RESEARCH ARTICLE OPEN ACCESS

Unwanted Message Filtration from User's Wall Using Naive Bayes

Punam K. Mohare [1], Prof. (Dr). S A Itkar [2]
Department of Computer Science and Engineering
P.E.S. MCOE
Savitribai Phule Pune University
Maharashtra - India

ABSTRACT
Nowadays online social network have become an essential part of daily life of human for communication, sending the messages, images and videos. Up to now online social network provide the very little support to prevent unwanted messages post users wall. The aim of present work to overcome the disadvantages proposed the filtered wall that automatically blocked the messages not the users and provide the more security and also provide the control the messages posted by the user on their wall. OSN provide very important role in daily human life. Naïve Bayes classifier is used for classification, it gives better result as compare to another classification technique. Machine Learning Text Categorization is also used to categorize the short text messages. Only unwanted messages will be blocked not the user to avoid this disadvantage blacklist mechanism is used. Blacklist mechanism is used to determine which user should be inserted in BL for how many period of time and when the user will be completed his blocking time. Machine learning technique is used to automatically assign the messages to its respective categories.

Keywords:- On-line Social Networks, Information Filtering, Short Text classification, Naïve Bayes.

I. INTRODUCTION

Online social network are todays one of the popular and important medium to communicate and share the text messages, images and videos. An OSN is a web-based service that allows individuals to:

1. Construct a public or semi-public problem within the service,
2. Articulate a list of other users with whom they share a connection,
3. View and traverse their list of connections and those made by others within the Service.

Information and communications technologies plays important and essential role in daily human life and in networked security. Nowadays security is important in the different communications technologies, particularly in online social networked. Information filtering can give to user's ability to automatically control the messages written on their own walls and filtering unwanted messages. The aim of present work to overcome the disadvantages proposed the filtered wall that automatically blocked the messages not the users and provide the more security and also provide the control the messages posted by the user on their wall. OSN provide very important role in daily human life. Naïve Bayes classifier is used for classification, it gives better result as compare to another classification technique. Machine Learning Text Categorization is also used to categorize the short text messages. Only unwanted messages will be blocked not the user to avoid this disadvantage blacklist mechanism is used. Blacklist mechanism is used to determine which user should be inserted in BL for how many period of time and when the user will be completed his blocking time. Machine learning technique is used to automatically assign the messages to its respective categories.

II. LIRERATURE REVIEW

In the paper of [1], Marco vanity proposed a system allowing 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 users to customize the filtering criteria to be applied to their walls, and a Machine Learning based soft classifier i.e. RBFN automatically labeling messages in support of content-based filtering. In the paper of [2], Document classification is a growing interest in the research of text mining. Correctly identifying the documents into particular category is still presenting challenge because of large and vast amount of features in the dataset. In regards to the existing classifying approaches, Naïve Bayes is potentially good at serving as a document classification model due to its simplicity. The aim of this paper is to highlight the performance of employing Naïve Bayes in document classification. Results show that Naïve Bayes is the best

classifiers against several common classifiers (such as decision tree, neural network, and support vector machines) in term of accuracy and computational efficiency.

In the paper of [3], a system that implements a content-based message filtering service for Online Social Networks (OSNs). They used Naïve Bayes algorithm for classification. It gives better result as compare to the other classifier like SVM, RBFN etc. Naïve Bayes advantages are the having low memory capacity and it is not potential sensitivity to input parameters. In the paper of [4], an overview done on classification process, document can be classified into three categorize such as unsupervised, supervised and semi supervised. Text classification process is assigning the text messages to the predefined classes automatically. In the paper of filtering wall am online social network [5], proposes a system that allows OSN users to directly control messages posted on their wall. This is achieved through flexible rule based system in support of a machine learning based short text classifier, user defined BL mechanism and an image comparison technique. In the paper of [6], system allows OSN user to have direct stop the unwanted messages posted on their wall. Content based filtering technique is used because messages are filtered by using content of the messages. Machine learning is used to automatically assigned the messages to the respective set of categorize. A machine learning based approaches [7] machine learning technique used for the classification purposed. In this had the comparison between the web feature (NN-WEB and SVM-WEB) and the keyword-based SVM approach (SVM-WORD) and the lexicon-based approach (LEXICON). Advantages of the web feature technique is that having performance with higher effectiveness than the SVM-WORD and LEXICON. In the paper of [8], an overview is done on three recommender system such as content based filtering in which user will prefer the item that similar user preferred in the past, collaborative filtering which based on the user preference and hybrid which is combination of both the recommender system.

III. ARCHITECTURAL VIEW

There are three layer of the filtered wall architecture (Fig.1) Out of three layers, first layer is Social Network Manager (SNM) which provides the basic OSN functionalities and maintains data regarding to the user profile. Software Social Network is the second and important layer for message categorization and it consists of CBMF and Short Text Classifier and also black list is maintained for the users. Third layer is Graphical User Interface consist filter wall and provides user wall on which user wants to post. The present work established in the SNAs and GUIs layer.

Processing Steps:

Step1: After entering the private wall of one of his/her contacts, the user tries to post a message, which is intercepted by FW.

Step2: Then Pre-processing take place and then A ML-based text classifier (Naïve Bayes) extracts metadata from the content of the message.

Step3: FW uses metadata provided by the classifier, together with data extracted from the social graph and users profiles, to enforce the filtering and BL rules.

Step4: Depending on the result of the previous step, the message will be published or filtered by FW.

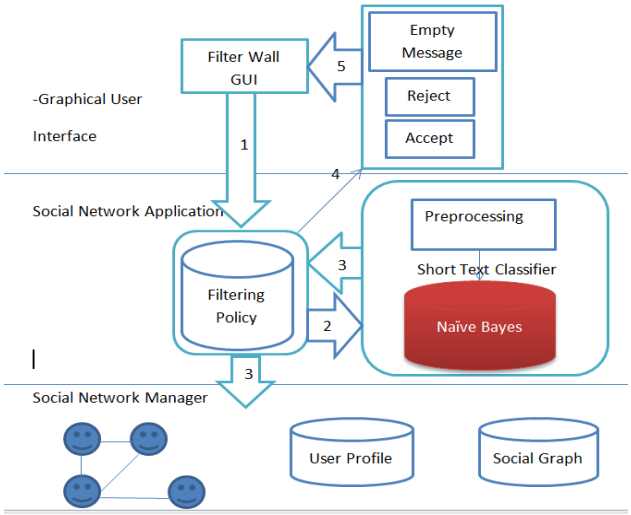

Fig. 1 System Architectural Diagram

IV. ALGORITHMIC ASPECTS

Radial Basic Function Network:

Short text classification is used as a classifier and is a two level hierarchical classification process. In this first level messages are classified in to neutral and non-neutral. Then second non-neutral classified into the respective categories is the second level. In this paper Radial Basic Function Network is used for the classification process. RBFN have a single hidden layer and RBFN commonly used as Gaussian functions. RBFN important advantage is that classification function is non-linear; drawback are the potential overtraining sensitivity to input parameter and potential overtraining sensitivity. Following steps for the RBFN:

Input: The message.

Output: Relate term with training category.

Step1: Input the message.

Step2: Filter the message as Punctuations characters, Exclamation marks, Question marks, Capital words.
Step3: The RBFN classified text messages into Neutral and Non-Neutral.
Step4: Calculate TFIDF of each term. Tf-idf formula
Step5: From each message mi extract the vector of features xi. The two sets training set and testing set are then transformed Tr and testing set as Te.
Step6: Then Test set Te trained with TrS1 with the aim to recognize whether or not a message is non-neutral.
Step7: Then finally relate term with training category.

Naïve Bayes:

Naïve Bayes is best classifier with having strong independence assumption and also posterior probability technique. Naïve Bayes is using the Bayes theorem. Naïve Bayes require the low storage requirement and also its result is fairly accurate due to this simplicity and good performance. Naïve Bayes gives accurate result as compare to the other classification technique. In this Naïve Bayes each document containing the term and probability depends on how many times this term will occur in this document.

Processing Steps:

Step1: Step1. Input the message.

Step2.Calculate the Prior Probability using following expression: the number of documents in class C / the total number of documents.

Step3. Apply Bayes Theorem using following expression:
P (A=B) = P (B=A)*P (A)/P (B)

Step4: Then finally relate term with category.

V. PERFORMANCE PARAMETER

Precision: tp \ (tp + fp)
Recall: tp \ (tp+fn)
F-Measure: (2* precision*recall) \ (precision+ recall)

VI. EXPERIMENTAL SETUP

Software Setup:
Operating System: Windows 7 and above
Language Used For Implementation: Java 1.8
Development Environment: Net beans 8.0.2
Hardware Setup:
RAM: 4GB and above
Processor: Intel CORE i3 and above

VII. RESULT

This section presents the performance of the RBFN and Naïve Bayes algorithm in the form of graph in terms of Precision, Recall and F-Measure. The results are obtained by using the dataset from website (www.macrovanetti.com/pages/wmsnsec/).

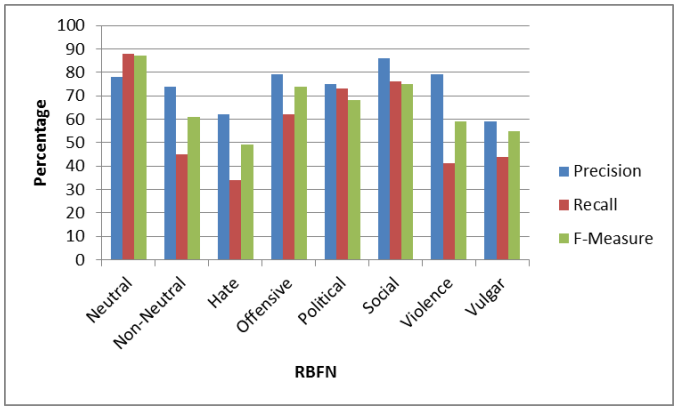

Fig. 2 Results of RBFN in terms of Precision, Recall and F-Measure

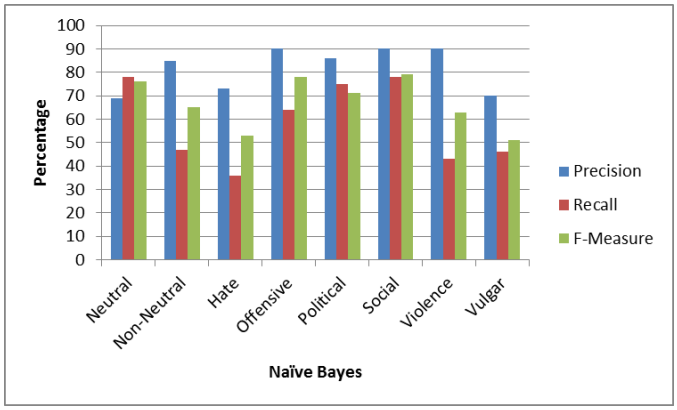

Fig.3 Results of Naïve Bayes in terms of Precision, Recall and F-Measure

VIII. CONCLUSION

In this paper, a system to prevent the display of unwanted messages posted on OSN user wall has been presented. The new system is called Filtered Wall (FW) that enable OSN users to directly control the messages posted on their private space by means of a flexible rule based system .In addition to this an OSN user can make the system more flexible by means of filtering rules. The Usage of Machine Learning has given efficient results to the system to find out the different categories of content present in the message such as vulgar,

hate, sex, violence, offensive. Naïve Bayes gives better result in terms of Precision, Recall and F-Measure and the system allows the users to customize their private space through the application of filtering criteria and a Machine Learning (ML) based soft classifier automatically labelling messages in support of content-based filtering

.

REFERENCES

[1] Marco Vanetti , M. carullo, "A System to Filter Unwanted Message from OSN user wall ," *IEEE Transaction knowledge and data engineering* vol:25,2013.

[2] S.L. Ting, W.H. Ip, Albert H.C. Tsang, "Is NaIve Bayes a Good Classifier for Document Classification?" *International Journal of Software Engineering and Its Applications* Vol. 5, No. 3, July, 2011.

[3] Dipali D. Vidhate, Ajay P. Thakare," An Avoid Unwanted Messages from OSN User Wall: Content Based Filtering Approach", *IJCSNS International Journal of Computer Science and Network Security*, VOL.15 No.5, May 2015.

[4] P. Pawar,"Comparative Study on Different Types of Approaches to Text Categorization", *International Journal of Machine Learning and Computing*, Vol. 2, No. 4, August 2012.

[5] Shruti. T, Greeshma.T, "Filterd Wall An Online Social Network", *International Journal of Science,Engineering and Technology Reserch(IJSETR)*, Volume3, Issue4, April 2014.

[6] 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.

[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. 482494, 2008.

[8] A. Adomavicius, G. Tuzhilin,"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. 734749, 2005.