IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

An Intelligent System to Filter Undesired Message from Online Social Media Wall

Miss. Radhika Fulzele[1], Prof. Vipul Dalal[2]

Student, Department of Computer Engineering, Vidyalankar Institute of Technology, Maharashtra, India[1]

Associate Professor, Department of Computer Engineering, Vidyalankar Institute of Technology, Maharashtra, India[2]

Abstract: Today's online social networks (OSN) have become a fairly common way to keep in touch with friends and family. OSN is an easy and interactive medium for communication; a number of people exchange messages with each other, these people may be desired or undesired. People may carry a wrong intent while sending messages to friends or unknown people, such scenario may distract the personal space of an OSN user. Cyber bullying due to online social network (OSN) is also on rise therefore the need of present situation is not only to avoid unwanted wall post but also to avoid undesired creator. OSN does not completely support this requirement OSN deals with the same issue, where a system is proposed, in which users directly can control the posted messages on their wall. A rule based system is designed which allows a user to manage their filtering criteria. Also the use of machine learning based soft classifier helps in automatically labelling the messages that supports content based filtering. Additionally we are implementing an automated strategy to blacklist user on a temporary or permanent basis.

Keywords: On-line Social Networks, Fuzzy c-Means, Information Filtering, Short Text Classification.

1. INTRODUCTION

People consider online social media is an interactive medium to share, communicate considered amount of human life information. They exchange several types of content like text, image, audio, video data. Many social networking sites are Facebook, twitter, google+, etc. Facebook provides platform to build social relations with people, share information with friends, finding old friends, colleagues whose account is active on social media website. OSN is a web based service through which an individual can send request to others and share information and create public profile on the web. In this, one can send data to other it may be unwanted messages which user do not show on public wall.

Web content mining strategies is generally used to uncover useful and relevant information from huge amount of data. It provides less security while posting unwanted content on OSN. There is the possibility of posting or commenting other posts on general walls i.e public/private walls. information filtering is done in which it explored the textual document or unrelated data . It can be used to give users the ability to control messages written on their own walls by filtering out unwanted messages but this service has not been provided so far. Though OSN do not completely support this requirements (eg: facebook). In Facebook, it gives permission to users to state that, is permitted to insert messages in their walls. Here, no content-based preferences are supported and hence it is not possible to prevent unwanted or undesired messages (such as political, vulgar) no matter of the user who posts them.

This service not only uses the earlier defined web mining techniques but also it requires an adhoc classification strategy as wall messages consists of maximum use of short texts but the methods currently in use produce the poor results in evaluating such messages.

2. LITERATURE SURVEY

Now-a-days [1] the web continues to grow, that has become very complicated to find related data with search engine. Some issues occur while searching data first location of relevant data and filtering irrelevant data. Each web page consist set of content based and link based features. They used two approach feed forward and back propagation neural network.

Support vector machine (SVM) and neural network [2] [3] is used for learning text categorization. This technique is used to classify news stories, to find interesting information on the web.

Earlier recommender systems were used to reach relevant products that mainly focused on user's likes and dislikes. In [4] describe content based book recommendation method .This method gives the information of the item itself for access to relevant products.

This system used information extraction and machine learning algorithm for text categorization.

Short text classification technique is used for text classification when the data set contains large documents [5]. We used neural learning strategy to semantically categorize short text. This technique is called as hierarchical strategy. Hierarchical strategy is better used to identify and eliminate the sentence by neutral and non-neutral.

OSN users allow controlling the messages posted on their walls. This is done by rule based system called filtering wall (FW). User sets some rule which if matched, and then the user is not allowed to post the messages. Here, content based filtering criteria are to be applied to their walls. This

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 6, June 2016

system blocks unwanted messages but will not block the sender from writing unwanted messages, which is the only drawback of the system.

3. PROPOSED WORK

The system analyses the social media data and automates the process of filtering the unwanted content from users and at the same time maintains the black list of the users. This system tracks the potential unwanted users and adds them to the black list. For maintaining blacklist, machine learning algorithm is used.

Black list (BL) mechanism is used to avoid messages from undesired creators irrespective of their message contents – the good or the bad. Thus the BL is directly required to be managed by the system and therefore, the system should be such that it should be able to differentiate between the users to be inserted in the blacklist and the ones to be retained in the system.

The proposed work is based on [6] in which blacklist process was handled manually. In this work, the blacklist process is automatically handled with the help of machine learning algorithm.

3.1 FUZZY c-MEANS

Fuzzy C-means (FCM) [7] is a data clustering technique. In this technique, we group data set into 'n' clusters with each data point in the dataset belonging to each cluster. The algorithm first calculates cluster centers and allocates points to these centers using a form of Euclidian distance. The process is continuously repeated until the cluster centers stabilize. Membership value is assigned to the data items for the cluster within range from 0 to 1. It supports the overlapping clusters. Steps of FCM algorithm are as follows:

1. Fix c (2<=c<n) and select a value for parameter 'm'. Initialize the partition matrix, U (0), Each step in this algorithm will be labeled as r, where r=0,1,2….

2. Calculate the c centers {Vij} for each step.

$$V_{ij} = \frac{\sum_{k=1}^{n} \mu_{ik}^{m'} \cdot x_{kj}}{\sum_{k=1}^{n} \mu_{ik}^{m'}}$$

3. Calculate the distance matrix

$$d_{ik} = \left(\sum_{j=1}^{m} \left(x_{kj} - v_{ij} \right)^2 \right)^{1/2}$$

4. Update the partition matrix for the rth step, U^r as follow

$$\mu_{ik}^{r+1} = \left(\sum_{j=1}^{c} \frac{d_{ik}^{r}}{d_{jk}^{r}}^{\frac{2}{[m'-1]}} \right)^{-1} \text{ Or}$$

$$\mu_{ik}^{r+1} = 0$$

5. If $\|U^{(r+1)} - U^r\| < €$ L then STOP: otherwise r=r+1 and return to step 2.

In step 5 we compare a matrix norm of two successive fuzzy partitions to a prescribed level of accuracy, €L ,to determine the solution is good enough.

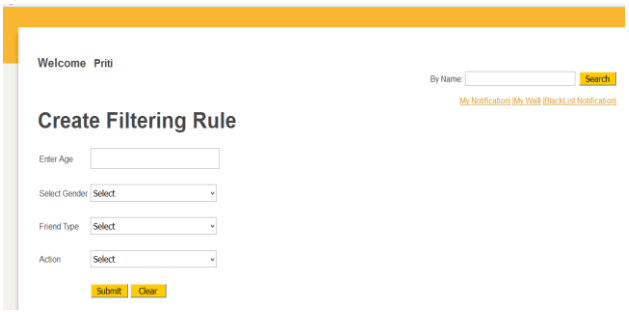

Fig 1. Set Filtering rules to block the user

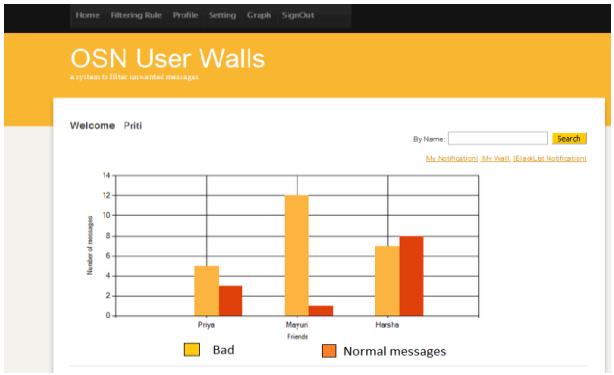

Fig 2.Graph analysis of Normal and Bad messages

3.2 Observations:

Fig 1 Shows to create filtering rules. This filtering rule (FRs) is set by wall owner to block the sender. The wall owner should set the rules for every user. If the rule is matched then that sender is automatically blocked by wall owner. It is automatically added into Blacklist (BL) from user.

Fig 2. Shows graph analysis of normal and bad messages. It counts the number of messages which is sent to the user. If the number of count of bad messages is more than threshold value then the user automatically is blocked, but it depends on wall owner to block that user on permanent or temporary basis.

4. CONCLUSION

The system is filtering unwanted message from OSN user walls. Filtering rules (FRs) is applied to filter content on OSN user walls. Machine learning techniques are used to classify the text in which it filters unwanted data. The system should be able to determine blacklist automatically and it should filter out the unwanted user. It is also able to determine automated strategy to blacklist user on a temporary or permanent basis with respect to bad post he wishes to post on OSN user wall. The system should automatically block the user.

DOI 10.17148/IJARCCE.2016.5615

REFERENCES

[1] 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.

[2] T. Joachims, "Text categorization with support vector machines: Learning with many relevant features," in Proceedings of the European Conference on Machine Learning. Springer, 1998, pp. 137–142.

[3] E. D. Wiener, J. O. Pedersen, and A. S. Weigend, "A neural network approach to topic spotting," in Proceedings of 4th Annual Symposium on Document Analysis and Information Retrieval (SDAIR-95), Las Vegas, US, 1995, pp. 317–332.

[4] 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.

[5] D. D. Lewis, Y. Yang, T. G. Rose, and F. Li, "Rcv1: A new benchmark collection for text categorization research," Journal of Machine Learning Research, 2004.

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

[7] Hung,Ming-Chuan Dept. of Inf. Eng., Feng Chia Univ.Taichung,Taiwan Yang, Don-Lin L. on "An efficient Fuzzy C-Means clustering algorithm " Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference, 2001

DOI 10.17148/IJARCCE.2016.5615