELATED WORK

Indeed, today OSNs provide very little support to prevent unwanted messages on user walls. For example, Face book allows users to state who is allowed to insert messages in their walls (i.e., friends, friends of friends, or defined groups of friends). 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, [21] this paper proposes a system enforcing content-based message filtering conceived as a key service for On-line Social Networks (OSNs).The system allows OSN users to have a direct control on the messages posted on their walls. This is achieved through a flexible rule-based system, that allows a user to customize the filtering criteria to be applied to their walls, and a Machine Learning based soft classifier automatically producing membership labels in support of content-based filtering.

Text Categorization technique to Filter Unwanted Messages in OSN plays a vital role in day to day life[4]. User can communicate with other user by sharing several types of contents like image, audio and video contents. A major issue in OSN (Online Social Network) is to preventing security in posting unwanted messages. Ability to have a direct control over the messages posted on user wall is not provided .Unwanted post will be directly posted on the public wall. Only the unwanted messages will be blocked. To avoid this issue Black List mechanism is proposed in this work which avoids undesired creators messages. Black list is used to determine which user should be inserted in BL and decide when the retention of the user is finished. Machine Learning Text Categorization is also used to categorize the short text messages.

This paper [22] relates to current interests of the user and adapt as they change over time. It must also explore newer domains for potentially interesting information. Thus, a learning approach to building personalized information filtering systems is proposed. The system is designed as a collection of information filtering interface agents. Interface Agents are intelligent and autonomous computer programs which learn users preferences and act on their behalf | electronic personal assistants that automate tasks for the user. This thesis presents the basic framework for personalized information filtering agents, and describes an implementation, "Newt", built using the framework. Newt uses a keyword based filtering algorithm. The learning mechanisms used are relevance feedback and the genetic algorithm. The user interface is friendly and accessible to both naive as well as power users.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 [21]. 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.

III. PROPOSED WORK

We exploit text categorization techniques to automatically assign with each short text message a set of categories based on its 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 discriminate features [16].The aim of the present work is therefore to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls. The solutions investigated in this work are an extension of those adopted in a previous work by us from whom we inherit the learning model and the elicitation procedure for generating pre classified data [10]. 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. In addition, the system provides the support for user-defined Blacklists (BLs), that is, lists of users that are temporarily prevented to post any kind of messages on a user wall [13].

In OSNs, information filtering can also be exploited for a dissimilar, more responsive, purpose. This is due to the fact that in OSNs there is the possibility of posting or commenting other posts on exacting public/private regions, called in common walls. Information filtering can therefore be used to provide users the capability to automatically control the messages written on their individual walls, by filtering out surplus communication. We believe that this is a key OSN service that has not been offered so far [8].
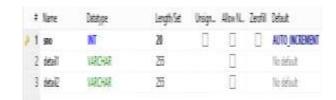
A. *Modules of the Proposed Work*

➢ **Authentication module**

Only the authorized person can enter by giving valid information. For invalid information the permission is denied. This module concentrates the security of the project from the unauthorized users.

➢ **New User Registration**

The new user has to register with the proper details. For getting the rights to accessing the features users have to register their identity to this system. Once registered, the system will provide the accessibility to the users.

Filter Message

In this social networking messages are sent .Messages are very short type .In this we can communicate each other. The messages are filtered using two techniques machine learning and short text classifier. If the messages are legal then it gets posted if it is illegal then it is blocked.

➢ **Admin**

Admin has all the user control and registration details. It also contains the filter performance details and addition of category based words in it. All these are stored in the database.

B. *Architecture*
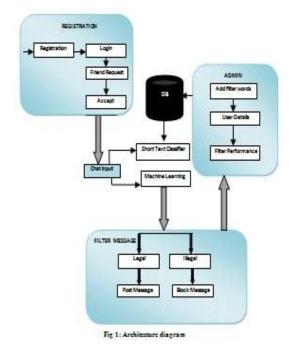

Fig 1: Architecture diagram

Table 1: New user Registration

IV. RESULT OF THE PROPOSED WORK

Thus, proposed system filters the unwanted messages, adds to the filter performance and analyses it. It blocks the suspicious word and adds to the particular category. So it doesn't get posted to the recipient. It increases the online chatting security and it is also used for investigation purposes.

Fig 2: Filtered Message

V. CONCLUSION AND FUTURE WORK

We have presented a system to filter undesired messages from Online Social Network walls. The system exploits a text classifier to enforce customizable content-dependent Filtering Rules. Moreover, the flexibility of the system in terms of filtering options is enhanced through the management of Black Lists. 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.

As future work, we intend to exploit similar techniques More improve the Black List and Filtering Rules. Additionally, we plan to study strategies and techniques that a user can do on the enforced filtering rules with the aim of bypassing the filtering system a message that should instead be blocked.

REFERENCES

• Abulaish M, Anwar T. A webcontent mining approach for tag cloud generation. In: Proc. of the 13th Int'l Conf. on IIWAS; 2011. p. 52e9.

• Abulaish M, Anwar T. A key phrase-based tag cloud generation framework to conceptualize textual data. Int J Adapt Resil Aut Syst (IJARAS) 2013; 4(2):72e93.

• Adams PH, Martell CH. Topic detection and extraction in chat. In: Proc. Of the IEEE Int'l Conf. on Semantic Computing, ICSC '08; 2008. p. 581e8.

- Agarwal S, Godbole S, Punjani D, Roy S. How much noise is too much: a study in automatic text classification? In: Proc. of the 2007 Seventh IEEE Int'l Conf. on Data Mining; 2007. p. 3e12.
- Al-Zaidy R, Fung BCM, Youssef AM, Fortin F. Mining criminal networks from unstructured text documents. Digit Investig 2012; 8(3e4): 147e60.
- Allan J, Wade C, Bolivar A. Retrieval and novelty detection at the sentence level. In: Proc. of the ACM SIGIR Conf; 2003. p.314e21.
- Anwar T, Abulaish M. Web content mining for alias identification: a first step towards suspect tracking. In: Proc. of the IEEE Int'l Conf. on ISI; 2011. p. 195e7.
- Anwar T, Abulaish M. Identifying cliques in dark web forums e an agglomerative clustering approach. In: Proc. of the IEEE Int'l Conf. on ISI; 2012.
- Anwar T, Abulaish M. Namesake alias mining on the web and its role towards suspect tracking. Inf Sci 2014; 276:123e45.
- Aw TA, Lee LH. Personalized normalization for a multilingual chat system. In: Proc. of thw 50th Ann. Meeting of the ACL; 2012. p. 31e6.
- Beebe NL, Clark JG. Digital forensic text string searching: improving information retrieval effectiveness by thematically clustering search results. Digit Investig 2007;4(Suppl.):49e54.
- Bengel J, Gauch S, Mittur E, Vijyaraghavan R. Chattrack: chat room topic detection using classification. In: Proc. of the 2nd Symposium on ISI;2004. p. 266e77.
- Bogdanova D, Rosso P, Solorio T. On the impact of sentiment and emotion based features in detecting online sexual predators. In: Proc. of the ACL 2012 Workshop on Computational Approaches to Subjectivity and Sentiment Analysis; 2012. p. 110e8.
- Chu H-C, Deng D-J, Park JH. Live data mining concerning social networking forensics based on a facebook session through aggregation of social data. IEEE JSel Areas Commun 2011;29(7).
- Elsner M, Charniak E. Disentangling chat. Comput Linguist 2010;36(3):389e409.
- Holmer T, Discourse structure analysis of chat communication, Language@ Internet 5.
- Iqbal F, Fung BCM, Debbabi M. Mining criminal networks from chat log.In: Proc. of the IEEE/WIC/ACM Int'l Jt. Conf. on Web Intelligence and Intelligent Agent Technology, WI-IAT '12; 2012. p. 332e7.
- Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM Comput Surv 1999;31(3):264e323.
- Kamal A, Abulaish M, Anwar T. Mining feature-opinion pairs and their reliability scores from web opinion sources. In: Proc. of the 2nd Int'l Conf. on WIMS; 2012.
- M.Vanetti, E. Binaghi.Content-based Filtering in On-line Social Networks.
- J.Anishya Rose. Machine Learning Text Categorization in OSN to Filter Unwanted text.
- Beerud Dilip Seth. A Learning Approach to Personalized Information Filtering.