Imperial Journal of Interdisciplinary Research (IJIR)
Vol-2, Issue-6, 2016
ISSN: 2454-1362, http://www.onlinejournal.in

A System to Filter Unwanted Messages from OSN User Walls

Viraj S. Kolapkar[1], Rohit V. Koul[2], Amit Mahajan[3] & Khushal C. Waghalikar[4]
[1,2,3,4]NDMVP's KBT COE, Nashik

Abstract*: One fundamental issue in today's Online Social Networks(OSNs) is to provide the users the ability to control the messages posted on their own private so that they can avoid the unwanted messages to be displayed/posted on their walls. The social networking sites are considered as the most popular medium to communicate with each other. They provide the best entertainment for younger generation in today's period. The OSNs help the individuals to connect with friends, family and the society in order to share experiences, thoughts and ideas. Now-a-days the problem faced by OSNs is the indecent messages that are posted on any of the individuals (user) wall which annoys the user on seeing them. Almost everyone of us are on Facebook and we know that Facebook allows users to state who is allowed to post messages into their walls(i.e., friends, friends of friends or a specified group). However content-based preferences are not supported in order to avoid unwanted messages such as political, vulgar, etc,. from other users. Our aim is to propose and experimentally evaluate an automated system called as the Filtered Wall(FW) which is able to filter unwanted messages from OSN user walls. The use of Machine learning Text Classifier is made to prevent unbearable messages on OSN user walls. It classifies the message into categories based on its content. We use a Blacklist mechanism which is used to avoid messages from undesired creators. The BL is used to determine that which user should be inserted in BL and decide that when the retention time of user has finished..*

1. Introduction

Online Social Networks (OSNs) offer a online medium for people to communicate with each other by sharing information. The shared information or content may be of the types such as text, audio, video, images, etc, [2][3]. Almost all the younger generation and most of the other age group people prefer and like the online social medium for interaction. Interaction includes sharing of thoughts, ideas, views, opinions and experiences with our closed ones. A huge and tremendous amount of data is being created and shared every month by several users according to the generated statistical analysis. Hence OSNs are the most popular and interactive medium used for communication. Today OSNs do not provide the ability for the users to have
control on the messages posted on their general walls. Users are not supported in terms of having control to the information or messages that are posted on their user walls. Hence, the use of information filtering becomes necessary for users who do not want unwanted messages to be displayed to them. Filtering of data/information helps in increasing the user security [3].

The goal of the presented work is to propose and experimentally evaluate an automated system called as the Filtered Wall (FW) which is able to filter unwanted messages from OSN user walls. We exploit Machine Learning (ML) text categorization techniques to automatically assign with each short text a set of categories based on its content [2]. The learning model includes the use of neural learning which is today recognized as one of the most efficient solution in text classification. We base the overall short text classification strategy by using a the Naïve Bayes Classifier[7]. They are preferred over all machine learning techniques because of their robustness and capabilities as acting as soft classifiers. We implement a neural model within a hierarchical two level classification strategy. In the first level, short messages are classified as Neutral(Positive) and Non-Neutral(Negative); in the second stage, Non-Neutral messages are classified producing gradual estimates of appropriateness to each of the considered category [1], [4].

In addition to classification provided, the system provides a powerful rule layer exploiting a flexible language to specify Filtering Rules (FRs), by which users can state what contents, should not be displayed on their walls. FR can contain different criteria that helps the users to set or customize the type of content that should not be displayed to them. FRs exploits use relationships user profiles as well as user defined Blacklists (BLs) [1].

2. RELATED WORK

Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno Carullo have

proposed "A System to Filter Unwanted Messages from OSN User Walls " which allows the user to have direct control over the messages posted on their user walls. In this research paper they discussed the way to filter and necessity of the filtering. Also given the existing system and told what "A system to filter unwanted messages from osn users wall" is. Hongyu Gao, Alok Choudhary, Yan Chen, Northwestern University Evanston, IL, USA , this group discussed the topic towards the online spam filtering in social networks. They have given the way to filter the messages by means of developing the plugins for the OSN. Everyone in a life time face this type of unfortunate nornade. The discussed how they got phished and affected their OSN account by the malware attack from the other users or most probably attackers.

3. SYSTEM ARCHITECTURE

The best way to model a Online Social Network is by using a directed graph. A graph contains nodes and edges. Similarly a graphical representation of social network contains nodes that represent network user and edges represent the relationship between two users. Relationship may be of the following types: friends, friend of friends, student of, parent of, etc contains categorization of messages according to content with the help of CBMF filters. A Blacklist is maintained for the users who send badwords in the message frequently[5].

3. Graphical User Interface(GUI):

In this layer FRs are used to filter unwanted messages and BL is provided for blocking the users who sends unwanted messages to the user. Type of unwanted messages are specified by the user.

i. After entering the private wall of one of his/her contacts the user tries to post a message which is intercepted by the FW.
ii. A Machine Learning based classifier extracts metadata from the contents of the message.
iii. The Filtered Wall uses this metadata together with the data extracted from social graph and user profile to enforce the filtering and blacklist rules.
iv. Based on the previous results the message will be filtered and shown to the user.

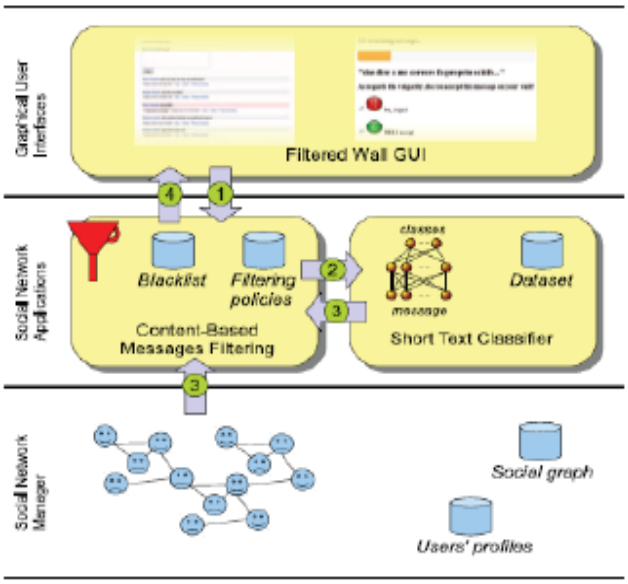

Figure :System Architecture.

4. PROPOSED METHODS/TECH

The following methods are used in the proposed system.

A. Short Text Classifier:

The existing text classification methods work well when the data is related to large documents such as newswires corpora. These techniques do not work properly when documents in corpus are short(i.e. short messages sent to each other while having a online chat).From Machine learning point of view we classify the task by using a two level strategy. It includes classifying sentences into "Neutral" and "Non-Neutral" by the class of interest. Our goal of study is to design, evaluate and implement representation techniques using neural learning strategy to classify short texts semantically [4]. The first level includes the classification of text or short text in which the short text are labeled with crisp neutral and non-neutral labels. Then the second level classifier is used to act on the non-neutral short text and a "gradual membership" is assigned to each of the conceived classes. Such type of grades is then used by further phases of filtering process.

B. Classification based on Machine Learning:

The text categorization technique is a two level hierarchical classification process. The first level classifier performs a binary hard categorization that labels the messages as neutral and non-neutral. Then the second level task performs a finer-grained classification. It includes a soft partition of non-neutral messages and assigns them with "gradual membership" of each of the non-neutral classes.

For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. Naive Bayes is a simple

technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable. Despite their naive design and requires a small amount of training data to estimate the parameters necessary for classification.

The steps of the Naive Bayes Algorithm for classification purpose are as follows:

Step 1: The task is to classify the new documents as they arrive i.e. to which class label they belong based on the currently existing sentence.

Step 2: After formulating our prior probability, we are ready to classify the new document,

Step 3: Then we calculate the number of points in the packets that keep changing for each record/word.

Step 4: Final classification is produced by combining information from both sources i.e. the prior and the posterior probability.

C. Filtering Rules And Blacklist Management
Filtering Rules:

For defining the language for FR specification, many issues are considered .First may be the issue of message with different meaning and significance of who sends it. Hence, the FR should allow the user to block the or restrict the message creators .The type, depth and trust value is recognised by the creator specification. When a user profile does not hold the values for attributes submitted by FR, the filtering rules will be applied. Then the user will be asked whether or not to block the other user whom the owner does not want to be allowed to send messages to him/her.

Blacklist:

By applying the Blacklist mechanism we can keep away the messages from undesired creators. Thus this will help to decide that which users will be inserted in the blacklist. It also decides the time for which the blocked user will stay in the blacklist. Such rules are called as BL rules and are used to improve the stiffness.

If the user has a bad opinion about the specific user then the user can be banned for an uncertain period of time.

D. Steps for Filtering of the message(Algorithm) :

- Input: A message which is posted by the other user.(It can contain only good words, only bad words or good words along with bad words).
- Output: If the message contains good words then post message and if it contains bad words then it rejects the bad words and posts the filtered message.

Step 1: Start

Step 2: A User tries post the message in wall.

Step 3: Machine learning checks each word of the message.

Step 4: If (Words = = Good Words)

Step 5: Message is posted on the wall.

Step 6: Else if(Words = = Bad Words)

Step 7: Reject Bad Words using Blacklist and post the filtered message on the wall

Step 8 : End. [6].

5. GOALS AND SCOPE

The goal of the proposed system is to provide the user of Online Social Network(OSN) with the flexibility to have control on the messages posted on their user walls. This will offer support to the user's security. The user will be able to avoid unwanted messages on his/her user walls according to their preferences.

6. CONCLUSION

In this paper, we have presented a system to filter unwanted messages from OSN user walls. We basically focus on the ability of the user to have direct control on the messages posted on their user walls. For doing this we first apply classification techniques for classifying text messages. It makes the use of neural learning for this purpose. Next we apply filtering rules and the use of Blacklists is done so that user can insert the undesired message sender to the blacklist. Hence, in such ways we can provide better privacy and security to the OSN users.

7. Acknowledgements

This research work was support by Prof. S. Talekar, NDMVP'S KBT COE Nashik. We thank him for guiding us and providing insight which greatly assisted our research work. We also thank Prof. B. S. Tarle, H.O.D. NDMVP'S KBT COE Nashik for his constant motivation. We would also like to show our gratitude to Dr. Prof. Jayant T. Pattiwar, Principal NDMVP'S KBT COE Nashik and Management of NDMVP Samaj for providing all

necessary facilities and their constant encouragement and support.

8. References

[1] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, and Moreno Carullo "A System to Filter Unwanted Messages from OSN User Walls"- IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 2, FEBRUARY 2013.

[2] Sujapriya S, Immanual Gnana Durai, Dr. C.Kumar Charlie Paul, "Filtering Unwanted Messages from Online Social Networks (OSN) using Rule Based Technique ", IOSR-JCE, (e-ISSN: 2278-0661), (p-ISSN: 2278-8727), volume 16, Issue 1, Ver. 1, PP 66-70, Jan. 2014.

[3] J.Anishya Rose, A. Pravin, "Machine Learning Text Categorization in OSN to Filter Unwanted Messages", (IJCSIT) International Journal of Computer Science and Information Technologies, ISSN: 0975-9646,Vol. 5 (1), 640-643, 2014.

[4] Vikrant Sanghvi, Amol Nanaware, Divya Nadar, Chitra Bhole, "A System to Filter Unwanted Messages from OSN User Wall", International Journal of Research in Advent Technology, E-ISSN: 2321–9637, Volume 1, Issue 5, December 2013.

[5] Amruta Kachole, S. D. Jondhale , "Unwanted Message Filtering System from OSNs User's Wall Using Customizable Filtering Rules and Black list echniques". IJETAE, Volume 4, Issue 2, ISSN 2250-2459, ISO 9001:2008 Certified Journal, February 2014.