Filtered Wall-An Online Social Network Filter

Sruthi.T, Greeshma.T.R

Abstract*—*In last few years Online Social Networks (OSNs) are extremely popular among Internet users. There is major issue in today's OSNs is to prevent the display of unwanted contents in user private space. So that there is an important need in today OSNs is to give users the ability to control the messages posted on their own private space to avoid that unwanted content is displayed. From the beginning to here, OSNs provide very little support to this requirement. To full fill this requirement, this paper proposes a system that allows OSN users to directly control messages posted on their wall. This is achieved through a flexible rule based system in support of a machine learning based short text classifier, user defined BL mechanism and an image comparison technique.

Index Terms—**Online social networks, information filtering, short text classification, image comparison**

I. INTRODUCTION

Online Social Networks (OSNs) plays a vital role in our day to day life. As we know, today everyone is using OSNs as an interactive medium to share, communicate, and distribute a significant amount of human life information. A user can communicate with other user by means of sharing several types of contents like text, image, audio and video. Therefore there is a chance in Online Social Networks (OSNs) of posting unwanted content on particular public/private areas, called in general walls. Here we can use an information filtering approach to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted messages [1]. OSN do not support any content based preferences to avoid these unwanted content display on user wall. For example, Facebook permits users to decide who is allowed to insert messages in their private walls (i.e., friends, defined groups of friends or friends of friends). Though, there is no content-based partialities are preserved and therefore it is not possible to prevent undesired communications, for instance vulgar or offensive ones, no matter of the user who posts them.

The aim of the present work is therefore to propose and implement an automated system, called Filtered Wall (FW).Filtered Wall is an OSN filter which is, able to filter unwanted messages from OSN user walls. This paper mainly focusesfiltering of unwanted text and image in a message from one user to any other user. Here we are exploiting Machine Learning (ML) text categorization techniques to automatically assign with each short text message a set of categories based on its content in support content based filtering for texts messages and we employ an image comparison technique for image filtering of unwanted images.

The remainder sections of this paper are organized as follows: Section 2 Literature survey, whereas Section 3 introduces the architecture of the proposed system, Section 4 Short text classifier, Section 5 Content based message filtering, Section 6 Black List, Section 7 Image Comparison and Section 8concludes the paper

II. LITERATURE SURVEY

Our goal is to design an online message filtering system that can filter unwanted messages from OSN user wall. The main part the system is an information filter which discards the unwanted information.There are number of applications use the concept of information filtering. In recent years Recommender systems [1] have become popular which is a type of information filtering system that predicts the preference of user.it give importance to user interest and recommends an item. Recommender system can works in two ways

- Collaborative filtering
- Content based filtering

A. Collaborative filtering

Collaborative filtering system mainly based on user's preferences, actions and predicts what users will like based on his similarities to other users. User likes and dislikes determine the item rating [2]. A collaborative filtering involves collaboration of multiple agents while filtering information and it requires large dataset. Cold Start, Sparsity, First Rater and Popularity Bias are problems related to collaborative filtering.

B. Content based filtering

Content based filtering focus on user interest and select items based on it. It suggests the best matched item based on previously chosen item. In content based system each user act as independently. In content based filtering there is no Cold Start, Sparsity and First rater problem. There are some disadvantages also present in content based filtering such as, it requires contents that can be encoded as meaningful features and user taste must be represented as learnable function of the content features.

Text classification is an important part of content based filtering. Content based filtering actually works well on the Machine learning based text classifiers. In a Machine learning approach learns from training data and creates

Manuscript received March, 2014.

***Sruthi.T**, Computer Science and Engineering, KMCT College of Engineering, Calicut, India, +9496050407.,*

***Greeshma.T.R**, Computer Science and Engineering, KMCT College of Engineering, Calicut, India,+9846086510.,*

ISSN: 2278 – 7798

classifiers for the classification of new dataset. The main task of text classification is to assign each text predefined category of text. The classification algorithms such as Support Vector Machines, Naive Bayes, Neural network, and Decision trees can be used for text classification.

The Support vector machines [3] are based on the Structural Risk Minimization principle and supervised learning model .SVM analyze and recognize patterns in the input and it's able to perform linear as well as nonlinear classification. SVM Classifier suitable for small amount of labeled and large amount of unlabeled data. The SVM Classifier is well suited for text classification because of it having high dimensional input space, irrelevant features, sparse document vectors and linearly separable text classification .

Naive Bayes classifier probabilistic classifier and it is based on Bayes theorem .Naive Bayes classifier use Bayes rule for calculation of probability.Probabilistic models, especially the ones based on the Naive Bayes theory, are the state of the art in text classification and in almost any automatic text classification task.[4]

Neural network converting an input vector into output .In which neurons are arranged in a layer. The multilayer feed forward network is most commonly used one in which a unit feeds its output to all the units of the next layer but there is no feedback to the previous layer. In Radial basis function network (RBFN)which is an artificial neural network which uses radial basis function as an activation function. The output of RBFN is a linear combination of radial basis functions of the inputs and neuron parameters [4].

Decision trees [4] act as a classifier by hierarchical decomposition of the data space. It works by means of determining the predicate or a condition depending on attribute value. Classification uses the class labels in leaf node. Pruning isrequired in decision tree in order to reduce over fitting of data. To personalize access to the websiteJ. Golbeck Offered an algorithm for trust computation called Tidal Trust which provide a trust value between [1,7] .

III. PROPOSED SYSTEM

In general, the architecture in support of OSN services is a three-tier structure (Figure. 1). These three layers are

- Social Network Manager (SNM)
- Social Network Application (SNA)
- Graphical User Interface (GUI)

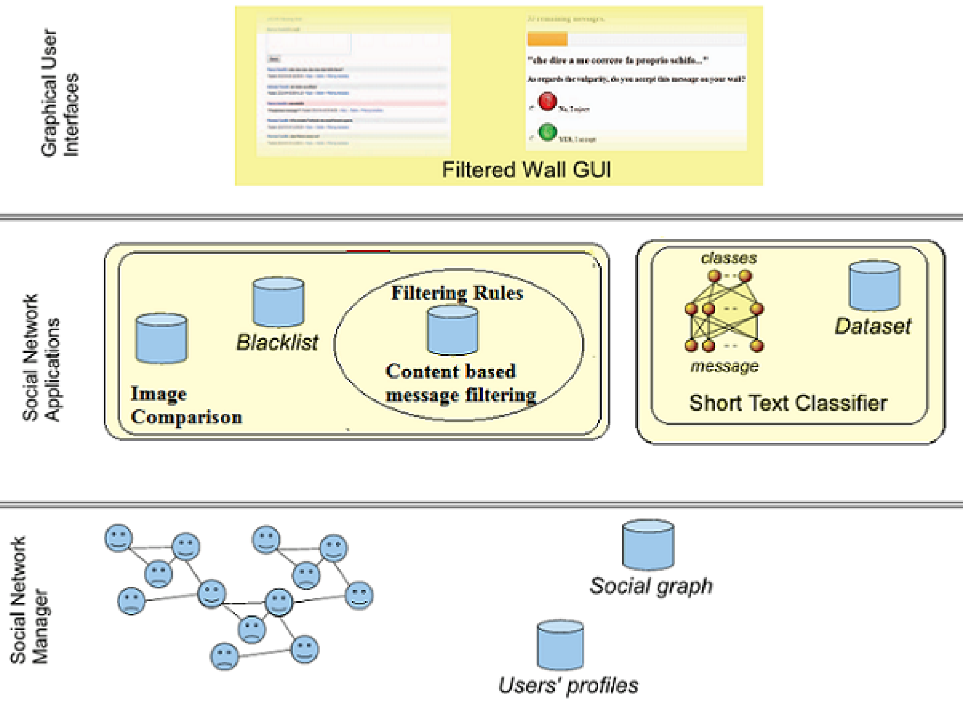

Figure 1: Filtered Wall Conceptual architecture

Social Network Manager (SNM) is the first layer, which is commonly aims to provide the basic OSN functionalities such as profile and relationship management. And it maintains the data related to user profile and provides the data to the second layer for applying filtering rules (FR) and blacklist. The second layer is Social Network Application (SNA) which provides support for external social network applications. SNA composed of Black List, Content Based Message Filtering (CBMF) [5], Image comparison and Short Text Classifier (STC) modules [6]. The STC classify the messages based on its contents CBMF filter messages According to filtering criteria. Graphical User Interface (GUI) is the third layer through which user provide the input and is able to see published messages. In addition to this GUI provide the facility to apply filtering rules and black list for the user in order to avoid unwanted message display. According to this architecture the system is placed in second and third layer.

Now we can consider the flow of messages follow, from writing to publication when it when it go through Filtered Wall.

1. A User entering the private wall of one of his/her contacts and then try to post some messages(text/image) which is intercepted by FW
2. Check whether the message creator is belongs in the BL or not ,if he/she is in BL immediately block the message without considering the contents of the message else go to next step
3. The message is passed to either STC or Image comparison module based on its format. The STC extract content of the message or the Image comparator check the image is same or not with the images that already in database
4. FW uses metadata provided by the STC togetherwith data extracted from the social graph and users' profiles, to enforce the filtering criteria or result from comparator
5. FW publish or filter the message based on previous step

IV. SHORT TEXT CLASSIFIER

The short classification module [7] composed of two main phases: Text representation and Machine Learning-based classification

A. Text representation

Text representation is a critical task because of it affects classification process.Many features are there for used in representation of text, but here we judge three types of features .we consider the two types of features, Bag of Words (BOW) and Document properties (DP), that are used in experimental evaluation to determine the combination that is most appropriate for short message classification are considered to be endogenous. Here we introduce Contextual Features (CFS) modeling information that are exogenous in nature and also characterizes the environment where the user is posting.

1. Correct words: It states the amount of terms. Correct words will be calculated.
2. Bad words: comparison to the correct words will be evaluated. Collection of dirty words will be determined.
3. Capital words: It will say about the amount of words written in message. Percentage of words in capital case will be calculated.
4. Punctuations characters: Percentage of punctuation character over the total number of character will be calculated.
5. Exclamation mark: Percentage of exclamation marks over the total number of punctuation characters will be calculated.
6. Question marks : Percentage of question marks over the total number of punctuation character will be evaluated.

The definitions which are used for CFS also will be used for BOW so that CFS and BOW are almost similar.

B. Machine Learning Based classification

In this section we can use any one the machine learning based text classification method which is best for short text classification. Here we suggest a Multilabel classifier based on Bayesian network.

1. Counting number of words
   The word counter algorithm implemented should find out the number of words (short texts) in the message.
2. Stop Word Removal Process
   After find out the number of short text we remove the stop words associated with message. In this step we reduce the content size but improve the quality of classification process when all the stop words present in the document are removed.
3. Removal of Special Character
   After Stop word removal we go to the process removing Special characters such as ".","?"etc.Which again reduce the size of message that is it reduces number of shorttexts. This improves the quality of STC.
4. Removal of Repeated Words:
   After the removal of Stop words ,Special characters we perform the removal of repeated or duplicate words this also increase the efficiency of Short Text Classifier.in this we keep frequency of occurrence of removed text for future use in computation probability of occurrence.
5. Multilabel classification
   Here we are introducing a Multilabel classifier based on concept of Bayes theorem .It canperform automatic multi labeling of messages .It can be easily implemented and it is very fast.

V. CBMF

CBMF is content based message filtering which consist of following sections.

A. Filtering Rules

Rules (FRs) are rules by which users can state what contents should not be displayed on their wall.

Definition 1(Creator Specification)

ISSN: 2278 – 7798

A creator specification denote as creator Spec which is a set of OSN users. Creator Spec can have one of the following forms, or it may be possibly combined:

1. A set of attribute constraints of the form an OP av where an is a user profile attribute name, av is a profile attribute value and OP is a comparison operatorrespectively, compatible with an domain.
2. A set of relationship constraints of the form (m; rt; minTrust; maxTrust), m denoting the OSN user who specify the rule within a relationship of type rt, having a depth greater than or equal to minTrust, and a trust value less than or equal to maxTrust.

Example 1: The creator specification CS1 = {Sex= Male, Age<18} Denotes all the males whose age is less than 18 years, whereas the creator specification CS2= {Alice; friends of; .1;.4} denotes all the users who are friends of Alice and whose trust level is less than or equal to 0.4. Finally, the creator specification CS3= {(Alice; friends of; 2; 0.4; Sex =male} selects only the male users from those identified by CS2

Definition 2 (Filtering rule)

A filtering rule FR is a tuple (author, creatorSpec, contentSpec, action)

- author is the user who specifies the rule;
- CreatorSpec is a creator specification, according to Definition 1;
- ContentSpec is a Boolean expression denoted as(c,ml) which is defined, where C is a class of the first or second level and ml is the minimum membership level threshold required for class C to make the constraint satisfied.
- action € {block notify} denotes the what action to be performed by the system when users identified by CreatorSpec.And the system on the messages matching contentSpec.

B. Online setup assistant for FRs threshold

OSA presents the user with a set of short text from messages that are selected from the dataset. For each message, the user tells the system the decision to accept or reject the message. OSA collect and processing the user decisions to get the minimum Membership level (ml) threshold required for class C.

Example 2: ((Bob; friend of; .2;.7), (vulgar; 0.80), block)

This example of FR show that Bob is an OSN user and he wants to block messages having a high degree of vulgar content. The threshold representing the user attitude for the vulgar class is obtained through OSA session is 0.8. Now the messages from user having relationship type "friend of "whose trust value is within the limit and whose messages are contains more than .80 of vulgar content is blocked

C. Trust Computation

Online Social Networks (OSNs) considered the trust as the assurance and confidence that information, people, behaves in expected way. In OSN trust may beMachine to machine, machine to human or human to human. At a deeper leveltrust is very much important in case of security or privacy in OSNs.

There are many algorithms are available for trust computation and which is used by different sites according to their trust value needs [8]. Here we are introducing a new algorithm for computing trust between users within an OSN based on the result from Filtered Wall. The basic concept of trust computation is that initially there exist a definite trust value between two users which isbased on the relationship type the trust value changes according to the filter result.

VI. BLACK LIST

Black List is list of user who are temporarily preventing from posting messages to wall of other user. The users whose messages are prevented independent from their contents called BL users. A Filtered wall user can add and remove another user contact in his/her to BL based on the result of filter.

VII. IMAGE COMPARISON

In this section we use a suitable image comparison technique for implementing image filtering module in the Filtered Wall. Here we a classic or default image check.

Classic Image Check

The classic or default image check compares the color value of every single expected and actual pixel. If at least one expected pixel differs from the actual pixel the check fails.

VIII. CONCLUSION

In this paper, we propose a new system called Filtered Wall that allows OSNs users to have a direct control on the messages posted on their private walls by means of a flexible rule based system.In addition to this a OSNs user can enhance the flexibility of the system through filtering rules, blacklist management and an image comparison technique. The system allows the users to customize their walls through the application of filtering criteria and a Machine Learning (ML) based soft classifier automatically labeling messages in support of content-based filtering. Here flexibility of the system can be enhanced through filtering rules and blacklist management. Our proposed system gives security to the Online Social Networks by means of preventing display of unwanted text and image contents. As a future work we are proposing an idea of content based image filtering in large scale based on similarity between images used as in Online Social Networks.

REFERENCES

[1] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, a Moreno Carullo, "A System to Filter Unwanted Messages from OSN User Walls", 2013.
[2] A. Adomavicius and G.Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the -Art and Possible Extensions, "IEEE Trans. Knowledge and Data Eng., vol.17, no.6, pp.734-749, June 2005
[3] 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.
[4] F.Sebastiani, "Machine Learning 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", 2010
[6] B.Sriram, D.Fuhry, E.Demir, H.ferhatatosmanoglu, and M.Demirbas, "Short Text Classification in Twitter to Improve Information Filtering," Proc.33rd Int'l ACM SIGIT Conf. Research and Development in Information Retrieval(sIGIR '10), pp.841-842,2010.
[7] V.Bobicev and M.Sokolova, "An Effective and Robust Method for Short Text Classification," Proc.23rd Nat'l Conf. Artificial Intelligence (AAAI), D.Fox and C.P.Gomes, eds., pp.1444-1445, 2008.
[8] J. A. Golbeck, "Computing and applying trust in web-based social networks," Ph.D. dissertation, PhD thesis, Graduate School of the University of Maryland, College Park, 2005.

Sruthi.T Pursuing M.Tech in Computer Science and Engineeringfrom KMCT College of Engineering, Calicut.Completed B.Tech in Information Technology from AWH Engineering College, kuttikattoor in the year 2012. Research area is Knowledge and Data Engineering.
Greeshma.T.R Working as Assistant. Professor in KMCT College of Engineering, Calicut. Completed M.Tech in Computer Science and Engineering from MES College of Engineering, kuttippuram. Research Area is Image Processing.

ISSN: 2278 – 7798