A System to Filter Unwanted Messages from Social Network User Walls

Sheetal M. Rajput[1], Nilima P. Patil[2]

PG Student, Department of Computer Engineering, SSBT's COET, Bambhori, Jalgaon, India [1]

Assistant Professor, Department of Computer Engineering, SSBT's COET, Bambhori, Jalgaon, India [2]

Abstract: Social Network is one of the most popular interactive medium to share, communicate and distribute information. It is the platform to build social relations among people. Users have ability to keep in touch with friends by exchanging different types of information or messages. Sometimes people post messages which may cause a serious problem to user's reputation. There is one fundamental issue in Social Network is to give users ability to control the messages posted on their own private space. To avoid the problem of unwanted messages, the proposed system called filter wall able to filter messages from user walls. In order to make social network user wall a secured wall, introducing a rule based system wall as well as Machine Learning approach, neural text classification applied on user wall.

Keywords: On-line Social Networks, Filtered Wall, Machine Learning, Short Text Classifier.

I. INTRODUCTION

Social Networks (SNs) are today one of the most popular interactive medium to communicate, share, and disseminate a considerable amount of human life information. Daily and continuous communications imply the exchange of several types of content, including free text, image, audio, and video data. Online Social Networks provide very little support to prevent unwanted messages on user walls. For example, Facebook allows users to state who is allowed to insert messages in their walls (i.e., friends, friends of friends, or defined groups of friends) [3][4]. To prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them. Sometimes people post messages which may cause a serious problem to user's reputation. So there is one fundamental issue in Online Social Network is to give users ability to control the messages posted on their own private space.

An automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls. It exploits Machine Learning (ML) text categorization techniques to automatically assign with each short text message. Providing this service is not only a matter of using previously defined web content mining techniques for a different application, rather it requires to design classification strategies. The major efforts in building robust short text classifiers (STC) are concentrated in the extraction and selection of a set of characterizing and discriminates features. The solutions investigated are an extension of those adopted in a previous work from which it inherits the learning model and the elicitation procedure for generating pre classified data. Besides classification facilities, the system provides a powerful rule layer to specify Filtering Rules (FRs), by which users can state what contents, should not be displayed on their walls. FRs

can support a variety of different filtering criteria that can be combined and customized according to the user needs. In addition, the system provides the support for user defined BlackLists (BLs), that is, lists of users that are temporarily prevented to post any kind of messages on a user wall.

A. Problem Statement

The most popular interactive medium to communicate, share, and disseminate a considerable amount of human life information is Online Social Networks. In Social Networks, information filtering can also be used for different, more sensitive, purpose. Sometimes people post messages which may cause a serious problem to user's reputation. There is one fundamental issue in Social Network is to give users ability to control the messages posted on their own private space. To avoid the problem of unwanted messages, the proposed system called filter wall able to filter messages from user walls. In order to make social network user wall a secured wall, introducing a rule based system wall as well as Machine Learning approach, neural text classification applied on user wall.

II. RELATED WORK

This work has relationship both with the state of the art in content-based filtering, as well as with the field of policy based filtering.

A. Content-Based Filtering

Information filtering systems are designed to classify a stream of dynamically generated information dispatched asynchronously by an information producer and present to the user those information that are likely to satisfy his/her requirements [1].

In content-based filtering, each user is assumed to operate independently. As a result, a content-based filtering system selects information items 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. While electronic mail was the original domain of early work on information filtering, subsequent papers have addressed diversified domains including newswire articles, Internet news articles, and broader network resources. Documents processed in content-based filtering are mostly textual in nature and this makes content-based filtering close to text classification. The activity of filtering can be modelled, in fact, as a case of single label, binary classification, partitioning incoming documents into relevant and nonrelevant categories. More complex filtering systems include multilabel text categorization automatically labelling messages into partial thematic categories.

Content-based filtering is mainly based on the use of the ML paradigm according to which a classifier is automatically induced by learning from a set of preclassi fied examples. A remarkable variety of related work has recently appeared which differ for the adopted feature extraction methods, model learning, and collection of samples. The feature extraction procedure maps text into a compact representation of its content and is uniformly applied to training and generalization phases. Several experiments prove that Bag-of-Words (BoW) approaches yield good performance and prevail in general over more sophisticated text representation that may have superior semantics but lower statistical quality. As far as the learning model is concerned, there are a number of major approaches in content-based filtering and text classification in general showing mutual advantages and disadvantages in function of application dependent issues. A detailed comparison analysis has been conducted con firming superiority of Boosting-based classifiers, Neural Networks, and Support Vector Machines over other popular methods, such as Rocchio and Nave Bayesian. However, it is worth to note that most of the work related to text filtering by ML has been applied for long-form text and the assessed performance of the text classification methods strictly depends on the nature of textual documents.

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 few attentions in the scientific community. Recent work highlights 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 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. This solution is inapplicable in our domain in which short messages are not summary or part of longer semantically related documents. A different approach is proposed by Bobicev and Sokolova that circumvent the problem of error-prone feature construction by adopting a statistical learning method that can perform reasonably well without feature engineering. However, this method, named Prediction by Partial Mapping, produces a language model that is used in probabilistic text classifiers which are hard classifiers in nature and do not easily integrate soft, multi-membership paradigms. In our scenario, we consider gradual membership to classes a key feature for defining flexible policy-based personalization strategies [1].

B. Policy-Based Personalization of OSN Contents

There have been some proposals exploiting classification mechanisms for personalizing access in OSNs. For instance, a classification method has been proposed to categorize short text messages in order to avoid overwhelming users of micro blogging services by raw data. The system described focuses on Twitter2 and associates a set of categories with each tweet describing its content. The user can then view only certain types of tweets based on his/her interests. In contrast, Golbeck and Kuter propose an application, called FilmTrust, that exploits OSN trust relationships and proveance information to personalize access to the website. However, such systems do 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. The filtering policy language allows the setting of 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. Moreover, system is complemented by a flexible mechanism for BL management that provides a further opportunity of customization to the filtering procedure [1] [2].

III. PROPOSED SOLUTION

The aim of the present work is therefore to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from user walls. Machine Learning (ML) text categorization techniques (Neural Text Classifier) used to automatically assign with each short text message. The major efforts in building a robust short text classifier are concentrated in the extraction and selection of a set of characterizing and discriminate features. The original set of features, derived from endogenous properties of short texts, is enlarged here including exogenous knowledge related to the context from which the messages originate. In the current paper the use of neural learning which is today recognized as one of the most efficient solutions in text classification.

A. Filtering Rules

In user walls like in everyday life, the same message may have different meanings and relevance based on who writes it. As a consequence, FRs should allow users to state constraints on message creators. Creators on which a FR applies can be selected on the basis of several different criteria; one of the most relevant is by imposing conditions on their profile's attributes. In such a way it is, for instance, possible to define rules applying only to young creators or to creators with a given religious/political view. Given the social network scenario, creators may also be identified by exploiting information on their social graph. This implies to state conditions on type, depth and trust values of the relationship(s) creators should be involved in order to apply them the specified rules.

B. Blacklist Mechanism

A further component of system is a BL mechanism to avoid messages from undesired creators, independent from their contents. Blacklists are directly managed by the system, which 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. To enhance flexibility, such information is given to the system through a set of rules, hereafter called BL rules. The wall's owners to specify BL rules regulating who has to be banned from their walls and for how long. Therefore, a user might be banned from a wall, by, at the same time, being able to post in other walls.
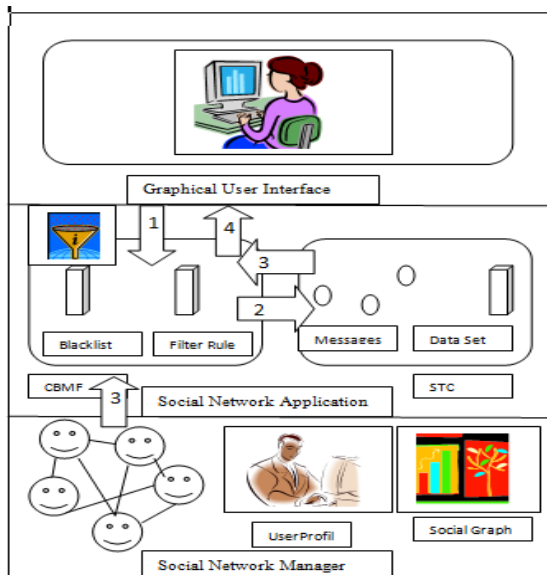
C. System Architecture

Fig. 1: Architecture diagram

The architecture in support of OSN services is a three tier structure .The first layer, called Social Network Manager (SNM), commonly aims to provide the basic OSN functionalities (i.e., profile and relationship management), whereas the second layer provides the support for external Social Network Applications (SNAs). The supported SNAs may in turn require an additional layer for their needed Graphical User Interfaces (GUIs). According to this reference architecture, the proposed system is placed in the second and third layers. In particular, users interact with the system by means of a GUI to set up and manage their FRs/BLs. Moreover, the GUI provides users with a FW, that is, a wall where only messages that are authorized according to their FRs/BLs are published. The core components of the proposed system are the Rule Based Messages Filtering and the Short Text Classifier modules. The latter component aims to classify messages according to a set of categories. In contrast; the first component exploits the message categorization provided by the Short Text Classifier module to enforce the FRs specified by the user. BLs [4] can also be used to enhance the filtering process. Below shows the steps for message display:

1) After entering the private wall of one of his/her contacts, the user tries to post a message, which is intercepted by FW.
2) A ML-based text classifier extracts metadata from the content of the message.
3) FW uses metadata provided by the classifier, together with data extracted from the social graph and users' profiles, to enforce the filtering and BL rules.
4) Depending on the result of the previous step, the message will be published or filtered by FW.

D. Text Classification Process

1. Preprocessing

Data mining is the process of extracting hidden pattern in a large dataset. Real world data is often incomplete inconsistent and lacking in certain behaviour and is likely to contain many errors. Data goes through a series of steps:

- Data cleaning: data is cleansed through processes such as filling in missing values, smoothing the noisy data, or resolving the inconsistencies in data.
- Data integration: data with different representation are put together and conflicts within the data are resolved.
- Data transformation: data is normalized aggregated and generalized
- Data reduction: this step aims to present a reduce representation of data
- Data discretization: involves the reduction of a number of values of a continuous attribute by dividing the range of attribute intervals.

2. Stop word removal

It is well recognized among the information retrieval experts that a set of functional English words (eg. the, a and that) is useless as indexing terms. These words have very low discrimination value, since they occur in every English document. Hence they do not help in distinguishing between documents with contents that are about different topics. The process of removing the set of

non content bearing functional words from the set of words produced by word extraction is known as stop words removal. In order to remove the stop words, this involves first creating a list of stop words to be removed, which is also called the stop word list. After this, the set of words produced by word extraction is then scanned so that every word appearing in the stop list is removed [6][7].

3. Porter stemming algorithm

Porter stemming algorithm is a process for removing the commoner morphological ending words in English. Rules in porter stemming algorithm are separated into distinct steps

- Gets rid of plurals and -ed or -ing.
- Turns terminal y to i when there is another vowel in the stem.
- Maps double su ces to single ones.
- Deals with ic-, full, ness etc.
- Takes o ant, ence .

4. Feature extraction/selection

Classification Feature extraction / selection helps identify important words in a text document. This is done using methods like TF-IDF (term frequency-inverse document frequency), LSI (latent semantic indexing), multi-word etc. In the context of text classification, features or attributes usually mean significant words, multi-words or frequently occurring phrases indicative of the text category.

IV. CONCLUSION

In this paper, present a system to filter unwanted messages from OSN walls. The system develops a ML soft classifier to implement customizable content-dependent FRs. In particular, it aim at investigating a tool able to automatically recommend trust values for those contacts user does not individually identified.

Consider that such a tool should propose expectation assessment based on users procedures, performances, and reputation in OSN, which might involve enhancing OSN with assessment methods. Though, the propose of these assessment based tools is difficult by several concerns, like the suggestions an assessment system might have on users confidentiality and/or the restrictions on what it is possible to audit in present OSNs. An introduction work in this direction has been prepared in the context of expectation values used for OSN access control purposes. The system proposed in this paper represents just the core set of functionalities needed to provide a sophisticated tool for OSN message filtering. To address this problem by investigating the use of online learning paradigms able to include label feedbacks from users .To enhance this system with a more sophisticated approach to decide when a user should be inserted into a Black Lists or not.

REFERENCES

[1] Macro Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati and Moreno carullo , A System to Filter Unwanted Messages from OSN User Walls, IEEE Transactions on Knowledge and Data Engineering. Vol 25 , No. 2. February 2013.

[2] Sujapriya, Immaanual Gnana Drai, Dr.C.Kumar ,Charlie Paul, Filtering Unwanted Messages from Online Social Networks Using rule based technique,IOSR-JCE,Vol.16 ,No. 66,Jan.2014.

[3] Taeho JO, NTC (Neural Text Categorizer): Neural Network for Text Categorization, IJIS Val.2, No. 2, April 2010.

[4] Anuratha Patra and Divakar Singh, Neural Network Approach for Text Classification using Relevance Factor as Term Weighing Metho, IJCA val.68, No. 17 April 2013.

[5] M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari,Content-based filtering in on-line social networks, in Proceedings of Workshop on Privacy and Security issues in Data Mining and Machine Learning (PSDML 2010), 2010

[6] B.Sriram, D. Fuhry, E.Demir, H. Ferhatosmanoglu and M. Demirbas, Short Text Classification in twitter to improve information filtering, in processing of the International ACM SIGIR Conference on Research and development of information Retrieval, SIGIR 2010,2010,pp, 841-842.

[7] Mita K. Dalal and Mukesh A. Zaveri, Automatic Text Classification: A Technical Review, Vol. 2, No. 2 August 2011.

[8] 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.

[9] 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.