International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2014): 5.611

Threshold Based Filtering System for OSN-Online Social Networks

Shareen S. Anthony[1], Nitin A. Shelke[2]

[1]M.E. final year CSE, GHRCEM, Amravati, India

[2]Assistant Professor, Department of Computer Science and Engineering, GHRCEM, Amravati, India

Abstract: ***Online Social Networks (OSNs) are today one of the most popular interactive medium to share, communicate, and distribute a significant amount of human life information. In OSNs, information filtering can also be used for a different, more responsive, function. This is owing to the fact that in OSNs there is the possibility of posting or commenting other posts on particular public/private regions, 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. OSNs provide negligible amount of support prevent undesired messages on user walls. To propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls. The proposed work deals with the prepossessing steps which is used to decrease the size of the database containing the abusive words.***

Keywords: Online Social Networks, Machine Learning, Filtering Rules, Content-based filtering.

1. Introduction

Information and communication technology plays a significant role in today's networked society. It has affected the online interaction between users, who are aware of security applications and their implications on personal privacy. There is a need to develop more security mechanisms for different communication technologies, particularly online social networks. OSNs provide negligible amount of support to prevent unwanted messages on user walls. With the lack of classification or filtering tools, the user receives all messages posted by the users he follows. In most cases, the user receives a noisy stream of updates. In this paper, an information Filtering system is introduced. The system focuses on one kind of feeds: Lists which are a manually selected group of users on OSN. List feeds tend to be focused on specific topics; however it is still noisy due to irrelevant messages. Therefore, we propose an online filtering system, which extracts such topics in a list, filtering out irrelevant messages [1]. Following is the conceptual architecture of Filtering System.

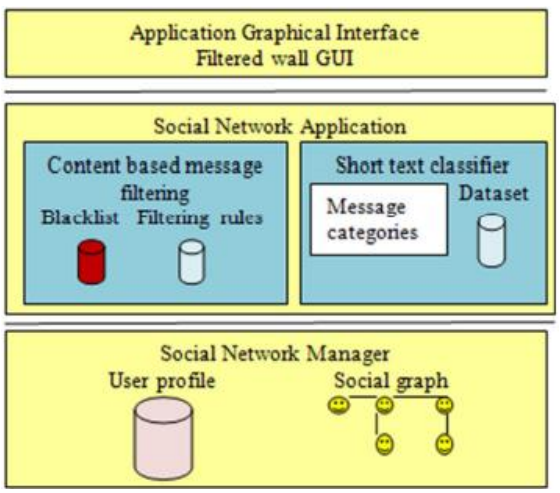

Figure 1: Filtering System Conceptual Architecture

In OSNs, 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. In the proposed system 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. 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. We exploit Machine Learning (ML) text categorization techniques [2] 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 are concentrated in the extraction and selection of a set of characterizing and discriminate features.

2. Proposed Work

Our goal is to design an online message filtering system that is deployed at the OSN service provider side. Once deployed, it inspects every message before rendering the message to the intended recipients and makes immediate decision on whether or not the message under inspection should be dropped.

2.1 Working Modules

2.1.1 Filtering Rules

A powerful rule layer exploiting a flexible language is provided by the system to specify Filtering Rules (FRs), by which users are able to state what contents, should not be displayed on their walls.

Licensed Under Creative Commons Attribution CC BY

2.1.2 Online setup assistant for FRs thresholds

OSA presents the user with a set of messages selected from the dataset discussed. For each message, the user tells the system the decision to accept or reject the message. The collection and processing of user decisions on an adequate set of messages distributed over all the classes allows to compute customized thresholds representing the user attitude in accepting or rejecting certain contents. Such messages are selected according to the following process. 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, for each message, the second level class membership values.

Suppose that Bob is an OSN user and he wants to always block messages having an high degree of vulgar content. Through the session with OSA, the threshold representing the user attitude for the Vulgar class is set to 0.8. Now, suppose that Bob wants to filter only messages coming from indirect friends, whereas for direct friends such messages should be blocked only for those users whose trust value is below 0.5. These filtering criteria can be easily specified through the following FRs5:
((Bob, friendOf, 2, 1), (Vulgar, 0.80), block)
((Bob, friendOf, 1, 0.5), (Vulgar, 0.80), block)

2.1.3 Blacklist

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 are given to the system through a set of rules, hereafter called BL rules.

BL rule:- A BL rule is a tuple (author, creatorSpec, creatorBehavior, T) where author is the OSN user who specifies the rule, i.e., the wall owner; creatorSpec is a creator specification; creatorBehavior consists of two components RFBlocked and minBanned.

RFBlocked = (RF, mode, window) is defined such that:– RF = #bMessages/#tMessages , where #tMessages is the total number of messages that each OSN user identified by creatorSpec has tried to publish in the author wall (mode = myWall) or in all the OSN walls (mode = SN); whereas #bMessages is the number of messages among those in #tMessages that have been blocked; window is the time interval of creation of those messages that have to be considered for RF computation; minBanned = (min, mode, window), where min is the minimum number of times in the time interval specified in window that OSN users identified by creatorSpec have to be inserted into the BL due to BL rules specified by author wall (mode = myWall) or all OSN users (mode = SN) in order to satisfy the constraint. T denotes the time period the users identified by creatorSpec and creatorBehavior have to be banned from author wall.
Example: The BL rule:

(Alice, (Age < 16), (0.5, myWall, 1 week), 3 days) inserts into the BL associated with Alice's wall those young users (i.e., with age less than 16) that in the last week have a relative frequency of blocked messages on Alice's wall greater than or equal to 0:5. Moreover, the rule specifies that these banned users have to stay in the BL for three days.

2.1.4 Preprocessing

The primary aim of the pre-processing phase is to remove from the input message all characters and terms that can possibly affect the quality of group descriptions.

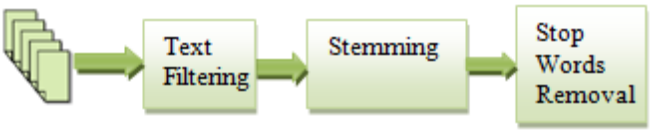

Figure 2: Pre-processing of Message

(a) Pre-processing steps:
/** Phase 1: Preprocessing */
for each document
{
do text filtering;
identify the document's language;
apply stemming;
mark stop words;
}

(b) Pre-processing phase pseudo-code
Algorithm 1:
1: d← input message
{STEP 1: Preprocessing}
2: for all d € D do
3: perform text categorization
4: if d!=null then
Filter text for unwanted symbols
5: apply stemming and mark stop-words in d;
6: end for

There are three steps to the preprocessing phase: Text filtering, Stemming and Stop words marking.

(1) Text filtering

In the text filtering step, all terms that are useless or would introduce noise in filtering process are removed from the input message. Among such terms are:
• HTML tags (e.g. <table>) and entities (e.g. &) if any.
• non-letter characters such as "$", "%" or "#" (except white spaces and sentence markers such as '.', '?' or '!')
Note that at this stage the stop-words are not removed from the input.

(2) Stemming

Stemming algorithms are used to transform the words in texts into their grammatical root form, and are mainly used to improve the Information Retrieval System's efficiency. To stem a word is to reduce it to a more general form, possibly its root. For example, stemming the term interesting may produce the term interest. Though the stem of a word might not be its root, we want all words that have the same stem to have the same root.

(3) Elimination of Stop Words

After stemming it is necessary to remove unwanted words. There are 400 to 500 types of stop words such as "of", "and",

Licensed Under Creative Commons Attribution CC BY

"the," etc., that provide no useful information about the message. Stop-word removal is the process of removing these words. Stop-words account for about 20% of all words in a typical document. These techniques greatly reduce the size of the searching and matching each word in message. Stemming alone can reduce the size of an index by nearly 40%. The following figure shows the DFD of Proposed Work

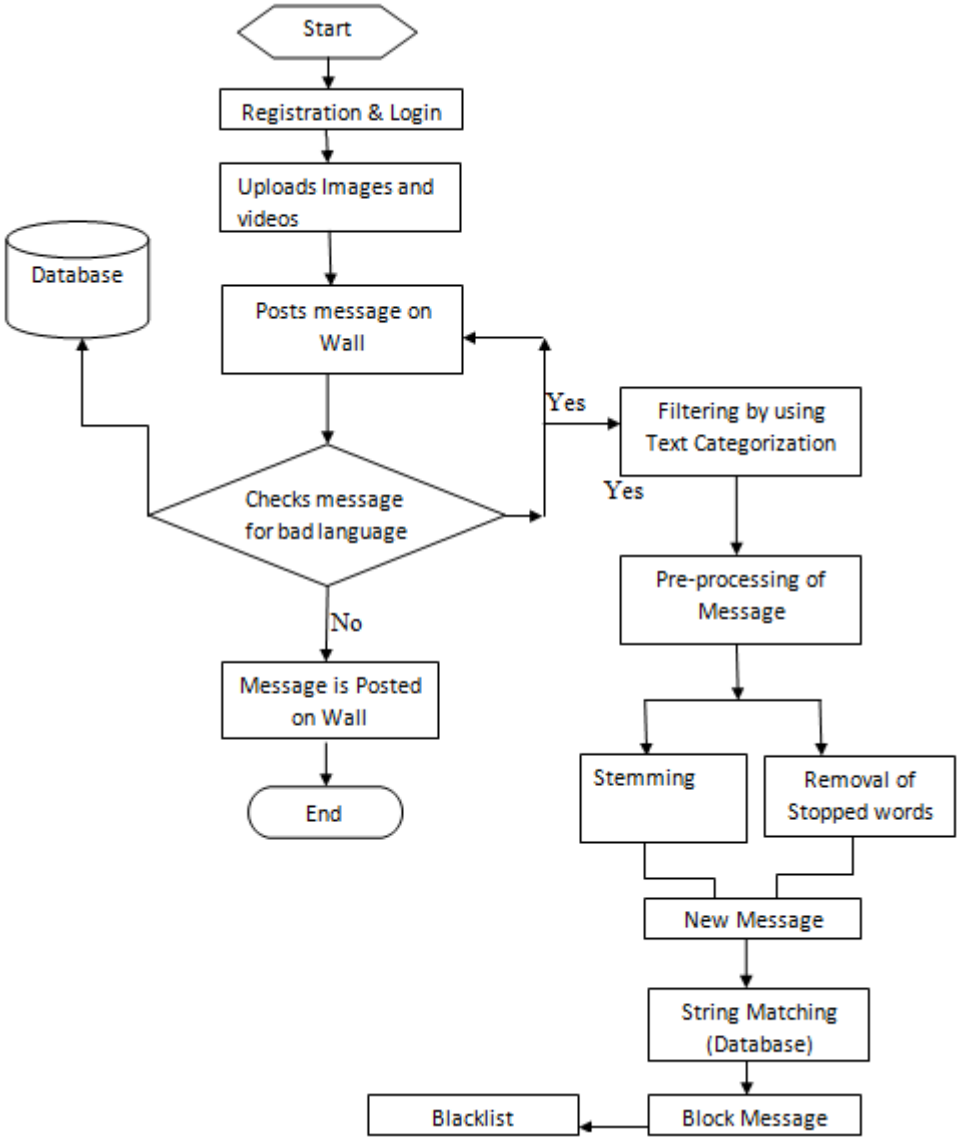

Figure 3: DFD Proposed Work

3. Conclusion

In this paper, we describe our work to provide unwanted message filtering for social networks. 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, management of BLs enhances the flexibility of the system in terms of filtering options. We would like to remark that the system proposed in this paper represents just the core set of functionalities needed to provide a sophisticated tool for OSN message filtering. Additionally, we studied strategies and techniques limiting the inferences that a user can do on the enforced filtering rules with the aim of bypassing the filtering system, such as for instance randomly notifying a message that should instead be blocked, or detecting modifications to profile attributes that have been made for the only purpose of defeating the filtering system.

.

References

[1] Miss. Shareen Anthony, Mr. Nitin Shelke, "A Survey of Filtering System for OSN (Online Social Networks)", IJCST, 2014.

[2] "Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno Carullo, "*A System to Filter Unwanted Messages from OSN User Walls*",2013.

[3] 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.

[4] J.Rose, A. Pravin *"Machine Learning Text Categorization in OSN to Filter Unwanted Messages",* IJCSIT, vol.5(1), 2014, pp.640-643, 2014.

[5] A.Adomavicius and G.Tuzhilin, *"Toward the Next Generation of Recommender Systems: A Survey of the State-of-the -Art and Possible Extensions , "* IEEE

Licensed Under Creative Commons Attribution CC BY

Trans. Knowledge and Data Eng.,vol.17, no.6,pp.734-749, June 2005.

[6] B.Sriram, D.Fuhry, E.Demir, H.ferhatatosmanoglu, and M.Demirbas, *"Short Text Classification in Twitter to Improve Information Filtering,"* Proc.33rd Int'l ACM SIGIT Conf. Research and Development in Information Retrieval(sIGIR '10), pp.841-842,2010.

Author Profile

Prof. N. A. Shelke has completed his M.Tech in Computer Science and Engineering from Government College of Engineering Amravati. He is currently working as an Assistant Professor in GHRCEM Amravati. His area of research includes Image Processing and Soft Computing.

Miss Shareen S. Anthony has completed her B.E. in Computer Science and Engineering from Sant Gadge Baba University Amravati. She is currently pursuing the M.E. degree at GHRCEM College Amravati,

www.ijsr.net
IJSR
ISSN (Online): 2319-7064

Licensed Under Creative Commons Attribution CC BY