

# Survey on Content-Based Message Filtering in On-Line Social Networks

Jignasha H. Patel<sup>1</sup> Prof. Jignasa N. Patel<sup>2</sup>

<sup>1</sup>Student <sup>2</sup>Assistant Professor

<sup>1</sup>Department of Information Technology <sup>2</sup>Department of Computer & Information Technology

<sup>1,2</sup>Shri S'ad Vidya Mandal Institute of Technology Bharuch, India

**Abstract**— Online social networks (OSNs) are very popular and one of the most communicating medium to share, communicate and exchange many types of information like text, image audio, video etc. But there is Major issue in OSN is to preventing security in posting unwanted messages. OSNs do not provide support to prevent unwanted messages on user walls. So there is a need to develop more security mechanisms in OSNs. In this paper we try to solve this problem and propose a system allowing OSNs users to have a direct control on the messages posted on their walls using Black list instead of Block list. Short text classification and content based message filtering, Machine Learning, Filtering rules and Blacklist these five modules are used for message filtering.

**Key words:** Online Social Networks, Content-Based Filtering, Short Text Classification, Black List

## I. INTRODUCTION

Internet is made up of several interlinked network which are related to Business, Government, Academics, etc. Millions of computers are connected to the internet and millions of people share the information, opinions, and important data with each other. If user wants information related to current news affairs then user must have to enter the URL (Uniform Resource Locator) of the web page. After that the web page of the current news will open. And user gets the information which he wants. Web sites are made up of multiple web pages. Web page is nothing but the collection of HTML (Hyper Text Markup Language) documents which is basically used to design the web page. There are other languages used to develop the web page such as Perl, Ajax, PHP, ASP etc. There are two types of web pages one is Static page and other one is Dynamic page. Static Page gives the same information after loading each time. But the Dynamic page gives new information after loading each time. The content of Static page doesn't change. But the content of the Dynamic page changes every time [1].

The one of the most important part of the internet is the Online Social Network (OSN) sites. These sites are also used to exchange the information, contents, data like audio, video clips, opinions related to particular topic, etc. Face Book, Twitter, Linked in, etc are the latest Social Networking Sites. On Face book if one user posts his picture then other user comments on that picture or they post a comment in the form of the statement. In this way people can share the information. But there are some disadvantages of using these Social sites and Web Pages. On social networking sites most of people post the useless comments as well as pictures. This can cause harm to other user. And also there are lots of unwanted content, texts, images and advertisements display on the web page. So to avoid this, we are going to remove the useless comments from the web page [1].

Using online social networks people to get together and share information in advanced ways. But OSN provide

least support to avoid unwanted messages from the wall. Information filtering techniques can be used to automatically control the unwanted messages posted in their wall. The key limitation of existing system is it just blocks the user who post unwanted message instead of content based message. So this paper is focus on content based filtering in OSN using machine learning techniques [2].

## II. FILTERED WALL ARCHITECTURE

The architecture carry OSN services are a three-tier structure which shown in Fig. 1. Filtered wall is used to filter messages that are useless, from Online Social Network user wall. Machine Learning is used for text categorization. The Short Text Classifier is used to extract and select the set of characterizing feature. Author developed the Filtered Wall architecture. Also Facebook application prototype, DicomFW is used. This application is work for wall only and not for group. By using this application messages are filtered from user wall [1].

The architecture to support of online social network services comprises of three major components (Fig 1): Social Network Manager, Short text classification and Content Based Filtering. Social Network Manager (SNM) offers the basic online social networks functionalities such as profile management, relationship management etc. Short Text Classification is employed to classify the incoming post messages. Content Based filtering offers the support for message filtration. Specifically, users interact with the system via a GUI to set up and manage their FRs/ BLs.

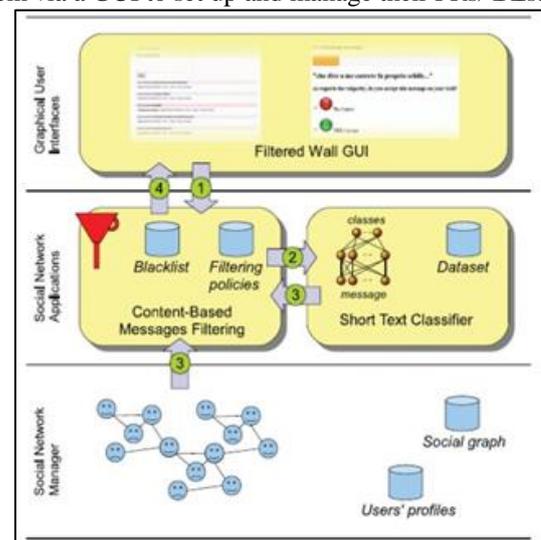


Fig. 1: Filtered wall architecture [4].

Moreover, the GUI provides users with a Filter Wall, that is, a partition where only messages that are allowed according to their FRs/BLs are published. As graphically shown in Figure 1, the path followed by a message, from its writing to the feasible final magazine can be summarized as follows:

- 1) Subsequent to incoming the private wall of one of his/her contacts, the user tries to post a message, which is intercepted by the Filtered Wall.
- 2) A Machine Learning-based text classifier extracts metadata from the content of the message.
- 3) Filtered Wall uses metadata provided by the classifier, together with data extracted from the social chart and users' profiles, to impose the filtering and Black List rules.
- 4) Depending on the result of the prior step, the message will be published or filtered by Filtered Wall.

The core components of the proposed system are the Content- Based Messages Filtering (CBMF) and Short Text Classifier modules [1].

### III. CONTENT-BASED FILTERING

Message filtering systems is typically designed to organize the data and dynamically generated information dispatched from the sender information producer to the user and the received information should likely satisfy the requirements of him/her. In the content-based filtering method, every user is assumed as independent. So, a content based filtering system selects the data based on the correlation between the user preferences and the content. This is completely differing to the mutual filtering system, which selects the items based on the correlation between the people with the similar user preferences [2].

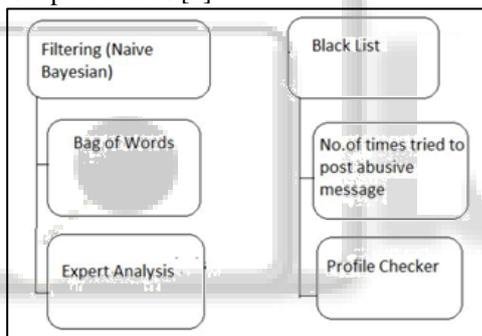


Fig. 2: Filtered wall architecture [2].

OSN provide a Block List in which if user doesn't want to share their ideas to particular friend or if the user doesn't wish to make known their details to their specific friends, then they can block permanently by this feature. But this feature not used to remove some unwanted post from their walls. Information Filtering is the best solutions for this issue. Content-Based Filtering system is used to classify the data and dynamically generated information and this system is similar to text classification and based on machine learning techniques. Most of the text filtering method by the Machine Learning has been applied to the long-form text [2].

Using 'Block list' user permanently block any friend or relative that's why its effect on relationship so overcome this problem by 'Black List' in which when the person continuously posts (more than 3 times) hateful messages on user's wall then users are blocked temporarily for few days [2].

#### A. Blacklist

The main goal of Blacklist Mechanism will keep away messages from undesired creators. BL are handled undeviating by the system. This will capable to make a decision the users to be inserted in the blacklist. And it also

decides the user preservation in the BL will get over [1][12]. OSN provide a Block List in which if user doesn't want to share their ideas to particular friend or if the user doesn't wish to make known their details to their specific friends, then they can block permanently by this feature. But this feature not used to remove some unwanted post from their walls. Information Filtering is the best solutions for this issue. Content-Based Filtering system is used to classify the data and dynamically generated information and this system is similar to text classification and based on machine learning techniques. Most of the text filtering method by the Machine Learning has been applied to the long-form text. Using 'Block list' user permanently block any friend or relative that's why its effect on relationship so overcome this problem by 'Black List' in which when the person continuously posts (more than 3 times) hateful messages on user's wall then users are blocked temporarily for few days [3].

### IV. SHORT TEXT CLASSIFIER

Automatic determination of the text relativity according to the real content under the given category topic is known as Text classification. In mathematics term, text classification is a map and map is the unspecified text to the given categories. And the map can be one-to-one mapping and it also can be a one-to- many mapping, because one text may be related to several categories. This can be demonstrated in a mathematical rule:

$$\emptyset = S \rightarrow C \dots \dots \dots (1)$$

Where, S is the text set to be classified, and C is the category set for the category topics [4].

Main goal of Short Text Classifier is classify the messages related to a set of categories. So, a classifier is set together to extract and select the selective features of the short text message. The task of short texts process comprised with two main phases Text representation and Machine Learning based classification [2].

In first phase, Vector Space Model is used to represent the short text message. VSM will indicate the text in suitable layout to extract its individual feature. In VSM, a short text message is represented as a vector of weights,

$$D_j = w^1_j, \dots, w^T_j \dots \dots \dots (2)$$

Where, T is the set of terms that occur at least once in at least one document of the set of document.  $T_r$  and  $w_{kj} \in [0; 1]$  indicate how much term  $t_k$  contributes to the semantics of document  $d_j$  [4]. In the Bag of Word (BOW) illustration, terms are recognized with words. Where non-binary weighting, the weight  $w_{kj}$  of term  $t_k$  in document  $d_j$  is computed according to the Standard Term Frequency - Inverse Document Frequency (TFIDF) weighting function is defined [5].

In second phase, a neural network classifier is sort out the received message. Neural network categorizes the short message into the appropriate class automatically; these messages are neutral or non neutral messages. Non neutral messages are additional to analyze to establish the correctness to each category [2].

### V. SYSTEM ARCHITECTURE

The user to get into the Online Social Network has to first register his/her details. Those details will be store into the

database. Then user can login with the registered username and password. The details entered by the user then will be verified with the data in the database. If the username and password matches completely with that of the registered one, then the user will be directed to the home page of the OSN. Otherwise, it will report a warning message to enter a valid username and password. There are many exiting features in OSN .The user can make friends and say their thoughts and views. The relationship between different people (like friend, family, etc.) and their belief percentage will also be stacked in the database [3].

The present work focuses mainly on the working of filtered wall which is briefly depicted in the below figure. First if the any user posts/comments any message then it will be channelized to the filtered wall, hen pops the functioning of two fabulous techniques namely “Text Mining” and “Filtering”. The turnout of text mining leads to the classification of neutral (risk-free) and non-neutral (rude/vulgar) messages. Fig 3 explains the entire working of the system. Then the non-neutral messages will be handled by various methods like expert analysis, filtering rules and by Naïve- Bayesian classifier. Filtering rules is comprised of antecedents (LHS) that specifies conditions for the rule and consequents (RHS) that specifies the action that has to be executed for that particular condition [3].

Eg: (Document=Contains word) ^ (Expert = Worst)

--> Non-neutral

Expert analysis has to do with the conclusion of the third party whether a particular word is neutral or non-neutral. Then the final filtering will be restricted by the Naïve-Bayesian classifier.

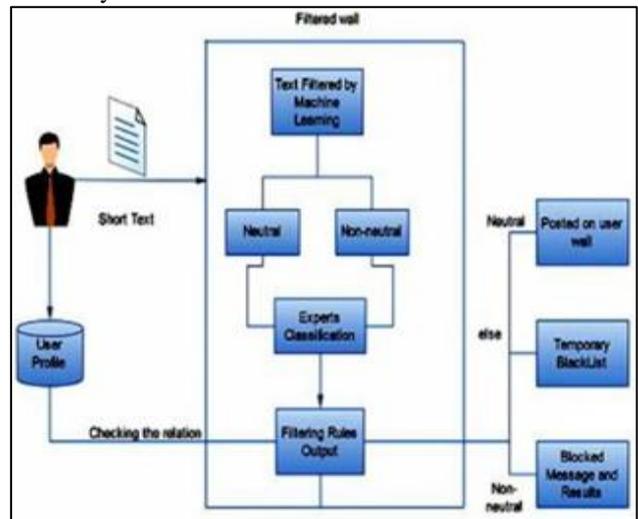


Fig. 3: Flow of the System [3].

## VI. LITERATURE SURVEY

Sr. No	Publication/ Year	Title	Review / Techniques	Advantages	Disadvantages/ Future Work
1.	IEEE-2014	Content Based Filtering in Online Social Network using Inference Algorithm [2].	The key limitation of existing system is it just blocks the user who post unwanted message instead of content based message. So this paper is focus on content based filtering in OSN using machine learning techniques. <b>Techniques</b> <ul style="list-style-type: none"> <li>Short Text Classifier.</li> <li>Content-Based Messages Filtering (CBMF).</li> <li>The forward - chaining inference algorithm.</li> </ul>	<ul style="list-style-type: none"> <li>Overcome the key limitation of the existing OSN.</li> <li>OSN cannot filter the incoming posts based on the content.</li> <li>Information is changing dynamically which can be handled by Inference Rule.</li> <li>Flexibility in management of black lists.</li> </ul>	<ul style="list-style-type: none"> <li>Propose to develop similar techniques to infer BL rules.</li> </ul>
2.	ELSEVIER - 2015	Automated Message Filtering System in Online Social Network [3].	This paper proposes a filtered wall to permeate offensive messages using rule based and text classification techniques in OSN. <b>Techniques</b> <ul style="list-style-type: none"> <li>Content-Based Filtering.</li> <li>Filtering Mechanism and Black List.</li> <li>Naive Bayes algorithm.</li> <li>Bayes Theorem</li> </ul>	<ul style="list-style-type: none"> <li>Apply Naive-Bayesian techniques are very easy to implement and an excellent performance with very less complexity.</li> <li>It is a very simple approach to provide an effective efficiency.</li> </ul>	<ul style="list-style-type: none"> <li>The limitation of this paper is that it cannot filter abusive images as well as videos.</li> <li>The future work concentrates on image and video filtering.</li> </ul>
3.	SPRINGER - 2011	Content-Based Filtering in On-Line Social Networks [5].	This paper is focus on filtering criteria and a Machine Learning based soft classifier in Content-Based Filtering. <b>Techniques</b> <ul style="list-style-type: none"> <li>Short Text Classifier</li> <li>Content-Based Filtering</li> </ul>	<ul style="list-style-type: none"> <li>Very good effectiveness, flexibility to changes in the domain and portability in different applications.</li> </ul>	<ul style="list-style-type: none"> <li>Difficulties arise in finding an appropriate set of features by which to represent short, grammatically poorly formed sentences.</li> </ul>

			with Blacklist.		<ul style="list-style-type: none"> <li>- As future work, intend to exploit similar techniques to infer BL and filtering rules.</li> </ul>
4.	IEEE - 2015	Notifying And Filtering Undesirable messages From Online Social Network (OSN) [6].	<ul style="list-style-type: none"> <li>- This paper defines an integrated system for filtering the unwanted results in the users OSN walls.</li> </ul> <p><b>Techniques</b></p> <ul style="list-style-type: none"> <li>- Machine Learning.</li> <li>- Filtering Rules.</li> <li>- Blacklist.</li> <li>- Short Text Classification.</li> </ul>	<ul style="list-style-type: none"> <li>- This system permits the users of OSN to take a directly supervision on the messages which are post on the OSN user's walls.</li> <li>- The obscenity of the users has been prevented.</li> </ul>	<ul style="list-style-type: none"> <li>- Using this proposed system the tractability of the method is improved in the factors of cleaning.</li> </ul>
5.	ELSEVIER - 2014	A user-centered and group-based approach for social data filtering and sharing [7].	<ul style="list-style-type: none"> <li>- In this paper, a user-centered and group-based approach for social data filtering and sharing and focus on two problem of SNSs, Information Overload and "Walled garden" Problem.</li> </ul> <p><b>Techniques</b></p> <ul style="list-style-type: none"> <li>- Social and Collective System (Soco-Sys).</li> </ul>	<ul style="list-style-type: none"> <li>- Overcome the problem of SNSs, Information overload and "Walled garden" Problem.</li> <li>- Proposed system more flexible and suitable to many users.</li> </ul>	<ul style="list-style-type: none"> <li>- Some other advanced features can also be added into the system like group analysis.</li> <li>- Try to add additional functionalities like Duplicated information hiding, sharing with SNSs and Notification.</li> </ul>
6.	IEEE - 2015	Towards Automatic Real Time Identification of Malicious Posts on Facebook [8].	<ul style="list-style-type: none"> <li>- This paper address the problem of automatic real time detection of malicious content generated in Facebook.</li> </ul> <p><b>Techniques</b></p> <ul style="list-style-type: none"> <li>- REST API.</li> </ul>	<ul style="list-style-type: none"> <li>- Existing clustering based spam detection techniques are able to detect less than half the number of malicious posts as compared to REST API.</li> <li>- The REST API is publicly accessible and a browser plug-in, that can be used to identify malicious content on Face book in real time.</li> </ul>	<ul style="list-style-type: none"> <li>- This paper does not state that our dataset is representative of the entire Facebook population.</li> <li>- The meritocratic nature of WOT makes it far more difficult for spammers to abuse.</li> <li>- In future, improve the performance and usability of browser plug-in.</li> </ul>
7.	IEEE - 2013	A System To Filter Unsolicited Texts From Social Learning Networks [9].	<ul style="list-style-type: none"> <li>- This paper is focus on Information Filtering which is categorized a flow of information and Parsing of the Text which is classification of texts to filter some unwanted message on OSNs.</li> </ul> <p><b>Techniques</b></p> <ul style="list-style-type: none"> <li>- Content-Based Filtering with Blacklist.</li> <li>- Pattern Matching.</li> </ul>	<ul style="list-style-type: none"> <li>- Sample learning website and blacklisted words table made the filter easier.</li> <li>- For blog comment and feedback comments are concerned this proposed system has worked successfully.</li> </ul>	<ul style="list-style-type: none"> <li>- Time-complexity is high.</li> <li>- In future, try to work on reduce time-complexity and also try to plans to incorporate strategies and technique.</li> </ul>
8.	IEEE - 2013	A System to Filter Unwanted Messages from OSN User Walls [1].	<ul style="list-style-type: none"> <li>- The main aim of this paper is to avoid useless data. They have used the concept of filtered wall which is used to filter messages that are useless, from Online Social Network user wall.</li> </ul>	<ul style="list-style-type: none"> <li>- This system allow OSN user to have direct control on the messages posted on their wall.</li> <li>- Content-Based Filtering using ML filtering are a very</li> </ul>	<ul style="list-style-type: none"> <li>- It is work for static Social Networking wall. It is not work for whole web page.</li> <li>- As future work, we intend to exploit similar techniques to infer BL rules and FRs.</li> </ul>

			<p><b>Techniques</b></p> <ul style="list-style-type: none"> <li>– Rule Based and Machine Learning based Soft Classifier.</li> </ul>	<p>good effectiveness, flexibility to changes in the domain and portability in different applications.</p>	
9.	SPRINGER -2016	Learning to Filter User Explicit Intents in Online Vietnamese Social Media Texts [10].	<ul style="list-style-type: none"> <li>– This paper present a machine learning approach to analyze users' posts and comments on online social media to filter posts or comments containing user plans or intents.</li> </ul> <p><b>Techniques</b></p> <ul style="list-style-type: none"> <li>– User intent filtering.</li> </ul>	<ul style="list-style-type: none"> <li>– Built a classification model based on the maximum entropy method to classify text posts/comments on online social media to determine which ones carry user explicit intents.</li> </ul>	<ul style="list-style-type: none"> <li>– Need to add better and higher level features for effectively highly ambiguous text posts/comments.</li> </ul>
10.	IEEE-2016	Effective and Adaptive Technological Solution to block Spam E-mails [11].	<ul style="list-style-type: none"> <li>– This paper focuses on the problem of spam E-mail and proposed new technology.</li> </ul> <p><b>Techniques</b></p> <ul style="list-style-type: none"> <li>– Machine learning based classifier (MLC).</li> <li>– Semantic similarity with edge based classifier (SSC).</li> <li>– Origin based filter (OBF).</li> <li>– Content based filter (CBF).</li> </ul>	<ul style="list-style-type: none"> <li>– Combination of adaptive OBF and adaptive CBF technique solve the problem of spam E-mail by proposing the technological solution.</li> <li>– It useful even with changing pattern of spamming since it is adaptive and effective in nature.</li> </ul>	

Table 1: A Survey on Message Filtering

## VII. CONCLUSION

Use of Internet and Online Social Network is increasing day by day. But lots of unwanted messages are also occurs on the net. These comments can cause harm or can create misunderstanding between peoples. Thus in this paper, the filtration of text or comments are done using short text classifier and content based message filtering. The user is blocked in black list for temporary. Overall system architecture also defines here. Due to this unwanted messages are filtered out and it will increase the efficiency of OSN user wall. In Future Work the abusive pictures or video can be deleted or blocked.

## REFERENCES

- [1] Macro Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati and Moreno Carullo "A System to Filter Unwanted Messages from OSN User Wall", Transaction on Knowledge and data Engineering, IEEE February. 2013
- [2] N.Thilagavathi, R.Taarika, "Content Based Filtering in Online Social Network using Inference Algorithm", IEEE 2014.
- [3] Subramaniaswamy Va, Logesh Rb,Vijayakumar Vc, and Indragandhi Vd. "Automated Message Filtering System in Online Social Network", Elsevier , 2015.
- [4] Zhou Faguo, Yang Bingru, Zhang Fan, Yu Xingang, "Research on Short Text Classification Algorithm Based on Statistics and Rules", IEEE – 2010.
- [5] Marco Vanetti, Elisabetta Binaghi, Barbara Carminati, Moreno Carullo, and Elena Ferrari, "Content-Based Filtering in On-Line Social Networks", Springer – 2011.
- [6] V.Vijeya Kaveri, S.Gopal, "Notifying and Filtering Undesirable Messages Fromonline Social Network (Osn)", IEEE – 2015.
- [7] Xuan Truong Vu, Marie-Hélène Abel, Pierre Morizet-Mahoudeaux, "A user-centered and group-based approach for social data filtering and sharing", ELSEVIER – 2014.
- [8] Prateek Dewan, Ponnuram Kumaraguru, "Towards Automatic Real Time Identification of Malicious Posts on Facebook", IEEE – 2015.
- [9] Sunil Yadav, S Das, D Rudrapal, "A System To Filter Unsolicited Texts From Social Learning Networks" IEEE – 2013.
- [10] Thai-Le Luong(B), Thi-Hanh Tran, Quoc-Tuan Truong, Thi-Minh-Ngoc Truong, Thi-Thu Phi, and Xuan-Hieu Phan, "Learning to Filter User Explicit Intents in Online Vietnamese Social Media Texts", SPRINGER – 2016.
- [11] Ajay U. Surwade, M.P.Patil, S.R. Kolhe, "Effective and Adaptive Technological Solution to block Spam E-mails", IEEE – 2016.
- [12] Gediminas Adomavicious, Alexander Tuzhilin, "Towards the next generation of recommender system: A survey of the state of the art and possible Extension", IEEE – 2005.