

RULE BASED MESSAGE FILTERING AND BLACKLIST MANAGEMENT FOR ONLINE SOCIAL NETWORK

V.Ezhilvani¹, K.Malathi², R.Nedunchelian³

¹Department of Computer Science and Engineering, Saveetha University, Chennai, India

²Department of Computer Science and Engineering, Saveetha University, Chennai, India

³Department of Computer Science and Engineering, Saveetha University, Chennai, India

Abstract

Online social network which play a most imperative role in today's world, but the major issue in online social network is displaying undesirable content in user walls. Main involvement of this paper is to propose an automated system to automatically filter the undesirable messages for online social network user walls using content based filtering .so, that the user can able to control directly the message which has been posted on their walls. In content based filtering the information item is been selected based on correlation between the content of the item and the user preferences.

Keywords: Social network, machine learning techniques, blacklist management

-----***-----

1. INTRODUCTION

Today online social network become most popular interactive medium to communicate with people and, share considerable amount of information and also gaining new friends over the internet. More than 300 social networking sites have common features in existence. The basic feature of social network is the ability to create and share a personal profile. This profile page includes a photo, and some basic personal information such as name, age, sex, location. Most of the social networks on the Internet also let you post videos, photos, and personal blogs on your profile page. One of the most important features of online social networks is to find and make friends with other site members. These friends also appear as links, so visitors can easily browse your online friend network. According to face book statistics 1million link, 2million friends are requested, 3million messages have been send for every 20 minutes on face book. The content present in social network is constituted by short text, and the notable example is the messages written by Online Social Network users on particular private or public areas, known as general walls. Proposing an automated system, to filter undesired comments from online social network is the aim of the present work. This is named as, Filtered Wall (FW).In previous work the wall owner does not have the control of their own private area and this shows there is no content based preferences. Therefore it is not possible to prevent unwanted messages such as political or vulgar once. To fill this gap, In proposed system ,user specify what are the contents not to be displayed on his or her walls by providing certain set of filtering rules .Thus, the present idea support content based preferences ,that is the user have the direct control of their walls. This is can be done using Machine Learning (ML) text categorization procedure .And, this ML have the capability to automatically assign with each messages as a set of categories based on contents. And, also using the blacklist management to control the message posted on user walls

2. PROBLEM DEFINITION

Today most of the people splurge more time in social network sites such as Face book, twitter, LinkedIn, Google plus etc...And the major issue in social network is user does not have a control over their walls because it does not support content based preferences .Therefore it is not possible to prevent undesired messages such as political or vulgar ones which is posted on the private space of the users. Likewise, huge volumes of data are extracted and posted to the social sites, so it becomes a sophisticated task to social network management. Most of the proposal is mainly focus on providing a classification mechanism to avoid useless data.

3. PROPOSED SYSTEM

The main focus of this paper is to develop a system for online social network user to prevent unwanted messages based on content given by the user. And, also using the machine learning system to provide the flexible way to control the incoming messages. Enforce blacklist rule to avoid unwanted messages from unauthorized users to keep the wall secure.

The three layer architecture in the below fig 1.In this three layer, first layer shows the social network manager, second layer shows the social network application and finally graphical user interface.

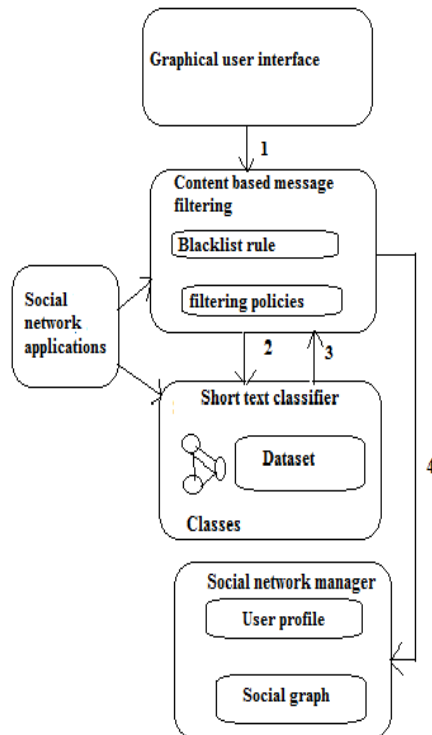


Fig 1: Filter wall system architecture

In the above Fig 1 first layer called social network manager which is used to provide the basic online social network functionalities that is the user profile and the relationship between the managements. Proposed system is placed in the second layer called as social network application in that it perform content based message filtering and short text classification based on the user profile. User interacts with the system through graphical user interface and this will manage the filter wall and blacklist management which is present in the social network application layer. Moreover, GUI provides a filter wall that is the wall will publish only the message that is authorized according to the filtering rule. The main components of the proposed system are content based filtering and short text classification. And the system makes use of blacklist management to improve the filtering actions. This blacklist is used to prevent unwanted messages from undesired users .so that the system can able to decide whose are user can enter into the blacklist and how long the message can present inside the blacklist.

4. METHODOLOGY

The main focus of this paper is to avoid and filter unwanted messages from online social network; this is supported by content based filtering and short text classifier.

4.1 Content Based Filtering

Assume that content based filtering user can operate independently, and the result shows that the information selected based on the correlation between the content of the item and user preferences. The information selected is textual in nature so it makes use of text classification. Content based filtering is present in the second layer in our

proposed work called as social network application layer .This filter focus on filtering the short text messages rather than wide range of topics. And this content based filtering fully based on machine learning paradigm.

4.2 Short Text Classifier

In our approach the short text classifier is called as multi class soft classification process and this process is composed of two phases called as text representation and machine learning based classification.

4.2.1 Text Representation

Extracting the particular features from a given set of information is critical task; this will affect the entire performance feature of classification. For our work we have taken two features such as bag of words and document property from literature[2].Text representation is called as vector model [3] in which text document is represented as d_j and the real weight of the $d_j = w1j, \dots, wTj$ t is the set of term that occur at least once from the collection of documents Tr and $wkj \in [0; 1]$ represent t_k contribute to document d_j .in back of words the terms are identified with the help of words. In case of non binary weight wkj , document d_j is computed according to the Inverse Document Frequency weighting function in literature[9] is defined as

$$tf - idf(tk, dj) = \#(tk, dj) \cdot \log \frac{|Tr|}{\#Tr(tk)}$$

Where $\#(tk, dj)$ denotes the number of times t_k occur in document d_j and $\#Tr(tk)$ denote document frequency tk .document property is adopted by collecting the correct words, bad words, capital words, punctuation character, exclamation marks etc.

Correct Words: This can be expressed in terms of $tk \in T \cap K$, where tk is the term of document d_j , and k denotes the known words.

Bad Words: This can be computed same as correct words feature, where here K denote the bad words

Capital Words: This will denote the amount of words used capital letters in the document and calculate the percentage for the capital words. For example document containing the word GOOD Work (here GOOD is capitalized and Work is not capitalized) percentage can be calculated according to the capital letter given in the document.

Punctuation Characters: This can be calculated by the ratio of total number of punctuation character in the message by total number of characters in the message. For example the documents contain the message as hi!!!How are you? is 4/15.

Exclamation Marks: calculated by the ratio between the total numbers of exclamation marks in the message by total

number of punctuation characters in the message. For example consider Hello!!What doing? Is 2/3

Question Marks: this can be calculated by the ratio between total numbers of question mark in the message by total number of punctuation character in the message. For example consider hi “Where are you?” is $\frac{1}{2}$

4.2.2 Machine Learning Technique

In previous section we have discussed about short text classification and it is called as hierarchical two level processes and it is named as first level and second level. In first level messages are classified as neutral and non-neutral. Second level partitions the non neutral messages. To understand the machine learning technique let us consider M1 and M2 as first level and second level classifier and y1 be the neutral class. The machine learning works as follows:

1. Let M_i be the message and it is processed such that x_i vector is extracted from the feature. Consider two sets such as $TrSd$ and $TeSd$ are transformed into $TrS = \{(x_i, y_i), \dots, (x|TrSD|, y|TrSD|)\}$ and $TeS = \{(x_i, y_i), \dots, (x|TeSD|, y|TeSD|)\}$.
2. Binary set is created for message M1 as $TrS1 = \{(x_i, y_j) \in TrS | (x_i, y_j), y_j = yj1\}$.
3. Multi class set is created for message M2 as $TrS2 = \{(x_j, y_j) \in TrS | (x_j, y_j), yjk = yjk + 1, k = 2 \dots, |\Omega|\}$
4. To find whether the message is neutral or non neutral for the message M1 is trained with TrS1 and Then the performance of message M1 is evaluated using TeS1
5. Performance of the M2 is evaluated using TeS2. for this message M2 is trained with Non neutral TrS2 message.

5. IMPLEMENTATION

Our proposed system is implemented to avoid unwanted messages from unauthorized users using blacklist and filter rule.

5.1 Blacklist Management

Blacklist rule is used to avoid unwanted messages created by the unauthorized users. This is the mechanism which is managed by the system, such that the system may able to decide who are users enter into the blacklist and how long the message can be present in the system. That is the wall owner must decide who are the users enter into his/her private wall and how long they present in the wall. The wall owner must specify the blacklist rules. This blacklist management will block the users based on the user profiles and the relationship in the online social networks such that the wall owner key entity to identify the users.

5.2 Filtering Rule

Filtering rule is defined as the author or creator who specifies the rules for example suppose rajev is an online social network user and he always wants to avoid the high vulgar messages. And rajev want to filter only the message

coming from indirect friends and not for the direct friends. This can be done using filtering rule as follows:

```
((rajev, friendof, 1,1),(vulgar,0.60),block)
((rajev, friend of, 1,0.3),(vulgar,0.70),block)
```

The values given in the above examples are trust value 0.60 and 0.70. if the friend of rajev with the trust value of 0.60 wants to publish the message as “h*v*ng *x*” on rajev’s private wall. After posting this message it will produce the grade value as 0.65 for the class vulgar message. Therefore, the message that containing huge degrees of vulgar content will be filtered from the system and the filtered message will not display in the user private wall. Hence, it reduces the unwanted content and improves the performance of the system. Moreover, filtering rule is the action performed by the system on the message to satisfy the rule.

5.3 Enhancement of Filtering Wall

- When the user wants to post a message on a private wall, he/she tries to enter into a wall and the user tries to post messages on the private wall but it is intercepted by the filter wall
- Secondly machine learning text classifier is used to extract the metadata from the given content of messages.
- Then the filter wall make use of this metadata which is provided by the short text classifier and along with the data extracted from the user profile by enforcing the blacklist rule and filtering rule.
- The filter walls publish or block the message depending on the result of previous step.

6. RESULTS AND DISCUSSION

The values of unwanted content such as neutral, non-neutral, violence and vulgar are determined based on the first level and second level category of the message.

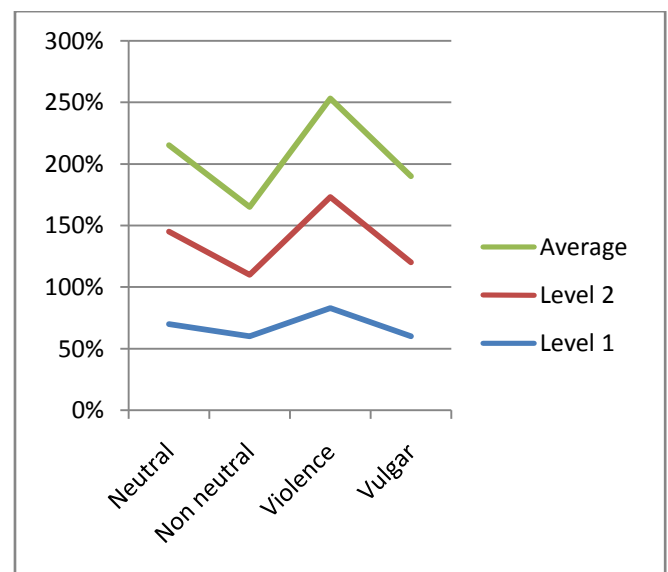


Fig 2: Test values for first and second level messages

Where set of classes such as neutral and non neutral belongs to first level, violence and vulgar belongs to second level of messages. And it is denoted by $\Omega = \{neutral, non - neutral, violence, vulgar\}$.

The above fig 2 shows the test result of first level and second level for neutral, non-neutral, violence and vulgar messages.

7. CONCLUSIONS

In this paper we have proposed a Filter wall to filter unwanted messages from online social network private walls. And we make use of blacklist management and filtering rule to provide rules to the system. Blacklist rule is used to avoid unwanted messages created by the unauthorized users. This is the mechanism which is managed by the system, such that the system may able to decide who are users enter into the blacklist and how long the message can be present in the system. So, we can avoid unauthorized person and we can keep our private wall secure.

REFERENCES

- [1]. Macro Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbana Carminati and Moreno Carullo "A system to filter unwanted messages from OSN user walls." , IEEE Transaction on knowledge and data Engineering, Feb 2013
- [2]. Carullo, M. Binaghi, E. Gallo " An Online document clustering technique for short web contents. In: Pattern Recognition letters" (july 2009)
- [3]. Manning, C., Raghavan, P., Schütze, H. "Introduction to information retrieval" (2008)
- [4]. M. Chau and H. Chen, " A Machine learning approach to web page filtering using content and structure analysis ", decision support system (2008)
- [5]. R. J. Mooney and L. Roy, "Content -based book recommending using learning for text categorization", ACM Conf. Digital Libraries (2000)
- [6]. A. Adomavicius and G. Tuzhilin, "Towards the next generation of recommender systems", IEEE Trans. Knowledge and Data Eng (june 2005)
- [7]. Sebastiani, F: Machine learning in automated text categorization. ACM Computing Surveys 34(1), (2002)
- [8]. N. J. Belkin and W. B. Croft, "Information filtering and information retrieval: Two sides of the same coin?" comm.. ACM (1992)
- [9]. Salton, G., Buckley, C. "Term - Weighting approaches in automatic text retrieval. Information processing and management (1988)