



## Implementation of Efficient Filtering System to Filter Unwanted Messages in Online Social Network

Rajshri Nikam\*, Prof. Deepak Gupta

Department of Computer Engineering  
Siddhant College of Engineering, Pune  
Savitribai Phule Pune University  
Maharashtra, India

---

**Abstract**— OSN's (Online Social Networks) are very important platform for human in today's life. Online Social Networks are used to share information, audio or video and also to share images interaction. Users can also post the messages on the wall of OSN. These posted text may contain some unwanted words like vulgar one and related to politics which shows some political or religious view. Due to this the environment gets disturbed. In order to avoid all these things, a system is proposed in which the posted unwanted message is filtered according to its content and if the message is normal then only it gets posted on the user's wall otherwise it gets filtered by the system. Because of this system, user becomes able to get control on the messages which are posting on their wall. If this happens frequently then that the unwanted user's may get blacklisted for specific time period due to which the user becomes unable to post a message on user wall. For that system text classification and filtering rules are used. Also if users post unwanted messages more than five times then they get blocked permanently automatically.

**Keywords**— Blocking, Filtering, Online Social Networking, User Wall.

---

### I. INTRODUCTION

Now days, people communicate each other due to OSN. OSN is online social networking. From the past few years it becomes a very fast, useful and efficient way to remain in contact with each other. By using OSN, user can share a lot of information related to their personal lives. They can also share their information related to their relationship with each other. The shared information may contain some text, images, audio or video messages. Also the information may contain some data which may be public or private related to any specific user. If we considered the Facebook in case of an OSN then while posting the messages on other user walls, may contain some which are wrong in nature, or they might show some political, or violence view. Because of this the environment gets disturbed.

In order to avoid all these things, we have proposed a system which filters out these undesirable messages. Using this system, user becomes able to restrict the unwanted messages which are going to post on their private wall called user wall. For that text classification and different filtering rules are taken into consideration.

In this system, whenever user posts the message on other user walls, that message gets classified into different classes by using machine learning classification for that RBFN algorithm is used. The main efforts are taken in building Short Text Classifier (STC) which is used for selection and extraction of data from the posted message. After that, Filter rules are applied on that messages. Filter rules are the rules are which are totally set by user and set different by different users[2]. If the

Facebook is considered then friends, friends of friends are come under trust relationship criteria. A threshold is calculated and according to that if the message is normal then only it gets posted on user wall otherwise gets filtered out by the system. A blacklist (BL's) is maintained to block the misbehaving users temporary or permanently. In case of temporary blacklisting, the users become unable to post on other user walls for some specific time period. While in case of permanent blacklisting, the user gets blocked permanently by the system [4].

### II. RELATED WORK

The main aim of filtering system is to filter unwanted content from input dataset before post on OSN user wall. This paper is the design of a system providing customizable content-based message filtering for OSNs, based on ML techniques. We are the first put to such kind of application for OSNs. However, our work has relationships both with the state of the art in content-based filtering, as well as with the field of policy-based personalization for OSNs and, more in general, web contents. Therefore, follows, we survey the literature in both these fields.

#### A. Content-Based Filtering

Each user is assumed to operate independently in content-based filtering a In content-based filtering system selects information item based on the correlation between the content of the items and the user preferences as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences. Documents processed in content-based filtering are mostly textual in nature and this makes content-based filtering close

to text classification. Content-based filtering use the ML paradigm. In which a classifier is automatically induced by learning from a set of preclassified examples. Some experiments prove that Bag-of-Words (BoW) approaches produce good performance and prevail in general over more sophisticated text representation that may have superior semantics but lower statistical quality [6], [7], [8]. The application of content-based filtering on messages posted on OSN user walls poses additional challenges given the short length of these messages other than the wide range of topics that can be discussed. Short text classification has received up to now little attention in the scientific community. Recent work indicate difficulties in defining robust features, essentially due to the fact that the description of the short text is concise, with many misspellings, nonstandard terms, and noise. Zelikovitz and Hirsh [14] attempt to improve the classification of short text strings developing a semi-supervised learning strategy based on a combination of labelled training data plus a secondary corpus of unlabeled but related longer documents.

**B. Policy-Based Personalization of OSN Content**

Now days there have been propose classification mechanisms for personalizing access in OSNs. This method has been proposed to categorize short text messages in order to avoid overwhelming users of micro blogging services by raw data .This system described in [9] focuses on Twitter2.In contrast, Golbeck and Kutersystemsdo not provide a filtering policy layer by which the user can exploit the result of the classification process to decide how and to which extent filtering out unwanted information. In our filtering policy language allows FRs according to a variety of criteria that do not consider only the results of the classification process but also the relationships of the wall owner with other OSN users as well as information on the user profile. In our system exploiting a flexible mechanism for BL management that provides a further opportunity of customization to the filtering procedure.

Our work is also inspired by the many access control models and related policy languages and enforcement mechanisms that have been proposed so far for OSNs (see[5] for a survey), since filtering shares several similarities with access control. A hierarchical two level classification is used in short text classification. In first level of a classifier labels the message into neutral and non-neutral. In second level of a classifier labels the messages into non neutral.

**III. PROPOSED SYSTEM**

The main purpose of this proposed work is to develop a system which filters out unwanted messages from user wall in case of Online Social Networking. This is done using different algorithms. The architecture of the system is shown in fig.1.

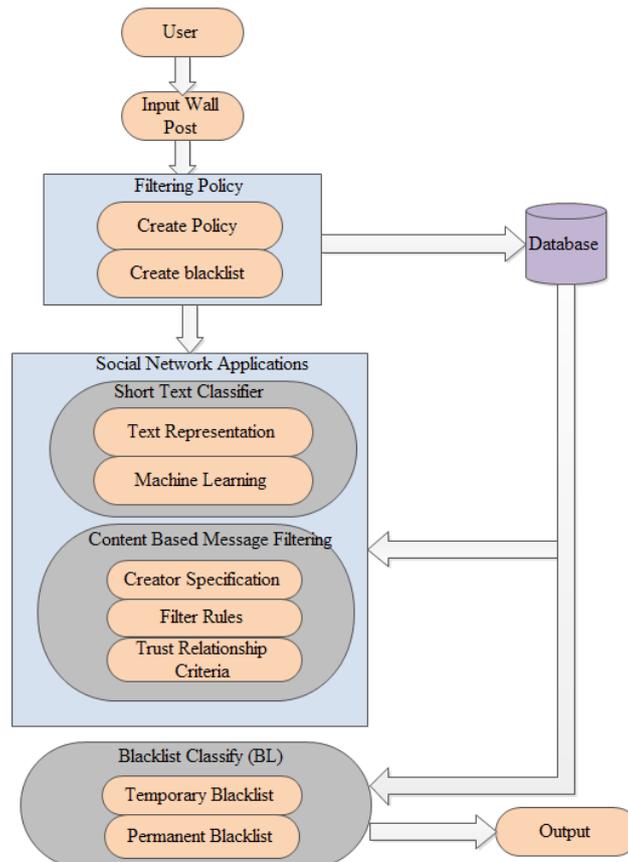


Fig 1: Proposed System

First of all user will enter the other user’s wall and post the message. After that, Machine learning classifier divides that message and classify into predefined classes. The filtering rules applied on that data. By using these criteria’s it is found that the posted message is normal then only it get posted on other user wall otherwise get filtered. If user post unwanted messages and images more than three times then they get automatically blocked permanently.

#### IV. IMPLEMENTATION

##### A. Module 1: Classification of posted text message

###### 1) Short Text Classifier

A hierarchical two level classification is used in short text classification. The first level of a classifier labels the message into neutral and non-neutral. In second level of a classifier labels the messages into non neutral.

###### 2) Text Representation

Text representation of a given document is important task strongly affecting the performance of classification process. It is done by extracting features for a given document. The investigation from [4][8][9] suggest three types of features important for text representation. Which are Bag of Words, Document properties (Dp) and Contextual Features (CF). The first two types of features are entirely derived from the information contained within the text of the message [4] whereas contextual features are exogenous. Representing text using endogenous. Terms are identified with words in Bag of Words representation. It is also important to use Feature which is extracted from outside the message content but related to message itself. A contextual feature is introduced in [4] that characterize the environment where the user is posting. According to [4]. Vector space Model is the model of text representation by which a text document is represented as a vector of binary or real weights. These three features are experimentally evaluated for short text classification in [9] for their appropriateness.

###### 3) Machine Learning-based Classification

As short text classification is hierarchical two level tasks and it should be robust to outliers hence RBFN is used for short text classification. An RBFN model is chose from the experimental evaluation in [11] among the other classifiers.

##### B. Module 2 : Used Filter Rule for classified message

After classification of posted message in different classes filter rules are applied. These are the rules set by user itself. These rules may be different for different users. The user may change the rules from user to user. e.g. if there are three users like Alice, Bob and Demon then the filter rules set by Alice for Bob may be different for that of Demon.

##### C. Module 3: Permanent blocking of unwanted user

If user post unwanted messages more than five times then they get automatically blocked permanently.

#### V. RESULTS AND DISCUSSIONS

The Bar graph in fig. 2 Shows the percentage Occurrences of the predefine class. A sample dataset is provided to system according to that probability for each class is calculated in percentage. The pre-defined classes are considered like Political, Offensive Vulgar, Violence, and Hate.

In this bar graph all the pre-defined classes are considered on X-axis and the probability value in percentage is considered on Y-axis.

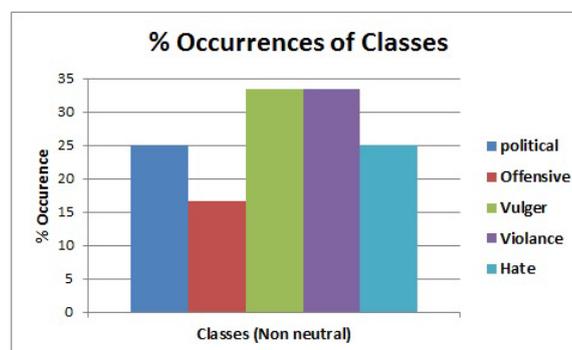


Fig 2: Percentage Occurrences of class

#### V. CONCLUSION

From this project, it is concluded that a better system can be implemented in case of OSN. Using this system, it is possible to control unwanted messages posted on other user walls. This can be done using different filtering rules Whenever user posts the message and image on other user walls at that time if that message satisfies all the criteria then only it get posted on other user walls otherwise it get filtered by the system itself. It is also possible to block the unwanted user temporarily so that he becomes unable to post on other user walls for a specific time period. If still the unwanted user not behaves properly then it may be the possibility that he/she get permanently blocked by the system itself.

#### ACKNOWLEDGMENT

I am very thankful to my guide for guiding me and heartily thankful to IJARCSSE for giving me such a wonderful opportunity for publishing my paper.

#### REFERENCES

- [1] J. Golbeck, "Combining Provenance with Trust in Social Networks for Semantic Web Content Filtering," Proc. Int'l Conf. Provenance and Annotation of Data, L. Moreau and I. Foster, eds., pp. 101-108,2006.

- [2] M. Vanetti, E. Binaghi, E. Ferrari, B. Carminati, and M. Carullo,, “A System to Filter Unwanted Messages from OSN User Walls”, IEEE Transactions on Knowledge and data Engineering, Vol. 25, No. 2, February 2013.
- [3] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari, “Content-Based Filtering in On-Line Social Networks,Proc.ECML/PKDD Workshop Privacy and Security Issues in Data Mining and Machine Learning (PSDML '10), 2010.
- [4] P. Tsang, A. Kapadia, “Nymble: Blocking Misbehaving Users in Anonymizing Networks”, IEEE Transaction on Dependable and Secure Computing, Vol. 8, No. 2, March-April, 2011.
- [5] R.J. Mooney and L. Roy, “Content-Based Book Recommending Using Learning for Text Categorization”, Proc. Fifth ACM Conf. Digital Libraries, pp. 195-204, 2000.
- [6] R. Schapire and Y. Singer, “Boostexter: A Boosting-Based System for Text Categorization”, Machine Learning, vol. 39, nos. 2/3, pp. 135-168, 2000
- [7] S. Dumais, J. Platt, D. Heckerman, and M. Sahami, “Inductive Learning Algorithms and Representations for Text Categorization,” Proc. Seventh Int'l Conf. Information and Knowledge Management (CIKM '98), pp. 148-155, 1998
- [8] S. Kotsiantis, V. Tampakas, “Increasing the accuracy of Hidden Naïve Byes model”, Advanced Information Management and Service (IMS), 6th International Conference, IEEE, pg. 247-252, 2010.
- [9] T. Joachims, “Text Categorization with Support Vector Machines: Learning with Many Relevant Features”, Proc. European Conf. Machine Learning, pp. 137-142, 1998.