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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014

Machine Learning system for Access Control in content sharing online social Networks

G.Vanitha [1], Mr. A.Bala Subramanian[2]

Research Scholar, Department of Computer Science, SNR Sons College, Coimbatore, India[1]

Associate Professor, Department of Information Technology, SNR Sons College, Coimbatore, India[2]

Abstract: Online social networks (OSNs) have experienced tremendous growth in recent years and become a de facto portal for hundreds of millions of Internet users. These OSNs offer attractive means for digital social interactions and information sharing, but also raise a number of security and privacy issues. The OSNs allow users to restrict access to shared data; Currently they do not provide any mechanism to enforce privacy concerns over data associated with multiple users. One basic issue in today's Online Social Networks (OSNs) is to give users the power to regulate the messages denote on their own non-public house to avoid the display of unwanted content is displayed. In the existing Apriori algorithm user cannot safeguard his or her blog, because unwanted message causes many problems. To overcome this problem Naive Bayes algorithm is proposed in this project. In the proposed system, the technique applied is novel machine learning algorithm for text classification based on Naive Bayes classifiers. The experimental results obtained in Naive Bayes classifier algorithm has improved accuracy.

Keywords: Data mining, Online Social Networks, Filtering Message, Apriori Algorithm and Naive Bayes Classifier Algorithm.

I. INTRODUCTION

Online Social Networks (OSNs) are today one of the most popular interactive medium to communicate, share, and disseminate a considerable amount of human life information. Daily and continuous communications imply the exchange of several types of content, including free text, image, audio, and video data. According to Face book statistics average user creates 90 pieces of content each month, whereas more than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) are shared each month.

The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data.

They are instrumental to provide an active support in complex and sophisticated tasks involved in OSN management, such as for instance access control or information filtering. Information filtering has been greatly explored for what concerns textual documents and, more recently, web content.

The aim of the majority of these proposals is mainly to provide users a classification mechanism to avoid the useless data. In OSNs, information filtering can also be used for a different, more sensitive, purpose.

This is due to the fact that in OSNs there is the possibility of posting or commenting other posts on particular public/private areas, called as general walls. Information filtering can therefore be used to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted messages. This is a key OSN service that has not been provided so far. Indeed, today OSNs provide very little support to prevent unwanted messages on user walls. For example, Face book allows users to state who is allowed to insert messages in their walls (i.e., friends, friends of friends, or defined groups of friends). However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them. Providing this service is not only a matter of using previously defined web content mining techniques for a different application, rather it requires to design ad hoc classification strategies. The aim of the present work is therefore to propose and experimentally evaluate an automated system, called Filtered Wall, able to filter unwanted messages from OSN user walls. It exploits Machine Learning text categorization techniques to automatically assign with each short text message a set of categories based on its content.

The original set of features, derived from endogenous properties of short texts, is enlarged here including exogenous knowledge related to the context from which the messages originate.

II. LITERATURE SURVEY

The following table provides the detailed background study of technique and limitation related to online social network

TABLE I
LITERATURE SURVEY

Author	Technique Used	Limitations
Mohamed Shehab[1]	"Access control for online social networks	An access control framework to manage third party

	third partyApplications"	applications. This framework is based on enabling the user to specify the data attributes to be shared with the application and at the same time be able to specify the degree of specificity of the shared attributes.
Mayuri Uttarwar[2]	"A Review on Customizable Content-Based Message Filtering from OSN User Wall"	A major task of today's online social network is information filtering. Using machine learning approach and a rule based system, text classification and customization of filtering criteria to be applied on user's wall is to be achieved.
Miss. Rashmi[3]	"Survey of Filtering System for OSN (Online Social Networks". Online Social Networks"	A new 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 (ML) based soft classifier automatically labelling messages in support of content-based filtering.
Sujapriya. G [4]	"Filtering Unwanted Messages from Online Social Networks (OSN) using Rule Based Technique] "	OSNs provide very little support to prevent unwanted messages on user walls .Though, no content -based partialities are preserved and therefore it is not possible to prevent undesired communications , for instance political or offensive ones, no matter of the user who posts them. To propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls.
M. Vanetti, E. Binaghi [5]	"Content-based Filtering in On-line Social Networks".	The system allows 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 a user to customize the filtering criteria to be applied to their walls, and a Machine Learning based soft classifier automatically producing membership labels in support of content-based filtering.

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014

Churcharoenkrung, N[6]	"Dynamic web content filtering based on user's knowledge".	The simple and efficient solution to this problem is to block the Web sites by URL, including IP address. However, it is not efficient for unknown Web sites and it is difficult to obtain complete block list. Content based filtering is suggested to overcome this problem as an additional strategy of URL filtering.
Dipali D [7]	"A System Approach to Avoid Unwanted Messages from User Walls".	One fundamental issue in today user wall(s) is to give users the ability to control the messages posted on their own private space to avoid that unwanted content is displayed. Up to now user walls provide little support to this requirement. To fill the gap, a proposed system allowing user wall users to have a direct control on the messages posted on their walls.
Victoria Bobicev[8]	"An Effective and Robust Method for Short Text Classification".	Classification of texts potentially containing a complex and specific terminology requires the use of learning methods that do not rely on extensive feature engineering. In this work we use prediction by partial matching (PPM), a method that compresses texts to capture text features and creates a language model adapted to a particular text. We show that the method achieves a high accuracy of text classification and can be used as an alternative to state-of-art learning algorithm.

III. METHODOLOGY

A. Apriori Algorithm

Apriori algorithm uses breadth-first search and a tree structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length k-1. Then it prunes the candidates which have an infrequent sub pattern. According to the downward closure lemma, the candidate set contains all frequent k-length item sets. After that, it scans the transaction database to determine frequent item sets among the candidates.

B. Apriori Algorithm Basics

Apriori Algorithm is an influential algorithm for mining frequent item sets for Boolean association rules.

- Frequent Itemsets: The sets of item which has minimum support (denoted by L_i for i^{th}-Itemset).
- Apriori Property: Any subset of frequent itemset must be frequent.
- Join Operation: To find L_k , a set of candidate k-itemsets is generated by joining L_{k-1} with itself.
- Find the frequent itemsets: the sets of items that have minimum support
- Use the frequent itemsets to generate association rules.

C. APRIORI PSEUDO CODE

Join Step: C_k is generated by joining $L_{k-1}$with itself
Prune Step: Any (k-1)-itemset that is not frequent cannot be a subset of a frequent k-itemset
Pseudo-code:
C_k: Candidate itemset of size k
L_k: frequent itemset of size k
L_1 = {frequent items};

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014

For (k = 1; L_k !=∅; k++) do begin
C_{k+1} = candidates generated from L_k;
For each transaction t in database do
Increment the count of all candidates in C_{k+1}
L_{k+1} = candidates in C_{k+1} with minimum_support

D. NAIVE BAYES: LEARNING ALGORITHM

* From training corpus, extract Vocabulary
* Calculate required $P(c_j)$and$P(x_k|c_j)$terms For each c_j in C do

$$P(c_j) \leftarrow \frac{|docs_j|}{|total\#documents|} *$$ $docs_j$←subset of documents for which the target class is c_j
$text_j$← single document containing all $docs_j$
for each word x_k in Vocabulary
* n_k← number of occurrences of x_k in $Text_j$

$$P(x_k|c_j) \leftarrow \frac{n_k+\alpha}{n+\alpha|Vocabulary|} *$$ Positions ← all word positions in current document which contain tokens found in Vocabulary
End Return$\cup_k L_k$;

$$c_{NB} = \underset{c_j \in C}{\operatorname{argmax}}\ P(c_j) \prod_{i \in positions} P(x_i|c_j)$$

$$c_{NB} = \underset{c_j \in C}{\operatorname{argmax}}\ P(c_j) \prod_{i \in positions} P(x_i|c_j)$$ Returnc_{NB}, where

FRAMEWORK OF PROPOSED SYSTEM

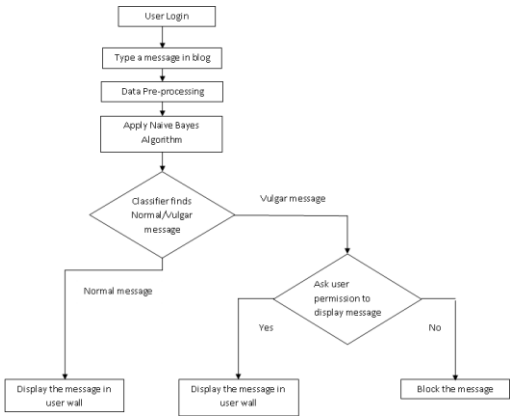

Figure : 1

STEPS FOR PROPOSED SYSTEM

STEP 1: User Login by entering his or her Login ID and Password.
STEP 2: Once the user is logged in, the user tries to post the message to his/ her contacts.
STEP 3: First step is Data pre-processing which consist of Stemming and Formatting the input data for removal of stop words.
STEP 4: Apply the Naive Bayes Algorithm, to find the number of occurrence and position of the text from which the data is classified accordingly.
STEP 5: If the classified data is a normal message then the user is allowed to post his/her message in the user wall without any alert.
STEP 6: If the classified data contains even a single vulgar/violent word then the input data is classified under the spam message.
STEP 7: Receiver of the message will be prompted with an alert asking whether to display the message which consist of vulgar/violent word sent by the particular user.
STEP 8: Receiver can decide upon whether to display the message or to block the message.

IV. RESULTS AND DISCUSSIONS

4.1 CLASSIFICATION USING APRIORI ALGORITHM
The main characteristics of the data sets used in the experiment are summarized and shown in Fig 4.1. The first column describes the set of classes i.e Angry, Happy, Information, Meeting, Violence, Vulgar, Offensive, Undefined, and Sad while the other column indicates the number of values contained in a particular class name using Apriori algorithm. In Apriori classification vulgar words are not classified

Class Name	Apriori obtain values
Angry	0
Happy	16
Information	291
Meeting	10
Offensive	0
Sad	0
Undefined	0
Vulgar	0

Figure 4.1 Result obtained from Apriori

The below graph describes the clasification of the dataset using Apriori algorithm. Only class names such as Happy, Information and Meetings are classified with values16,291,10 respectively.

Angry,Offensive,Sad,Vulgar classes which leads to the spam messages are not classified.

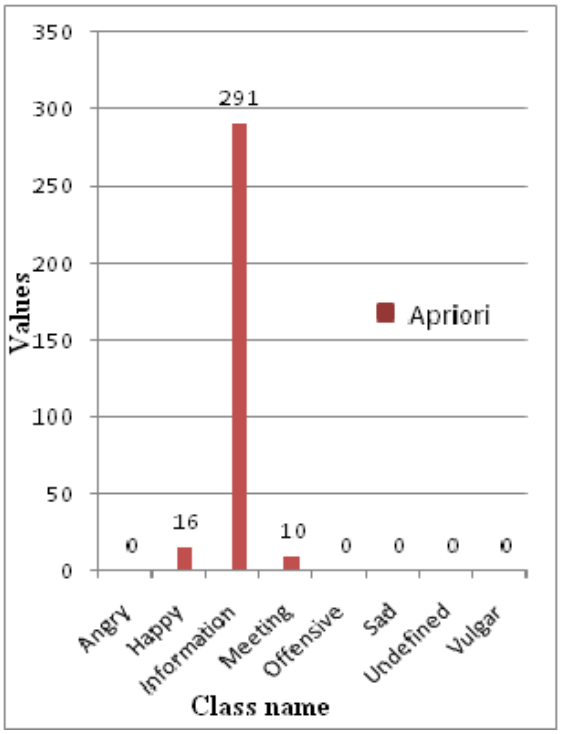

Figure 4.2 Classification chart in Apriori algorithm

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

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 6, June 2014

4.3 CLASSIFICATION USING NAIVE BAYES ALGORITHM

The below table describes the values obtained using Naive Bayes algorithm. Spam filtering technique which uses Naive Bayes classifier is used to classify the class name. In the figure 4.3, class names such as angry, offensive,sad and vulgar are classified with values 1,2,2,25 repectively which are not found in Apriori. Undefined class contains the messages which does not fit in any of the class category. In the above experiment class Happy contains 16 values but in Naïve Bayes 14 values are considered under class Happy. This is because eventhough if the text is a happy classifier it may contain some of the vulgar/violent words, so those text are classified under vulgar or any of the class that leads to spam messages.

Class Name	Naive Bayes Obtain values
Angry	1
Happy	14
Information	191
Meeting	31
Offensive	2
Sad	2
Undefined	51
Vulgar	25

Figure 4.3 Result obtained from Naive bayes

The below graph describes the clasification accuracy of the dataset obtained using Naive Bayes algorithm. Values are compared against the class names and plotted.The classification and values found is accurate in Naive Bayes algorithm.

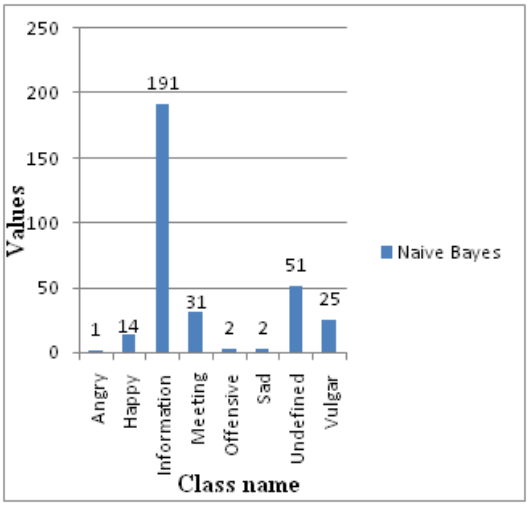

Figure 4.3 Classification chart in Navie Bayes algorithm

V. CONCLUSION

The system exploits a Machine learning soft classifier to enforce customizable content-dependent Filtering Rules. Moreover, the flexibility of the system in terms of filtering options is enhanced through the management of Black List. This work is the first step of a wider project. The existing problem is clearly find out in the project. The problem is solved using Naive Bayes algorithm. The proposed approach handles very well with high-quality database. This high dimensional database is much better, efficient and it can do more work and gives better result. The user can safeguard the blog by avoiding unwanted message or spam message. The classification clearly classified comparing to existing method. The user can identify who have sent unwanted message without opening his or her blog. So, maximum security is achieved by the proposed method.

REFERENCES

[1] Mohamed Shehab and Anna Squicciarini"Access control for online social networks third partyApplications "of the International Conference on online social network 11 July 2004.

[2] Mayuri Uttarwar, Prof. Yogesh Bhute " A Review on Customizable Content-Based Message Filtering from OSN User Wall. MayuriUttarwar et al,International Journal of Computer Science and Mobile Computing Vol.2 Issue. 10 , October-2013

[3] Miss. Rashmi R. Atkare Prof. P.D.Soni" Survey of Filtering System For OSN "(Online Social Networks)"National Conference 21 August 2003.

[4] Sujapriya. G. Immanual Gnana Durai" Filtering Unwanted Messages from Online Social Networks (OSN) using Rule Based Technique]" OSR Journal of Computer Engineering (IOSR-JCE) 1 Jan. 2014.

[5] M. Vanetti, E. Binaghi, , B. Carminati, M. Carullo and E. Ferrari" . Content-based Filtering in On-line Social Networks" VLDB ConferenceVarese, Italy 2005.

[6] Churcharoenkrung, N., Kim, Y.S., Kang, B.H." Dynamic web content filtering based on user's knowledge". International Conference on Information Technology: Coding and Computing 1, 184–188 (2005).

[7] Dipali D. Vidhate , Ajay" A System Approach to Avoid Unwanted Messages from User Walls".International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Volume 3 Issue 2, February 2014.

[8] Victoria Bobicev, Marina Sokolova "An Effective and Robust Method for Short Text Classification". CHEO Research Institute 401 Smyth Road, Ottawa, Ontario, Canada. Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (2008).