FUHM: A SYSTEM TO FILTER UNDESIRED HINDI MESSAGES FROM OSN USERS WALL

[1]SAVITHA MATHAPATI, [2]GAJANAND SHARMA, [3]S H MANJULA, [4]VENUGOPAL K R

Department of Computer Science and Engineering, University Visvesvaraya College of Engineering, Bangalore University, Bengaluru, India.
E-mail: hiremathsavitha@gmail.com

Abstract- Online Social Networks (OSN) are popularly used for sharing data, information or knowledge among the people having similar interests. Todays OSN like, Facebook classifies the messages based on sender's relationship with the receiver. The challenge is to give OSN users the ability to filter a message posted on their own private space based on its content, like blocking a political message. In this paper, we propose Filtering Undesired Hindi Messages (FUHM) Algorithm that allows users to define their own filtering criteria for messages posted on their wall. This is attained by a rule based system and a text classifier. Time taken in classification process adds to the efficiency of overall system. Proposed algorithm is time efficient as compare to previously available methods as we use multiclass classification, classifying mes-sage into different categories in one step.

Index Terms- Content based Filtering, Message Filtering, Naive Baye's Classifier, Online Social Network, Text Classification.

I. INTRODUCTION

Online Social Networks play an important role in building social link among people having common interest. Here people share or post interesting news, photos, audio, video and their feelings on different topics. Variety and volume of messages posted on the wall are huge and all messages posted on a particular user's wall may not be useful as the interest differ among users. Solution to this problem is to give users the ability to block the messages that they are not interested in. This prevents user's wall overwhelmed by unwanted data. Facebook provides the users a mechanism to control what kind of contents can be posted by specifying who are allowed to post on their wall, e.g., friends, family, mutual friends and group of friends. This does not stop undesired/unwanted messages to be posted on their wall. Hence Information Filtering mechanism is required to automatically control the undesired messages from getting posted on the users wall. Content-Based filtering provides the users an ability to filter unwanted messages depending on the content, such as political or offensive.

Motivation: Vanetti et al., [1] proposed a system (FUM) called Filter Unwanted Messages from User Wall that provides the users a flexibility to filter undesired messages that are in

Italian language from getting posted on their wall. Filtering model uses two level hierarchical strategy for content based filtering. This helps identifying and filtering the undesired messages from the private space of an OSN user. Labeling a message with a category in one step and again labeling it with a sub category in the next step takes more time as compared to time taken by a classifier performing whole process in one step.

Contribution: In this paper, we propose automated system called "A System to Filter Undesired Hindi Messages from OSN Users Wall" (FUHM). The aim of our work is to build a filter wall that filters undesired messages in Hindi. We use single-level classification over two level classification, where messages are classified into different categories in one stage. This reduces the time taken to classify messages improving the time complexity.

The paper is organized as follows. Survey of related work is discussed in section 2. Section 3 gives the problem definition of the proposed work. Machine learning based classification method is discussed in section 4. The process of Filtering Rules and Blacklist Management is illustrated in section 5. Algorithm to filter undesired hindi messages from OSN user wall is discussed in section 6. Performance evaluation of the system is analyzed in section 7. Finally, section 8 gives the conclusions of the paper.

II. RELATED WORK

Content Based Message Filtering (CBMF) system for online Social Network(OSN) is mainly based on Machine Learning approach. Our work is related to both Content-Based Filtering and the policy based personalization for On-line Social Net-works. We survey the different related works on both of these areas.

A. Content-Based Filtering

Information producer generate information and send to the other users without considering whether it is liked or required by the user or not. Filtering system is required to classify these stream of information and post only those type of messages that satisfy the user preference. Content based Message Filtering (CBMF) selects information item based on the interrelationship between user's preferences and the

content of the information item [2]. Most of the time, CBMF is used in filtering textual documents that make it more reliable for text filtering. Earlier work of information filtering was on e-mail and later work continued on internet news articles, newswire articles and network resources.

Content-Based filtering system uses Machine Learning based algorithms. Here classifier is designed using some preclassified messages. Lot of work has been done on different methods of feature extraction, model learning and collection of samples [3]. In Feature extraction, text messages are mapped to a compact representation of its content and then devoted for training the classifier. Number of experiments prove that Bag-of Words approach gives a good performance and prevail over more sophisticated text representation that may have superior semantics but lower statistical quality [4]. Sebastiani et al., [5] analysed different methods of learning model and concluded that Boosting-based classifier, Support Vector Machine and Neural Networks outperform other methods. Boosting-based classifier used BoosTexter algorithm for text categorization technique. Support Vector Machine classify text using hyper-plane that classify the different classes [6].

B. Policy-Based Personalization

Now a days, as there is huge amount of data available in On-line Social Network, personalized access is taking importance. Sriram B et al., [7] proposed a method where each tweet is associated with predefined categories that decribe evey tweet contents. Based on his / her interest, user read certain type of tweets. Golbeck [8] proposed Film Trust, an application that combines original information and OSN trust relationships to calculate the trust value of each user on relationship graph. Drawback of such system is, they do not provide users an ability to vary the level of filtering process. This drawback is taken care in our system by providing users to set up Filtering Rules according to variety of criteria.

In the field of Online Social Networks, effective and usable access control frame works are mainly required. OSN's policy combines the individual choices of users into filtering rules. L Kagal et al., [9] worked on the system where semantic web automate their discovery and communication, because they must autonomously conclude what information is to be exchanged and how. Topology-Based Acces control is mainly used in access control models. In these models, requirements of access control are the relationships that the user have with the user who is requesting. Same method is used to find the users to whom Filtering rules applies [10]. P Bonatti et al., [11] introduced a concept called Protune, a trust negotiation frame work features a powerful declarative metalanguage for driving some critical negotiation decisions to monitor negotiations.

III. PROBLEM DEFINITION

The objective of the proposed work is to develop an automated system, that is able to identify and filter undesired messages from OSN user wall based on its content and user preferences. This is achieved using Naive Bayes classifier to attain content based filtering along with a flexible rule based
system. OSN users have direct control on the message content appearing on their wall. In this work we use a single step classification that classifies the input message with efficient time complexity.

IV. MACHINE LEARNING-BASED CLASSIFICATION

For content based filtering of text messages, we propose a single level classification process. The classifier performs the classification and assigns one of the five categories to each message. The categories defined are neutral, political, vulgar, offensive and violence. For probabilistic classification of text messages, we choose Naive Bayes Classifier, because building this classifier model is fast. Another advantage of Naive Baye's classifier model is that, it is easy to train with new data whenever required without rebuilding the model. Na¨ıve Bayes Classifier is a probability based classifier that uses Bayes theorem along with naiıve (strong) independence assumptions between the features. Naive Bayes Classifier is a well-known classifier used for text categorization i.e., labeling a given document or text with a category name. It uses bag-of-words model where a sentence or text document as bag of its words. The bag-of-word model doesn't consider the grammar or word order, insted it keeps the word count in track. The word frequency is taken as the feature in order to train a classifier.

The performance of overall classification process is highly affected by the pre-classified data that is used for training the classifier. To get good performance, the classifier is trained with adequately complete, noiseless and uniform data. A particular document is classified into different categories ac-cording to the decision taken by the experts who are manually classifying the text. The training data is given to 6 experts from different fields and asked to classify that data into above mentioned five categories. After this manual classification done by all these experts, the majority of categories assigned to a text or document is taken as the actual label for that text or document.

We now formally describe the strategy behind the overall classification. Let is the set of five categories i.e., neutral, political, vulgar, offensive and violence. Each message that is posted to any user's wall or personal inbox is assigned to one of these five categories. Each message m_i of the supervised collected set of messages, D_H is defined in Equation 1.

$$D_H = \{(m_i, y_i) \ldots (m_{|D_H|}, y_{|D_H|})\} \quad (1)$$

D_H is composed of the text message m_i and its respective label y_i will be {0, 1} for all categories, indicating the belongingness to each of the defined category. We are going to partition the set D_H into the training dataset T r_D and the test dataset T e_D.

Let C_r be the classifier and y_i is the belongingness to the different categories. Learning phase work as follows:

1) Features x_i are extracted from each message m_i. $T r_D$ and $T e_D$ are then converted into $Tr = \{(x_i, y_i) \ldots (x_{|T rD|}, y_{|T rD|})$ and $Te = \{(x_i, y_i) \ldots (x_{|T eD|}, Y_{|T eD|}\}$ respectively.

2) A multiclass training set Tr = $\{(x_j, y_j)$ belongs to Tr / (x_j, y_j), $y_{jk} = y_{jk+1}$, k = 1…|Ω| is created for C_r.

3) Classifier C_r is trained with the Tr messages and per-formance of the C_r is evaluated using test set Te.

V. FILTERING RULES AND BLACKLIST MANAGEMENT

In this section, we explain the rule service designed to filter unwanted messages. There are two parts of the rule service. First we discuss the filtering rules and later we describe the purpose of blacklists.

The social network is considered as a directed graph, where each node represents a user of social network and the edge connecting any two nodes represents the relationship between users. Each edge is having a label associated to it, that represent the type of the relationship (friend, teacher, colleague, family member, mutual friends etc..) between these two users. Each type of relationship is having a predefined trust value that indicates how trustworthy the user considers another user with whom he / she is going to establish the relationship. When the user selects a relationship type while adding another user as a friend, the trust value associated to that relationship type is assigned to the edge between these two users. For different type of relationships, we are taking a number in the range [0, 1] as the trust value. From this explanation we say that, in social graph of network, there exits an edge connecting any two users with labels R_T and T. This represents a direct relationship of type R_T and trust value T between these two users. On the other hand, two users are having an indirect relationship of type R_T if there exists a path consisting of more than one edge connecting these two users in the equivalent social graph of the network.

A. Filtering Rules

While defining the structure of filtering rules, we have to consider an issue that affect the process of message filtering. The issue that affect the classification of message filtering is, one message may have different meanings and interpreted in different ways based on the sender of message. We highly focus on giving users the ability to create some rules and constraints on the message senders. The sender on whom the filtering rules are imposed are selected based on some criteria. A very common and important criteria is to impose constraints on the attributes of user's profile. Using this, a user define rules applicable on the senders of a specific age group or senders belonging to a particular gender. The filtering rules are defined based on the information gathered from the social graph such as depth value, trust value or type-of the relationship. All the above mentioned options are formulated as a sender specification, defined as follows:

Definition 1 (Sender Specification) A sender specification senderSpec indicates a set of social network users. senderSpec have any one of the forms described below or a combination of these forms.

corresponding to the attribute and OP is comparison operator that is compatible with the domain of attrname.

2) A set of constraints applicable on user relationship type of the form (x, R_T , trustVal, depthVal). It denotes all the users of the social network who are interacting with user x having relationship of type R_T and having a trust value greater than or equal to trustVal and depth value lesser than or equal to depthVal.

Another important requirement that is fulfilled by the Fil-tering Rules is to support the specifications of content-based filtering of the text messages. For content based filtering, we are using a single level classification strategy. We are able to identify the possible category of the message (i.e., the category for the message that has highest degree of membership). Along with this idea, we provide users the ability to manually define a threshold value for each of the defined category. A category is associated with the message only if the membership level of message to particular category is greater than the user defined threshold. User define a Filtering Rule to filter out the messages belonging to a particular category out of defined categories i.e., neutral, political, vulgar, offensive and violence. Above mentioned idea is formulated as a Filtering Rule, defined as follows:

Definition 2 (Filtering Rule). A filtering rule FR is defined as a tuple creator, senderSpec, messageSpec, action, where

1) creator is the network user who defines the rule.

2) senderSpec is the sender specification defined in Definition 1.

3) messageSpec is a Boolean expression that is defined on the constraints of message content. The content constraint is of the form (C, cutoff) where C is the category that the message is assigned and cutoff is the minimum user defined cutoff value for category C to satisfy the constraint.

B. Blacklists

Another important component is a blacklist mechanism that is used to restrict the messages sent by undesired senders. Blacklist mechanism is controlled and managed by the system where the OSN user has to insert the user id of their friends from whom he / she do not want to receive any message. User block a particular sender for any time duration and after blocking, sender is not able to post / send any content on the user wall. This provide flexibility to the user to create the blacklist rule according to his / her choice. The blacklist rules work like above described filtering rules. They help users to identify undesired senders and block them for a specific time period. We have introduced an automatic blocking process that is based on the principle that if a user is sending an objec-tionable content (the content belonging to a category blocked by the receiver) to a particular user more than a predefined threshold , it indicates that he / she is very interested in that type of content. Thus it is advisable to block that user for a

1) A set of constraints applicable on user profile attributes considerable amount of time. To implement this feature, we of the form (attrname OP attrval), where attrname stands used number of failure attempts (when sender tried to send a for attribute name of user's profile, attrval is the value message and it is filtered by filtering rules defined by receiver)

as a key parameter. It assists the system to identify the sender whose messages are continuously failing the receiver defined filtering rules. After stopping the objectionable content to be sent, the system maintains a counts of such attempts. If the count exceeds the predefined threshold value, system automatically add the sender in receiver's blacklist. Thus the blacklist is defined in both ways, by the user himself / herself or automatically by the system. Hence a blacklist rule is defined as follows-

Definition 3 (Blacklist Rule) A blacklist is defined as a tuple (creator, sender, senderBehavior, TimePeriod) where-

1) creator is the network user who specifies the blacklist rule.

2) sender is the network user who sends or tries to send a
message to the creator.

3) senderBehavior is examined based on number of attempts
the sender's messages fails the filtering rules defined by the
receiver. It is defined in the form (C, attemptCount), where C
is the non-neutral category that the message is assigned to
and it is already blocked by the receiver and attemptCount is
number of attempts sender tries to send the message.

4) TimePeriod is the number of days the sender is blocked to send any message to particular receiver.

VI. FILTERING UNDESIRED HINDI MESSAGES FROM THE OSN USER WALL

In this section, we are going to describe the Filter Unwanted Hindi Messages (FUHM) algorithm to filter undesired mes-sages from the OSN user wall. When a user, say 'X' tries to send a message to another user to say 'Y' then, first we need to check whether user 'Y' wants to receive messages from user 'X' or not. If user 'Y' is not interested to receive any message from user 'X', then he has to block user 'X'. The first thing we have to check is whether 'Y' is blocked by the user'X'. For this purpose we have to scan the blacklist of user 'Y' i.e., BL(Y). If BL(Y) contains user id of user 'X', it means 'Y' has blocked 'X'and he / she no longer wish to receive any kind of message from 'X'. In this case any further step is not required and the process is stopped. On the other hand, if BL(Y) doesn't contain user id of 'X' then we scan the rule database of 'Y', $DB_r(Y)$ to check if there is any rule defined by 'Y'. If $DB_r(Y)$ is non-empty, it means there are some rules defined in the rule database of user 'Y'. Then we need to check whether user 'X' passes all these rules or not. The rules is defined based on age, gender of sender or trust value, depth value of the relationship they share. If user 'X' fails any of the rule then he / she need to be blocked to share any kind of content with 'Y'. If user 'X' passes all the rules or the rule database of 'Y' is empty, then the system goes to process further steps.

After passing the rule service, the message is taken as input. The trained classifier read the message, tokenize it and classify into five defined categories. A probability distribution is calculated for each message how strongly it belongs to a particular category. The category that is assigned with highest probability is taken as the message category i.e., CAT(M).

If CAT(M) is neutral then need not to do any further step and the message is sent to the receiver 'Y'. If CAT(M) is non-neutral, means out of offensive, violence, political or vulgar, then the system scans the blocked category list of user 'Y', $BC_L(Y)$. There are three possibilities now-

1) $BC_L(Y)$ is empty:

In this case, user 'Y' has not blocked any category of non-neutral messages. So system has nothing to do. It proceeds for further steps. The message is delivered to user 'Y'.

2) $BC_L(Y)$ is non-empty but current message category is not present in it:
If $BC_L(Y)$ is not empty then it means user 'Y' has blocked some of the non-neutral message categories. But according to statement, current message category is not present in $BC_L(Y)$. It means user 'Y' does not want messages of that category to be blocked. So system proceeds to deliver the message to user 'Y'.

3) $BC_L(Y)$ is non-empty and current message category is present in it:
In this case, $BC_L(Y)$ contains the category of non-neutral message that the current message belongs to. It means user 'Y' does not want to receive message of that category.

Now again two possibilities are there based on the proba-bility distribution assigned to message i.e. probDist(M). The first possibility is the probability distribution of message i.e. probDist(M) is less than user defined cut-off, cutof f_C (Y). In this case, according to user 'Y', that message is not falling under that category and it is okay to receive such content. The message is delivered to user 'Y'.The second possibility is the probability distribution of message, probDist(M) is greater than or equal to user defined cut-off cutof f_C (Y). Then, according to user 'Y' that message is falling under that blocked category and user did not want to receive such content. So the message is blocked instantly.

After blocking every message, the system create a database entry $DB_E(X,Y)$, with the category of message. When count of database entry $DB_E(X,Y)$ exceeds 3, then the sender 'X' tried to send a content that the receiver 'Y' don't want to receive. It is considered that user 'X' is not respecting the privacy and feelings of user 'Y', system block the user 'X' for a period of 5 days and forbid him / her to send any message to 'Y'.

VII. PERFORMANCE EVALUATION

In this section, we describe the performance evaluation of this message classification and filtering modules. We start with the description of data set and further continue with the text classification process.

A. Description of Data Set

Tools for content based filtering of text documents or messages are available for some languages. These language specific tools are useful only for the respective languages. For these languages, the dataset is also made available by the researchers who developed these tools. While talking about

Algorithm 1: Filtering Undesired Hindi Messages

```
Input : Message from user X trying to post on wall of
        user Y
Output: Decision on posting the message or not on the
        wall of user Y
begin
    User 'X' logs into a social network and selects
    another user 'Y' to send a message
    if X 2 BL(Y) then
        Block the user 'X' from sending the message
    else if rule database DB_r(Y) is empty then
        Take message 'M' as input
    else if user 'X' does not pass all the rules defined in
    DB_r(Y) by user 'Y' then
        Block the user 'X' to send the message
        Goto End
    else
    Classify the message 'M' into five categories using
    classifier
    if message category 'CAT(M)' is 'neutral' then
        post the message
        Goto End
    else
    Scan the BC_L(Y) list of user 'Y'
    if BC_L(Y) list is empty then
        Post the message
    else if CAT(M)  BC_L(Y) then
        if probDist(M)  cutoffC(Y) then
            Block the message
            Goto End
        Post the message
    Create a database entry DB_E(X, Y) for each message
    'M' with its category, that failed to post
    Select count DB_E(X, Y) from database
    if count DB_E(X, Y) 3 in last 7 days from the
    current date then
        block the sender 'X' for 5 days to send message
        to receiver 'Y'
    End
```

TABLE I
COMPARISON BETWEEN TIME TAKEN BY TWO METHODS IN CLASSIFICATION PROCESS

Fraction of Dataset	20%	40%	60%	80%	100%
Time taken by FUHM in seconds	2.3066	3.9577	6.0836	8.7337	11.2625
Time Taken by FUM in seconds	2.4260	5.5558	10.1434	15.3925	20.1788

TABLE II
COMPARISON BETWEEN TIME TAKEN BY TWO METHODS IN CLASSIFICATION PROCESS FOR DIFFERENT CATEGORIES

Category of Message	Neutral	Offensive	Violence	Political	Vulgar
Time taken by FUHM method in seconds	3.3778	3.3209	3.4023	3.3565	3.3674
Time Taken by FUM Method in seconds	4.3330	5.7271	5.7709	5.7148	5.7475

Hindi language, it is one of the most popular language, that is used for communication by nearly 490 million people around the globe. We planned to implement such tool for Hindi lan-guage to filter out unwanted messages based on their content. To solve the problem of unavailability of Hindi dataset, we created a dataset D_H with considerable amount of data. While assigning the category i.e., Λ = {neutral, political, vulgar, offensive, violence} to individual messages, we considered the fact that a message is assigned to multiple categories based on the context, classifier's domain and interests. Thus it creates intrinsic ambiguity sometimes whether to assign a message to different categories. To solve this problem, we created a survey and asked six different experts to label the messages into one of the five available categories. The category for a particular message was then chosen according to majority labeling basis. Hence for a message M_i, if majority of experts assigns category A_i belongs to $\{0, 1\}^{|\Lambda|}$, then message M_i, is taken under category A_i. The experts are chosen keeping in mind that person's work domain, religion, gender and other factors that affect the decision making process. Thus six experts has been chosen from different domains having different religious background and educational status. This is to ensure the best category for a message. A message has been considered in neutral category if message is not falling under any of the defined non-neutral category, like political, offensive, violence and vulgar. We say that neutral = Λ- {political, vulgar, offensive, violence}.

B. Evaluation of Text Classifier

In our experiment, we considered a set of five Hindi mes-sage categories. They are neutral, political, vulgar, offensive, violence. Total messages in dataset DH contain 20% each of neutral, political, vulgar, offensive and violence message categories. We randomly select a set of messages and given as input to the "A System to Filter Unwanted Messages from OSN User Wall" (FUM) [1] and proposed method (FUHM). Simulation results of the (FUM) method and proposed method(FUHM) for different fraction of dataset(20%, 40%, 60%, 80%, 100%) is tabulated in Table 1. It depicts that for

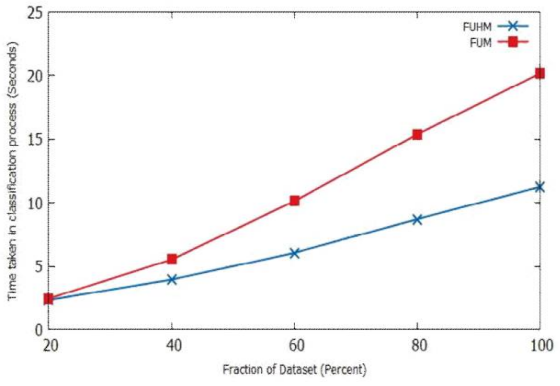


Fig. 1. Comparison between Time Taken by Two Methods in Classification Process

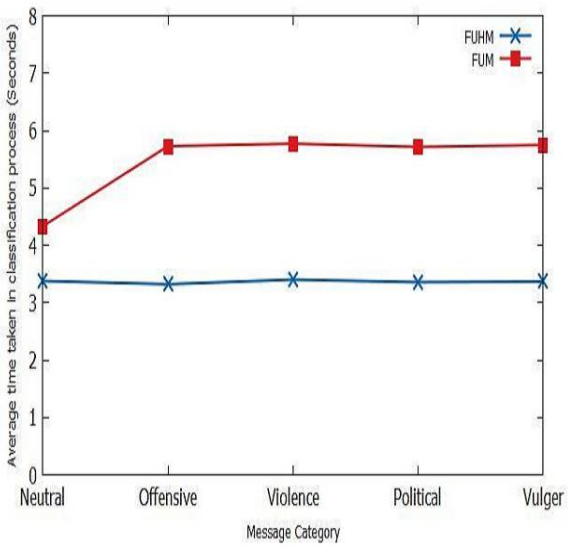


Fig. 2. Comparison between Time Taken by Two Methods in Classification Process for Different Categories

each fraction of dataset, FUHM method takes less time as it classify messages in single step compared to FUM [1] method that takes two steps to classify messages as shown in Figure 1. Hence FUHM perform better compared to FUM method.

Another comparison we performed is between the time taken to classify the messages of different categories. We randomly selected five messages of each category, making a total of 25 messages. These messages are given as the input to both the classifiers, base method classifier and proposed method classifier. Table 2 shows the time taken in classification of different messages belonging to different categories. Time shown here is the average time taken to classify five messages of a particular category. From Figure 2, we clearly analyze the time taken by the classifiers. The time taken for classification by the proposed FUHM method is almost straight that depict

that for each category (Neutral, Offensive, Vio;ence, Political and Vulgar) the classifier is taking almost same amount of time. Table 2 lists the time taken for the classification of different categories for both FUM and FUHM method. Time taken to classify is more for FUM while classifying non-neutral categories compared to FUHM method. This is because in FUM, first message is classified into neutral and non-neutral and further classification is to perform non-neutral message into four sub categories. If the message falls under neutral category, in that case we are not moving to second level classification. So time taken in classification process is less in this case. On the other hand if message falls under non-neutral category then second level classifier classify the message in four sub categories of non-neutral category. In this two level classification process(FUM), the time taken by classifier is slightly more compared to proposed FUHM method.

CONCLUSIONS

In this paper, we presented an automated system to remove unwanted messages of Hindi language getting delivered to OSN users. We used a machine learning based soft classifier to apply user defined content-based filtering rules. Proposed single level classification of FUHM method reduced the time taken to classify the non-neutral messages. When 100% of the Dataset is used for training the classifier, there is 55% saving of time in training the classifier using proposed FUHM method compared to FUM method. We gave the flexibility to define, change and remove the filtering rules along with blacklist management.

REFERENCES

[1] 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, pp. 285–297, 2013.

[2] 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.

[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. ACM, pp. 195–204, 2000.

[4] C. Apte,´ F. Damerau, and S. M. Weiss, "Automated Learning of Decision Rules for Text Categorization," ACM Transactions on Information Systems (TOIS), vol. 12, no. 3, pp. 233–251, 1994.

[5] F. Sebastiani, "Machine Learning in Automated Text Categorization," ACM Computing Surveys (CSUR), vol. 34, no. 1, pp. 1–47, 2002.

[6] T. Joachims, "Text Categorization with Support Vector Machines: Learn-ing with Many Relevant Features," European Conference on Machine Learning, pp. 137–142, 1998.

[7] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short Text Classification in Twitter to Improve Information Filtering,"

[8] Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 841–842, 2010.

[9] J. Golbeck, "Combining Provenance with Trust in Social Networks for Semantic Web Content Filtering," in International Provenance and Annotation Workshop, Springer, pp. 101–108, 2006.

[10] L. Kagal, T. Finin, M. Paolucci, N. Srinivasan, K. Sycara, and G. Denker, "Authorization and Privacy for Semantic Web Services," IEEE Intelli-gent Systems, vol. 19, no. 4, pp. 50–56, 2004.

[11] F. Bonchi and E. Ferrari, "Privacy-aware Knowledge Discovery: Novel Applications and New Techniques," 2010.

[12] P. Bonatti and D. Olmedilla, "Driving and Monitoring Provisional Trust Negotiation with Metapolicies," Constraints, vol. 14, p. 15, 2005.

★★★