Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 6, June 2015, pg.92 – 98

RESEARCH ARTICLE

A Novel Technique to Filter Unwanted Messages from Online Social Network

Janavee Jambhulkar, Prof. Deepak Kapgate
Department of Computer Science and Engineering, GHRAET, Nagpur, India
Department of Computer Science and Engineering, GHRAET, Nagpur, India
Janavee16@gmail.com, deepakkapgate32@gmail.com

Abstract- Internet has a great impact on the life of the people in positive way. The use of internet has increased immensely. In present years Online Social Networks also evolved and plays an equivalent role. Online social networks are perfect for exchanging public opinions, views and ideas. Hence Online Social Networks should be extremely secure and should protect the individual's privacy. The Online Social Network provides the security measures but they were limited. While Socializing the user can access the profile of other members which are involved in social sites and even share data such as images, text, videos etc. One critical issue in user wall is to give users the capability to control the messages posted on their own personal space in order to avoid unwanted content to be displayed on their wall. In this paper, we apply filtering methods by using machine learning technique to filter the unwanted messages on the user walls.

Keywords-Internet, filtering, online social network, security, socialization

I. INTRODUCTION

Social networking sites are a part of daily life and they have brought inventive changes in communication between people. These sites provide diverse resources such as instant messages and email at single place. Miscellaneous type of data can be shared such as images, music, video etc., through the social networking sites. Social network allows user to connect to a variety of pages on the network that includes some useful sites like business, online shopping, marketing, e-commerce, education. Availability of these resources makes the communication easier and faster. Social networking sites help in developing connection with people, friends and relatives. Those People having similar professions can make groups like students, writers, lawyers, social workers, doctors etc.

An online social networking site is a platform to create social networks between people who share interests, activities or connections. An online social networking site consists of a profile of each user, links of the user, and a variation of additional services. Online Social networks are web-based services which allow individuals to create a public profile, to create a list of users with whom they can share and view the connections in the system.

An Information filtering is a composition that removes repetitive unwanted data from immense collection of information using automated and semi automated methods before the presentation of a human user. In order to execute this, the user's profile is compared to some associating characteristics. Information filtering is used here to manage the massive data from

the online social networks. Nowadays online social network provide restricted support to avoid unwanted messages on user walls. For example, Facebook permits users to manage which user is allowed to insert messages on their walls on the basis of relationship based filtering (i.e., friends, friends of friends, or defined groups of friends). However, content-based techniques are not used and therefore it is not possible to avoid undesired messages, such as religious ones, without taking into consideration about the user who posts them. Content based filtering is preferable for the short texts that occur in messages.

In this paper, our main aim is to analyze the classification technique and to design the system to filter the undesirable messages from OSN user wall. The aim of the present work is to suggest and experimentally estimate an automated system which is called as Filtered Wall (FW) that should be able to filter unwanted messages from OSN user walls. Machine Learning (ML) text categorization techniques are evolved to automatically assign with each short text message based on its content by using a set of categories.

In addition, the system will use a flexible language to demonstrate the filtering rules (FRs), with the help of it the users can decide what contents should be displayed on their walls. The FRs can be personalized according to the users need. Along with it there are user defined blacklists (BLs) which will temporary intercept users to post any type of message on user walls.

II. RELATED WORK

Recommender systems works in three main ways, the content-based filtering, Collaborative filtering, policy-based personalization

A. Content based filtering

Content based filtering which is also known as cognitive filtering, recommends items for a user based on the representation of previously evaluated items and information available from the content. The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document. a content-based filtering system 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 [1]. Selection of item is based on user interest.

The recommended systems previously use social filtering methods that consider recommendations on other users choice. On the contrary R. J. Mooney et.al describes a content-based book recommending system that takes advantage of machine-learning algorithm for text categorization and information mining. Thus, improve access to relevant information [3].

In the content based filtering, the systems is capable of learning from user's actions related to a particular content and use them for other content types which is the main advantage. Filtering concept is enforced to the Online Social Network user wall using rule based text categorization technique. The latest experiments emphasize complexities, efficiently as short text is brief, with a variety of misspellings, nonstandard conditions, and noise. Zelikovitz et.al tried to improve the classification of short text strings by developing a semi-supervised learning policy based on a combination of labelled training data and a secondary amount of unlabeled but related longer essays.[2] This declaration is inappropriate in our field in which short text messages are not part of long semantically associated documents. A different approach is intended by Bobicev et.al by adopting a statistical learning method that can perform well by avoiding the problem of error-prone quality construction [2].

B. Collaborative filtering

Danyel Fisher et.al presented a framework based on java, SWAMI (Shared Wisdom through the Amalgamation of Many Interpretations) for studying and structuring collaborative filtering systems. It consists of three components: a prediction engine, an evaluation system, and a visualization component. They verified comparison of three prediction algorithms: a traditional Pearson correlation-based method, support vector machines, and a new precise and scalable correlation-based method based on clustering technique. It was demonstrated that new pearson clustered correlation predictor go with current state of art methods, with benefit of scalable performance.[4]

C. Policy based personalization filtering

Policy based personalization has been useful in various Context. It acclimatizes a service in particular context as per the user defined policies.

In Twitter, communication policy can be defined between two communicating parties. It allocates a category to each tweet and exhibit only those tweet which are of concern to the user. In this situation, policy based personalization signify the ability of the user to filter out messages on wall according to filtering criteria specified by user. In contrast, Golbeck et.al proposes an application, named FilmTrust, which make use of OSN trust relationship and derivation information to personalize access to website. Though these types of system does not provide a policy layer for filtering by which the user can exploit the result of the classification to decide to which extent the unwanted information is filtered out. In contrast, filtering policy language allocate

the setting of FRs according to different criteria, which will consider the relationships of the wall owner with other OSN users as well as information on the user profile and output or results of the classification process. Furthermore, our system is accompanied by a flexible mechanism for BL management which provides opportunity of customization to the filtering process [10].

The work by Boykin et.al that presented an automated anti-spam tool that can recognize unsolicited e-mail, spam and messages related with known people of user. However, the strategy stated does not make use of ML content-based techniques [11].

Foltz et.al researched tested methods for predicting which Technical Memos (TMs) best match people's technical interests. This was totally based on previous feedback. There was no individual filtering.

D. Text representation

S. Dumais et.al evaluate the efficacy of five different automatic learning algorithms for text categorization in terms of learning speed, real-time classification speed, and classification accuracy and they came up with the conclusion that Linear Support Vector Machines (SVMs) are most accurate classifier, fastest to train, and quick to evaluate. They used SVMs for categorizing Web pages and email messages .They wish to extend their work by including the extra structural information about documents, as well as knowledge-based features for classification accuracy and automatically categorize items into hierarchical grouping structures [5]. R. E. Schapire et.al depict an implementation, called BoosTexter, the new boosting algorithms for text categorization and also compare its performance with a a range of other text-categorization algorithms on a different tasks.[7].

The merely social networking service present for providing filtering abilities to its users is MyWOT social networking service which gives its users the ability to: 1) rate resources on basis of four criteria: truthfulness, trader or vendor reliability, privacy, and safety of child 2) state preferences determining whether the browser should block access to specified resource, or should simply give a warning message according to the specified rating. Though there are some similarities, the method adopted by MyWOT is different from ours. It supports filtering criteria which are less flexible than those of Filtered Wall as they are only based on four criteria mentioned above. Furthermore, no automatic filtering method is provided to the end user [1].

III. PROPOSED SYSTEM

The aim is to propose and experimentally assess an automated system, Filtered Wall (FW) which enable to filter unwanted messages from OSN user walls. We make full use of Machine Learning (ML) text categorization which automatically assigns with each short text message a set of categories based on its content. Using a hierarchical two stage classification strategy, we insert the neural model. In the first stage, the RBFN categorizes short messages into Neutral and Non-neutral. In the second stage, Non-neutral messages are classified to produce gradual evaluations of appropriateness to each of the reviewed category. Besides classification facilities, the system provides a dominant rule layer exploiting a flexible and workable language to specify filtering rules (FRS), using these users can state what contents, will not be exhibit on their walls. FRS can support different filtering standards that can be combined and modified according to the user needs. FRS exploits user profiles, the output of the ML categorization procedure to state the filtering standards to be enforced as well as user relationships. In addition, the system calls for black lists (BLS) defined by users, lists of users that are temporarily blocked to post any kind of messages on a user wall and provide the spam.

IV. SHORT TEXT CLASSIFICATION

The main function of the proposed system is the content-based message filtering (CBMF) and short text classifier. In addition it supports the classification of message based up on the category set. On datasets with large documents like newswires corpora, already established techniques used for text classification work well but fails when the documents in the corpus are short. In this context, critical aspects are the definition of a set of characterizing and different features allowing the delineation of underlying concepts and the collection of a complete and consistent set of governed examples. From a ML point of view, we approach the task of short text categorization by defining a hierarchical two stage strategy assuming that it is better to identify and eliminate neutral sentences, then classify non neutral sentences. The first stage task is considered as a hard classification where short texts are labeled with crisp Neutral and Non-Neutral labels. The second stage soft classifier acts on the crisp set of non-neutral 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 successive phases of the filtering process.

V. FILTERING RULES AND BLACKLIST MANAGEMENT

In this section, we introduce the rule layer adopted for filtering unwanted messages. We start by describing FRs, then we illustrate the use of BLs.

A. Filtering Rules

In defining the language for FRs specification, we examine three main issues that, should affect a message filtering decision. First of all, the same message may have non-identical meanings and closely connected based on who writes it. As a result, FRs should allow users to impose constraints on message creators. FR applied creators can be selected on the basis of different standards; one of the most promising is by forcing conditions on their profile's attributes. For example, possible to specify rules applying only to creators with a given religious or political view or young creators. Given the social network scenario, creators may also be identified by misusing information on their social graph. This indicate to state conditions on depth, type and trust values of the relationships creators should be involved in order to apply them the defined rules. All these issues are formalized by the notion of creator specification.

Online setup assistant for FRs thresholds:

As mentioned in the previous section, we deal with the problem of setting thresholds to filter rules, by formulating and implementing within FW, an Online Setup Assistant (OSA) procedure. OSA confers the user with a set of messages which has to select from the dataset. For each message, the user tells the system his decision to reject or accept the message. 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. Such messages are selected according to the defined process. A certain amount of non neutral messages taken from a portion of the dataset, not belonging to the training or test sets, are arranged by the ML in order to have the second stage class membership values for each message.

B. Blacklists

We make use of a Blacklists mechanism to avoid messages from unwanted creators. BL is administered directly by the system, which is able to:
(1) Identify who are the users to be placed in the Blacklists,
(2) Block all the messages
(3) Decide when user withholding in the BL is finished.

The Blacklist mechanism has to be directed with some rules in order to make the system able to automatically perform these tasks.
In particular, these rules aim to specify
(a) How the Blacklist (BL) mechanism has to identify users to be banned and
(b) For how long the banned users have to be withhold in the BL, i.e., the withholding or retention time.

VI. FILTERED WALL ARCHITECTURE

The architecture of OSN is a three-tire structure of three layers. These three layers are

- Social Network Manager(SNM)
- Social Network Application(SNA)
- Graphical User Interface(GUI)

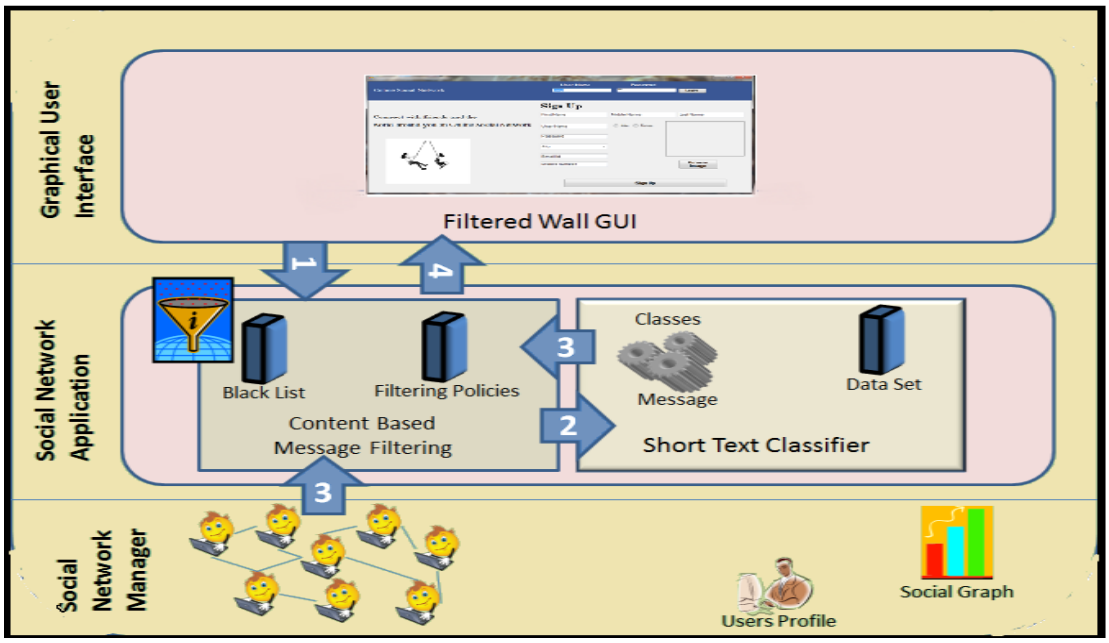

Figure 1.filtered wall architecture

In general, the architecture in support of OSN services is a three-tier structure. Additionally, some OSNs provide an additional layer allowing the support of external Social Network Applications. Finally, the supported SNAs may require an additional layer for their needed graphical user interfaces (GUIs).

A. Social Network Manager

The first layer is Social Network Manager layer which provides the necessary OSN functions such as profile and relationship administration. It also maintains all the data of the user profile. After administrating and maintaining all users data will be provided for second layer for applying Filtering Rules and Black lists.

B. Social Network Application

In second layer, Content Based Message Filtering and Short Text Classifier is composed. This is important layer for the message categorization as per CBMF filters. Blacklist is managed for the user who frequently sends bad words in message.

C. *Graphical User Interface*

The third layer consists of Graphical User Interface to the user who wants to post his messages as a input. Here, Filtering Rules are used to filter the unwanted messages and provide Black list for the users who are temporally banned to post messages on user's wall. The GUI consists of Filtered Wall where the user is able to see his desired messages.

1. After entering the private wall of one of users, the user tries to post a message captured by Filtered wall

2. A ML-based text classifier extracts data from the message content.

3. Filtered wall uses data given by the classifier, along with data extracted from the users profiles, to implement the filtering rules and blacklists techniques.

4. Considering the result of the previous step, message will be filtered.

VII. RESULT ANALYSIS

The user interface login form is designed for new user to register by filling the details in the form. The user who is already registered can login by entering the details such as username and password. This form is connected to the database so that the data entered while registering is stored in database. When the registered user enters the correct login details which gets matched with the database, it will display a message showing login successful. Whenever the user tries to post unwanted message on the wall of another user, the word or the message will be filtered and it will display a message. The message will show the unwanted text which the user does not want it to be posted on his wall.

A. Verifying the Detection of Bad words

Different number of words will lead to different predictions. The Experiment is done with different number of words and the number of detected bad words is analyzed. The experiment is carried out with various numbers of words such as 50,100,150,200 and is plotted on graph for different cases. The resultant graph is shown below.

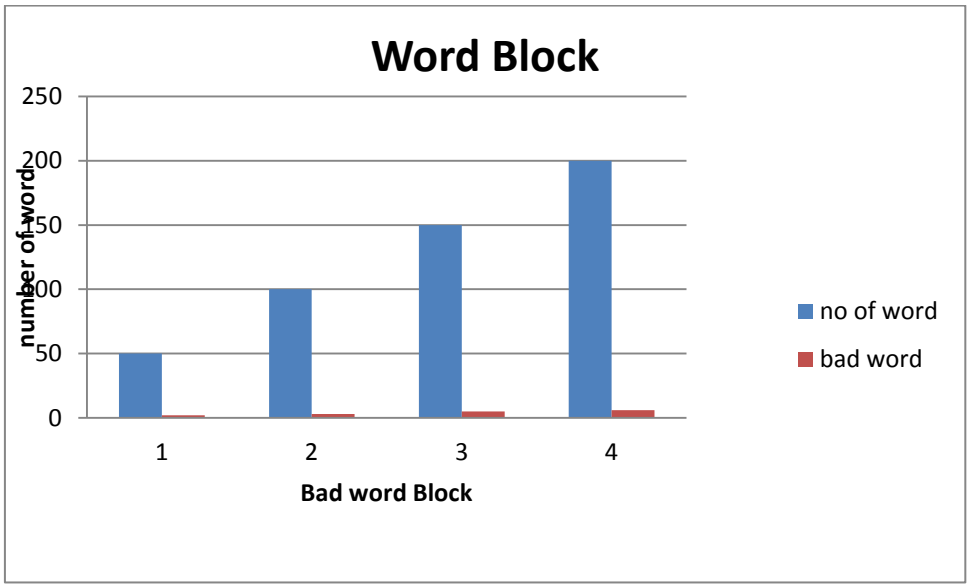

Figure V-8: Graph Showing bad word detection

In above graph, abscissa represents different bad word block, and vertical axis represents the number of words.

VIII. FUTURE SCOPE

The future work may include Image Filtering Techniques. In our system we can only filter the text messages. So Image filtering can be tried in future system. Furthermore, the flexibility of the system in terms of filtering options is enhanced through the management of BLs. We can address this problem by investigating the use of on-line learning paradigms able to include label feedbacks from users in future work. We can use the same concept in other social forums. It never block all social medias instead we can apply this on selected space.

IX. CONCLUSION

In this paper, we have proposed a system to filter undesired messages from OSN walls. The system exploits a ML soft text classifier to impose customizable content-dependent FRS. Besides, the flexibility of the system in terms of filtering criteria is enhanced through the management of BLs. This work is the further step of a wider project. The early motivating results we have obtained on the classification procedure instigate us to continue with other work that will aim to enhance the quality of classification. In particular, future plans contemplate a extensive investigation on two interdependent tasks. The current batch learning strategy, based on the preparatory collection of the entire set of labeled data from experts, allowed an accurate

experimental evaluation but needs to be developed to include new operational requirements The development of a GUI and a set of related tools make easier BL and FR specification is also we plan to investigate, since usability is a key requirement for such kind of applications.

REFERENCES

[1] Marco Vanetti et. Al "A *System to Filter Unwanted Messages from OSN User Walls"* University of Insubria, Italy IEEE Transactions On Knowledge And Data Engineering Vol:25 Year 2013

[2] Adomavicius et. Al, "*Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions*," IEEE Transaction on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[3] F. Sebastiani, "*Machine learning in automated text categorization*," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.

[4] N. J. Belkin et. Al, "*Information filtering and information retrieval: Two sides of the same coin?*" Communications of the ACM, vol. 35, no. 12, pp. 29–38, 1992

[5] A. D. Swami et. Al, "A *Text Based Filtering System for OSN User Walls*" International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 2, February 2014

[6] Amruta Kachole et.Al, "*Unwanted Message Filtering System from OSNs User's Wall Using Customizable Filtering Rules and Black list Techniques"* IJETAE Volume 4, Issue 2, February 2014.

[7] 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.

[8] R. J. Mooney and L. Roy, "Content-*based book recommending using learning for text categorization*," in Proceedings of the Fifth ACM Conference on Digital Libraries. New York: ACM Press, 2000, pp. 195-204.

[9] M.Carullo, E.Binaghi, and I. Gallo, "An *Online Document Clustering Technique for short Web contents*," Pattern Recognition Letters, vol.30, pp.870-876, July 2009.

[10] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari,"Content-*based filtering in on-line social networks*," in Proceedings of ECML/PKDD Workshop on Privacy and Security issues in Data Mining and Machine Learning (PSDML 2010), 2010.

[11] R. E. Schapire and Y. Singer, "*Boostexter: a boosting-based system for text categorization*," Machine Learning, vol. 39, no. 2/3, pp. 135–168, 2000.

[12] P. W. Foltz and S. T. Dumais, "*Personalized information delivery: An analysis of information filtering methods*," Communications of the ACM, vol. 35, no. 12, pp. 51-60, 1992.

