THE INTERNATIONAL JOURNAL OF SCIENCE & TECHNOLEDGE

Enabling Security Service Using Content Based Message Filtering System on OSN Userwalls

Jebeen M.
SCSVMV University, Kanchipuram,Tamil Nadu, India
Poorva Devi R.
Assistant Professor, SCSVMV University, Kanchipuram,Tamil Nadu, India

Abstract:
At the present day,Online Social Networks gains more and more popularity , since it provides the most easiest interactive medium to share the informations among the group of users.However,it provides a little support to control the messages that are posted on their own private user walls.In this paper,an automated information filtering system is proposed that can filter the posted messages based on the filtering criteria.It allows OSN users to customize the filtering criteria on the messages and also the undesired users might be banned from a wall for the specified time period.The proposed approach uses two level text analysis algorithm.In the first level,posted messages are classified using the improved text classification algorithm and the messages are labeled.The second level filter the message based on the Filtering Rules(FR) that are stated by the user wall.Finally,the undesired messages that are posted by the group user more than the threshold value can be placed in the BlackList based on the BlackList(BL) rule which is also customized by the user wall.The BL users can be banned for the specified time period from the user wall.Besides,the efficiency of our approach is manifested by the preliminary experiment.

Key words*: Content-based message filtering, Text classification, Filtering rules(FR), Blacklist(BL) management*

1. Introduction

1.1. Security

The tradeoff of security applies can lie in technology, accounts, network access, equipment and content. Often, the default settings provided by the online social media is greatest of ease of use but are also the least secure.It is never a good idea to trust someone else has you security in mind, so you should check and configure these settings yourself.One of the security settings of restrict content from appearing on your profile page and prevent people from accessing photos and other content.

Today,OSNs provide very little support to prevent the unwanted messages on the user walls.It allows users to specify who is allowed to post messages in their walls.However,it does not support the content based preferences and therefore it is not possible to prevent undesired messages,no matter of the user who posts them.Therefore the aim of the present work to propose and experimentally evaluate an automated system able to filter unwanted messages from the user walls in order to prevent the undesired users to enter in our user wall.

1.2. Security Reviews

Threats change, so security needs to change and address them. Social media sites often make changes to their security to address identified or potential issues. While some sites notify you of updates and new settings, others may implement them without your knowledge, leaving you unaware of what's happened. How often this happens, and whether you're notified depends on the site. Because of this, you should review your settings from time to time to ensure they're configured the way you intended.

When a social media site changes its security, it can affect the options that are available. While you may have thought security was set up properly, the options may have changed. In some cases, the changes may reset your security settings to their default settings or provide additional options that may need to be set. The site may decide to turn on a setting that you don't want or make the option available and turned off. To benefit from the available security options, you need to review them and make sure they're set properly.

1.3. Security Strategies

The purpose of a social media security strategy is to give people the ability to do what's needed without compromising security. In creating one, you need to identify what areas need to be secure, how security will be achieved, and who will be responsible. The strategy should encompass any areas related to using social media, inclusive to the corporate workstations people may use, mobile devices issued to employees, network security, and firewall restrictions.

In this proposed approach, security can be given on the content to be posted on the user walls and also restrict the undesired user to post their message on other user walls.

2. Automated Content-Based Filtering System

The proposed system is the design of a system providing customizable content-based message filtering for the social media networks.It has to perform the following tasks.

2.1. Data Preprocessing

Data preprocessing is preparation for data mining and it mainly includes data scrubbing,data integration,data conversion,data reduction,etc.,The entire process begins with collection of evidence acquired from various data sources.Figure 1 shows the data collection from different sources and the preprocessing like stemming can be performed on the collected text messages.Here we have collected the text messages from the various news websites,since the classes to be categorized like politics,sports,general.In the ideal situation,the data should be of low-dimentionality,independent and discriminative so that its values are very similar to characteristics in the same class but very different in features from different classes.

2.2. Learning and Classification

Learning is the term used to describe the actual process of training the classification model. One can distinguish three learning strategies: supervised, unsupervised and reinforcement learning.

The automated filtering system supports supervised learning where the learning algorithm is given a labelled training set to build the model on.It is called " supervised " since the outcome of the new data item can returned on the basis of which the model learns how to return the best solution to the given problem.

Figure 1 illustrates the learning in the third step that can learn by scan the messages and identify the features to be used to represent the particular category message. The feature selection is the process of selecting a specific subset of the terms of the training set and using only them in the classification algorithm. The feature selection process takes place before the training of the classifier.The main advantages for using feature selection algorithms are the facts that it reduces the dimension of our data, it makes the training faster and it can improve accuracy by removing noisy features. As a consequence feature selection can help us to avoid overfitting.The basic selection algorithm for selecting the k best features is presented below (Manning et al, 2008):

```
SELECTFEATURES(D, c, k)
1  V ← EXTRACTVOCABULARY(D)
2  L ← []
3  for each t ∈ V
4  do A(t, c) ← COMPUTEFEATUREUTILITY(D, t, c)
5     APPEND(L, ⟨A(t, c), t⟩)
6  return FEATURESWITHLARGESTVALUES(L, k)
```

The trained model learn from the labelled instances in the training dataset.A part of the available data has to be left out for testing purposes,which further narrows down the amount of data to be used for a proper training of the classifier.In order to measure the performance of the classifier,we can use the new dataset with unknown instances.

Let TR={t1,t2,…tn} be the training set,where ti,i=1,2,…n are set of text messages.The text classifier train with TR messages and ci be the belongingness class of the message ti. Naive Bayes methods are a set of supervised learning algorithms based on applying Bayes' theorem with the "naive" assumption of independence between every pair of features. Given a class variable y and a dependent feature vector $x_1$through x_n, Bayes' theorem states the following relationship:

$$P(y \mid x_1, \ldots, x_n) = \frac{P(y)P(x_1, \ldots x_n \mid y)}{P(x_1, \ldots, x_n)}$$

Using the naive independence assumption that$P(x_i|y, x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) = P(x_i|y)$,
for all i,

The filtering system for short-text message classification extracts three types of features,these are word features,single characters features,statistical features.These features extraction can be performed using the modified stemming algorithm to extract the above features.

This combination of features identification gave the best generalization accuracy for posted user wall messages.The features can be represented in two possible ways; binary(i.e,true or false,indicating that a particular feature simply exists in the text or not) and numeric(i.e,a number representing the frequency of a particular feature in the text).

The proposed system uses the Naïve Bayes classifier would be "independent feature model" that assumes that the presence(or absence) of a particular feature of a class is unrelated to the presence(or absence) of any other feature.For example,a message may be considered to be a political message if it has the words related to the politics.Even if these features depend on each other or upon the existence of the other features,a naïve Bayes classifier considers all of these properties to independently contribute to the probability that this message is politics.

Let we have seven text messages and among them two messages are politics: "Assembly." And "Assembly Court Won." and the other two messages are considered as sports: "Match Won.", "Court Match" and the last three messages are considered as General:"Hello","Hello Won" and "Hello Won Match" in Table I. However, for training the system construction of vector table is very important and need to train the system through the vector table. Initially we have only one feature extraction process which

breaks down each message into individual words and produces words by separating the words by space or comma(,) or full stop(.) or exclamatory sign(!). So after the feature extraction process the words become the word vocabulary: "Assembly", "Court","Match","Won" and "Hello".This feature selection process considers only few number of words.But we have to take into account a large number of possible political,sports and general messages in order to improve the accuracy of the classification process.

Message No	Type	Word attributes				
		Assembly	Court	Match	Won	Hello
1	Politics	1	0	0	0	0
2	Politics	1	1	0	1	0
3	Sports	0	0	1	1	0
4	Sports	0	1	1	0	0
5	General	0	0	0	0	1
6	General	0	0	0	1	1
7	General	0	0	1	1	1

Table 1: Vector Table

As the Naïve Bayes is the probabilistic classifier, we don't need to know the total number of words in each SMS, thus the vector table can be replaced by the word occurrence table which is demonstrated in Table II,

Word Attribute	Politics Occurences	Sports Occurences	General Occurences
Assembly	2	0	0
Court	1	1	0
Match	0	2	1
Won	1	1	2
Hello	0	0	3

Table 2: Word Occurences

Therefore, to classify the unknown posted messages we can demonstrate the Naïve Bayes classification as:
P(Politics|Assembly,Court,Won)
=P (Politics) X P(Assembly|Politics)
X P (Court|Politics)XP(Won|Politics)
Likewise,other Sports and General messages can be classified by calculating the probability of Politics,Sports,and General.After calculating the final probability for each messages under each category we can finally make the decision of being Politics,Sports and General depending on their majority value.
If the proportion of Politics exceeds the proportion of other messages, then it has a greater chance to be a "Politics" and the same can be applied for other messages.

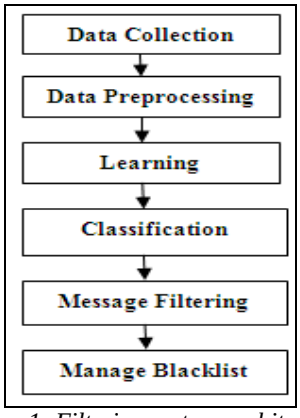

Figure 1: Filtering system architecture

2.3. Text filter with Filtering Criteria
Besides classification facilities, the system provides a facility to filtering the undesired messages based on the specified filtering rules(FRs),by which users can state what contents should not be displayed on their walls.
In order to filter the message based on the content, filtering criteria can be set as the category of the message. Hence, the automated system filter the posted messages based on the filtering criteria and display on the user wall.
For example, OSN provides the facility to post messages like politics, cinema, sports, etc. on other user walls without consider the content of the messages.The user can specify the category of messages that they don't want to display on their walls.So,the owner

of the user wall can set the filtering criteria on their profiles that can filter the incoming messages based on the specified filtering criteria. The Filtering Rule(FR) can be specified as part of the user profile and that can be dynamically changed by the user.

2.4. Blacklist

The BL mechanism can be supported by the proposed system to avoid messages from undesired creators, independent from their contents. It should be able to determine who are the users to be inserted in the BL and decide, when user's retention in the BL is finished.

The BL mechanism can use the BL rules that can also be set by the user wall owner who can decide to specify the BL rules regulating who has to be banned from their walls and for how long. In our proposed system, the BL rules can be set dynamically by the user wall owners able to identify users to be blocked according to their profiles as well as their relationships in the OSN.

3. Implementation Method

The internal operation of the filtering system can be depicted graphically in Figure 2. The filtering system consists of two layers, namely GUI (Graphical User Interface) and FW (Filtered Wall).

The top layer (GUI) provides the users with their wall where only messages that are authorized according to their FRs and BLs are published.

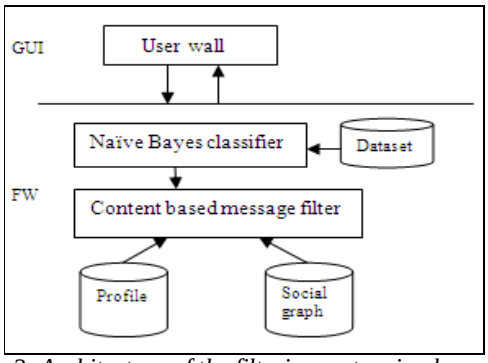

Figure 2: Architecture of the filtering system implementation

The bottom layer (FW) performs the message classification based on the learn model and filtering based on the filtering criteria in the user profile which is created by the user and managed by the filtering system.

The internal operation of the filtered system can be depicted graphically in Figure 2.It shows the flow path followed by a message from its writing to the possible final publication can be summarized as follows.

- The user enters into their private wall and tries to post a message.
- The FW intercepts the message and classify using naïve bayes text classification and forward to the message filter.
- The message filter learns a profile of the user's interest based on the features presented in the database.Filtering rules are applied on the classified message and generate the result.
- Depending on the result of the filter, the message will be published (or) filtered.

4. Experiment and Result Analysis

After the collection and classification of the text messages from various sources (like news websites), the experimental database consists of more than 600 messages and less than 1000 that are split into two groups, training and test data sets.

It's very difficult to make a judgment that a classification algorithm is better than another because it may work well in a certain data environment, but worse in others. Evaluation on performance of a classification algorithm is usually on its accuracy. However, other factors, such as computational time or space are also considered to have a full picture of each algorithm. This project will evaluate performance of various classification techniques on three different data environment, which are noisy, un-noisy, and computational-intensive.

Classification technique can be classified into five categories, which are based on different mathematical concepts. These categories are statistical-based, distance-based, decision tree-based, neural network-based, and rule-based. Each category consists of several algorithms, but the most popular from each category will be used and compared on 3 different classification tasks. The algorithms are C4.5, Naïve Bayes, K-Nearest Neighbors.

The performance of the proposed system is purely depends on the accuracy of the classification technique.The second part of the FW provides the filtering strategy that depends only on the user profiles that compares user interests on the posted wall messages and filter the message if the posted message comes under the filtering criteria.

4.1. Algorithm comparison

4.1.1. Naïve Bayes Classifier

As we can see in the calculation formula of posterior probability, the algorithm is designed for categorical data. This is one disadvantage in Naïve Bayes. However, there is a way to overcome this problem where continuous data is divided into ranges.

Then the probability of a value is considered the range's probability. Naïve Bayes is chosen as a representative of statistical-based category because it works very well in a non-linear problem domain. A non-linear problem occurs when the model's classes can not be divided linearly, just as demonstrated in Figure 3.
The program used in this project divides continuous data into range with a predefined number of values. For example, in the experimental dataset, each attribute is divided into 3 ranges: the first range with lower bound and upper bound consists of the first 200 instances of all combination of Politics, Sports and General messages, the second range with lower bound and upper bound consists of the next 400 instances, and the third range with lower bound and upper bound consists of the last 600 instances.

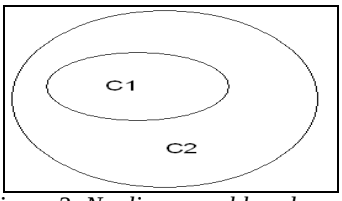

Figure 3: Nonlinear problem domain

4.1.2. K-Nearest Neighbors
How K-Nearest Neighbors works is very simple:
Finding the K nearest neighbors to an input instance in the population space and assign the instance to the class the majority of these neighbors belong to. The "nearest" measurement refers to the Euclidean distance between two instances. For example, the Euclidean distance between t_i and t_j is

$$D(t_i, t_j) = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2}$$

with p denoted as the number of attributes in each data instance. Since the class assignment is based on the majority, K must be odd to avoid situation when there is an equal number of each class. Besides, the value of K must be greater than the number of classes in the problem.

4.1.3. C4.5
C4.5 is the most popular and the most efficient algorithm in Decision tree-based approach. A Decision tree algorithm creates a tree model by using values of only one attribute at a time. At first, the algorithm sorts the dataset on the attribute's value. Then it look for regions in the dataset that clearly contain only one class and mark those regions as leaves. For the remaining regions that have more than one classes, the algorithm choose another attribute and continue the branching process with only the number of instances in those regions until it produces all leaves or there is no attribute that can be used to produce one or more leaves in the conflicted regions.
Some effective decision tree algorithms choose which attribute to make the partitions by calculating the information gain for each attribute. The gain is calculated by subtracting the entropy of the attribute from the entire dataset entropy. The entropy measure the amount of disorder in a set of values. One disadvantage of calculating an attribute's entropy is that it relates to categorical data. However, just like in Naïve Bayes, this can be overcome by dividing continuous data into ranges. The entropy of an attribute is then determined by firstly calculating the entropy in each range

$H(\text{range, attribute}) = \sum p(\text{class}_j \mid \text{range}) \times \log [p(\text{class}_j \mid \text{range})]$

(Where $p(\text{class}_i \mid \text{range})$ is the probability of each class appearing in the calculated range)
The entropy of the entire attribute is determined as following:

$H(\text{attribute}) = \sum p(\text{range}_i) \times H(\text{range}_i, \text{attribute})$

($p(\text{range}_i)$ is the probability of an attribute's range in the entire dataset)
The information gain of that attribute is then:

$\text{Gain (attribute, dataset)} = H(\text{dataset}) - H(\text{attribute})$

with $H(\text{dataset}) = \sum p(\text{class}_j \mid \text{dataset}) \times \log [p(\text{class}_j \mid \text{dataset})]$

($p(\text{class}_j)$ is the probability of class j in the entire dataset)
We have segmented the database as follows and each algorithm is running with the same data set messages.At first we train our system by first 700 messages and then test our system by next 200~300 messages as depicted in Table III.

Number of messages	Naïve Bayes Accuracy (%)	K Nearest Neighbors Accuracy (%)	C4.5 Accuracy (%)
1~200	94	93.3	86.7
201~400	95.3	94	86
401~800	95.5	93.5	84.3

Table 3

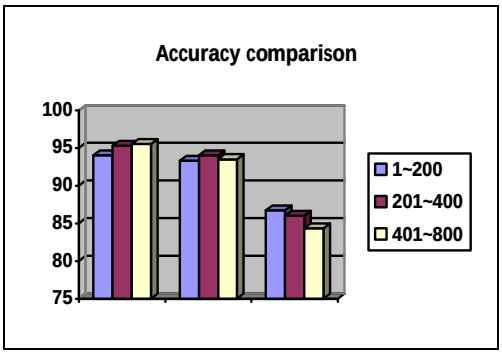

Figure 4

By increasing the size of the testing data set,the performance and accuracy of the algorithm can be determined.

4.2. Training and Testing time

If a dataset contains millions of training data with many attributes, and the number of training loops on the set is high, the network will take a very long time on a typical computer to arrive with a model. The proposed system provides the feature of banning the undesired users from their user walls and this system needs some background databases like SQL to keeps track the users and their activity in order to find the blacklist.The other algorithms run very fast not only in any data environment.

5. Conclusion

In this paper a new way of security is provided on the user wall by filtering the unwanted messages from other users.This work can be used by all the websites which wants to provide the content-based security to the users.The system exploits the data mining classification technique to enforce the customizable content-based FRs.The previous work only considered the FRs on the posted message.This new system can also provide the sophisticated approach to decide when a user should be inserted into a BL and how long they can be stayed in BL.It also can be customized by the wall user.The users who are in the BL cannot post their messages on the other user walls. The proposed system learns based on the pre-classified data that can be no longer used,since the content of the sources can be changed dynamically.In future work,we have to include the extra approach to learn the source content dynamically.

6. References

1. Adomavicius, G.and 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. 734–749, 2005.
2. 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.
3. 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.
4. F. Sebastiani, “Machine learning in automated text categorization,” ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.
5. 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.
6. N. J. Belkin and W. B. Croft, “Information filtering and information retrieval: Two sides of the same coin?” Communications of the ACM, vol. 35, no. 12, pp. 29–38, 1992.
7. P. J. Denning, “Electronic junk,” Communications of the ACM, vol. 25, no. 3, pp. 163–165, 1982.
8. 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