International Journal of Pharmaceutical Science Invention
ISSN (Online): 2319 – 6718, ISSN (Print): 2319 – 670X
www.ijpsi.org Volume 3 Issue 5 ǁ May 2014 ǁ PP.20-25

Filtered wall: A System to filter unwanted messages from OSN user wall using content based filtering

Miss Janhavi A. Patokar [1], Prof. V.T.Gaikwad [2]
Information Technology Department,
Sipna College of Engineering and Technology, Amravati.

ABSTRAC: *In today On-line Social Networks (OSNs), one fundamental issue is to give users the ability to control the messages posted on their own private space to avoid that unwanted content is displayed. Up to now OSNs provide little support to this requirement. To fill the gap, we built a system which allows OSN users to have a direct control on the messages posted on their walls. This is achieved through a flexible rule-based system, that allows a user to customize the filtering criteria to be applied to their walls, and a Machine Learning based soft classifier automatically labelling messages in support of content-based filtering.*

KEYWORDS: *On-line Social Networks, Information Filtering, Short Text Classification, Policy-based personalization.*

I. INTRODUCTION

In the last years, On-line Social Networks (OSNs) have become a popular interactive medium to communicate, share and disseminate a considerable amount of human life information. Daily and continuous communication implies the exchange of several types of content, including free text, image, audio and video data. The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data and then provide an active support in complex and sophisticated tasks involved in social networking analysis and management. A main part of social network content is constituted by short text, a notable example are the messages permanently written by OSN users on particular public/private areas, called in general walls[1].The aim of the present work is to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter out unwanted messages from social network user walls. The key idea of the proposed system is the support for content- based user preferences. On-line Social Networks (OSNs) 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.

According to Facebook statistics 1 average user creates 90 pieces of content each month, whereas more than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) are shared each month. The huge and dynamic character of these data creates the premise for the employment of web content mining strategies aimed to automatically discover useful information dormant within the data. They are instrumental to provide an active support in complex and sophisticated tasks involved in OSN management, such as for instance access control or information filtering. Information filtering has been greatly explored for what concerns textual documents and, more recently, web content. However, the aim of the majority of these proposals is mainly to provide users a classification mechanism to avoid they are overwhelmed by useless data. In OSNs, information filtering can also be used for a different, more sensitive, purpose. This is due to the fact that in OSNs there is the possibility of posting or commentingother posts on particular public/private areas, called in general walls. Information filtering can therefore be used to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted messages. We believe that this is a key OSN service that has not been provided so far. Indeed, today OSNs 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). However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them.

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 ad-hoc classification strategies. This is because wall messages are constituted by short text for which traditional classification methods have serious limitations since short texts do not provide sufficient word occurrences.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 OSN user walls. We exploit Machine Learning (ML) text categorization techniques [2] to automatically assign with each short text message a set of categories based on its content. To the best of our knowledge this is the first proposal of a system to automatically filter unwanted messages from OSN user walls on the basis of both message content and the message creator relationships and characteristics.

II. LITERATURE REVIEW

The main contribution of this paper is the design of a system providing customizable content-based message filtering for OSNs, based on ML techniques. As we have pointed out in the introduction, to the best of our knowledge we are the first proposing 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, in what follows, we survey the literature in both these fields.

2.1 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 [3]. 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. 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 modeled, in fact, as a case of single label, binary classification, partitioning incoming documents into relevant and non relevant categories [4]. More complex filtering systems include multi-label text categorization automatically labeling 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 pre-classified examples. A remarkable variety of related work has recently appeared, which differ for the adopted feature extraction methods, model learning, and collection of samples [5], [6], [7], [8],[9]. The feature extraction procedure maps text into a compact representation of its content and is uniformly applied to training and generalization phases. 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 attention 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, non standard terms and noise. Focusing on the OSN domain, interest in access control and privacy protection is quite recent. As far as privacy is concerned, current work is mainly focusing on privacy-preserving data mining techniques, that is, protecting information related to the network, i.e., relationships/nodes, while performing social network analysis [5]. Works more related to our proposals are those in the field of access control. In this field, many different access control models and related mechanisms have been proposed so far (e.g., [6,2,10]), which mainly differ on the expressivity of the access control policy language and on the way access control is enforced (e.g., centralized vs. decentralized). Most of these models express access control requirements in terms of relationships that the requestor should have with the resource owner. We use a similar idea to identify the users to which a filtering rule applies. However, the overall goal of our proposal is completely different, since we mainly deal with filtering of unwanted contents rather than with access control. As such, one of the key ingredients of our system is the availability of a description for the message contents to be exploited by the filtering mechanism as well as by the language to express filtering rules. In contrast, no one of the access control models previously cited exploit the content of the resources to enforce access control. We believe that this is a fundamental difference. Moreover, the notion of black- lists and their management are not considered by any of these access control models.

2.2 Policy-based personalization of OSN contents

Recently, there have been some proposals exploiting classification mechanisms for personalizing access in OSNs. For instance, in [11] a classification method has been proposed to categorize short text messages in order to avoid overwhelming users of microblogging services by raw data. The system described in

[11] 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 [12] propose an application, called FilmTrust, that exploits OSN trust relationships and provenance 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. In contrast, our 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, our system is complemented by a flexible mechanism for BL management that provides a further opportunity of customization to the filtering procedure. The only social networking service we are aware of providing filtering abilities to its users is MyWOT, a social networking service which gives its subscribers the ability to: 1) rate resources with respect to four criteria: trustworthiness, vendor reliability, privacy, and child safety; 2) specify preferences determining whether the browser should block access to a given resource, or should simply return a warning message on the basis of the specified rating. Despite the existence of some similarities, the approach adopted by MyWOT is quite different from ours. In particular, it supports filtering criteria which are far less flexible than the ones of Filtered Wall since they are only based on the four above-mentioned criteria. Moreover, no automatic classification mechanism is provided to the end user. 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, since filtering shares several similarities with access control. Actually, content filtering can be considered as an extension of access control, since it can be used both to protect objects from unauthorized subjects, and subjects from inappropriate objects. In the field of OSNs, the majority of access control models proposed so far enforce topology-based access control, according to which access control requirements are expressed in terms of relationships that the requester should have with the resource owner. We use a similar idea to identify the users to which a FR applies. However, our filtering policy language extends the languages proposed for access control policy specification in OSNs to cope with the extended requirements of the filtering domain. Indeed, since we are dealing with filtering of unwanted contents rather than with access control, one of the key ingredients of our system is the availability of a description for the message contents to be exploited by the filtering mechanism. In contrast, no one of the access control models previously cited exploit the content of the resources to enforce access control. Moreover, the notion of BLs and their management are not considered by any of the above-mentioned access control models.

III. ANALYSIS OF PROBLEM:

The use of effective and appropriate methods in facilitating projects enhances its effectiveness and efficiency. The method will be applied in system analysis and design method where an existing system is studied to proffer better options to solving existing problems. Indeed, today OSNs 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). However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones, no matter of the user who posts them. However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, no matter of the user who posts them. 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 ad hoc classification strategies. This is because wall messages are constituted by short text for which traditional classification methods have serious limitations since short texts do not provide sufficient word occurrences.

IV. PROPOSED WORK:

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 OSN user walls. We exploit Machine Learning (ML) text categorization techniques [2] to automatically assign with each short text message a set of categories based on its content. 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 solutions investigated in this paper are an extension of those adopted in a previous work [13] from which we inherit the learning model and the elicitation procedure for generating pre-classified data. 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. As far as the learning model is concerned, we confirm in the current paper the use of neural learning which is today recognized as one of the most efficient solutions in text classification [2]. In particular, we base the overall short text classification strategy on Radial Basis Function Networks (RBFN) for their proven capabilities in acting as soft classifiers, in managing noisy data and intrinsically vague

classes. Moreover, the speed 2 in performing the learning phase creates the premise for an adequate use in OSN domains, as well as facilitates the experimental evaluation tasks.

4.1 System Architecture:

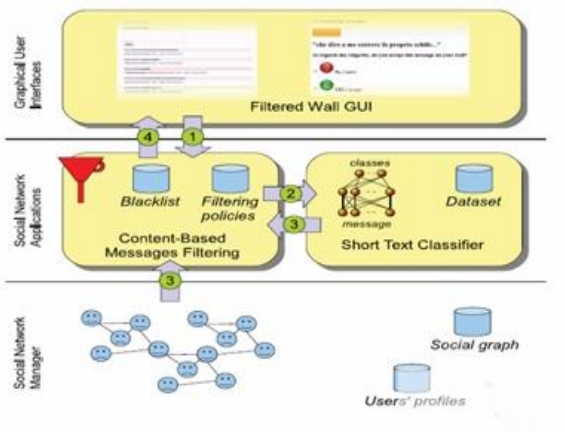

Fig 1: Filtered Wall Conceptual architecture

The architecture in support of OSN services is a three-tier structure (Figure 1). 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).4 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 Content-Based Messages Filtering (CBMF) and the Short Text Classifier (STC) 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 STC module to enforce the FRs specified by the user. BLs can also be used to enhance the filtering process. As graphically depicted in *Fig1*, the path followed by a message, from its writing to the possible final publication can be summarized as follows:

- Step 1- After entering the private wall of one of his/her contacts, the user tries to post a message, which is intercepted by FW.
- Step 2- A ML-based text classifier extracts metadata from the content of the message.
- Step 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.
- Step 4- Depending on the result of the previous step, the message will be published or filtered by FW.

4.2 Filtering rules:

In defining the language for FRs specification, we consider three main issues that, in our opinion, should affect a message filtering decision. First of all, in OSNs 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. In general, more than a filtering rule can apply to the same user. A message is therefore published only if it is not blocked by any of the filtering rules that apply to the message creator. Note moreover, that it may happen that a user profile does not contain a value for the attribute(s) referred by a FR (e.g, the profile does not specify a value for the attribute Hometown whereas the FR blocks all the messages authored by users coming from a specific city). In that case, the system is not able to evaluate whether the user profile matches the FR. Since how to deal with such messages depend on the considered scenario and on the wall owner attitudes, we ask the wall owner to decide whether to block or notify messages originating from a user whose profile does not match against the wall owner FRs because of missing attributes.

4.3 Online setup assistant for FRs threshold:

We address the problem of setting thresholds to filter rules, by conceiving and implementing within FW, an Online Setup Assistant (OSA) procedure. OSA presents the user with a set of messages selected from the dataset. For each message, the user tells the system the decision to accept or reject the message. The collection and processing of user decisions on an adequate set of messages distributed over all the classes allows to compute customized thresholds representing the user attitude in accepting or rejecting certain contents. Such messages are selected according to the following process. A certain amount of non neutral messages taken from a fraction of the dataset and not belonging to the training/test sets, are classified by the ML in order to have, for each message, the second level class membership values.

4.4 Blacklists:

A further component of our system is a BL mechanism to avoid messages from undesired creators, independent from their contents. BLs 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 users retention in the BL is finished. To enhance flexibility, such information are given to the system through a set of rules, hereafter called BL rules. Such rules are not defined by the SNM, therefore they are not meant as general high level directives to be applied to the whole community. Rather, we decide to let the users themselves, i.e., 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.Similar to FRs, our BL rules make the wall owner able to identify users to be blocked according to their profiles as well as their relationships in the OSN. Therefore, by means of a BL rule, wall owners are for example able to ban from their walls users they do not directly know (i.e., with which they have only indirect relationships), or users that are friend of a given person as they may have a bad opinion of this person. This banning can be adopted for an undetermined time period or for a specific time window. Moreover, banning criteria may also take into account users' behavior in the OSN. More precisely, among possible information denoting users' bad behavior we have focused on two main measures. The first is related to the principle that if within a given time interval a user has been inserted into a BL for several times, say greater than a given threshold, he/she might deserve to stay in the BL for another while, as his/her behavior is not improved. This principle works for those users that have been already inserted in the considered BL at least one time. In contrast, to catch new bad behaviors, we use the Relative Frequency (RF) that let the system be able to detect those users whose messages continue to fail the FRs. The two measures can be computed either locally, that is, by considering only the messages and/or the BL of the user specifying the BL rule or globally, that is, by considering all OSN users walls and/or BLs.

Advantages Of proposed System:

- ✓ A system to automatically filter unwanted messages from OSN user walls on the basis of both message content and the message creator relationships and characteristics.
- ✓ The current paper substantially extends for what concerns both the rule layer and the classification module.
- ✓ Major differences include, a different semantics for filtering rules to better fit the considered domain, an online setup assistant (OSA) to help users in FR specification, the extension of the set of features considered in the classification process, a more deep performance evaluation study and an update of the prototype implementation to reflect the changes made to the classification techniques.

V. APPLICATION:

- Filtering the unwanted messages enforces protection and productivity policies for businesses, schools, and libraries to reduce legal and privacy risks while minimizing administration overhead. Filtering provides network administrators with greater control by automatically and transparently enforcing acceptable use policies.
- The Content filtering is designed to control what content may or may not be viewed by a reader, especially when used to restrict material delivered over the Internet via the Web, e-mail, or other means. Restrictions can be applied at various levels:
- A government can attempt to apply them nationwide.
- An employer to its personnel.
- A school to its teachers and/or students.
- A library to its patrons and/or staff.
- An individual to his or her own computer.

VI. CONCLUSION

In this paper, we have presented a system to filter out undesired messages from OSN walls. The system exploits a ML soft classifier to enforce customizable content-depended filtering rules. Moreover, the flexibility of the system in terms of filtering options is enhanced trough the management of BLs. the system can automatically take a decision about the messages blocked because of the tolerance, on the basis of some statistical data (e.g., number of blocked messages from the same author, number of times the creator has been inserted in the BL) as well as data on creator profile (e.g., relationships with the wall owner, age, sex). As future work, we intend to exploit similar techniques to infer BL and filtering rules. As future work, we intend to exploit similar techniques to infer BL and filtering rules.

REFERENCES:

[1] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo Department of Computer Science and Communication University of Insubria 21100 Varese, Italy

[2] F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.

[3] N. J. Belkin and W. B. Croft, "Information filtering and information retrieval: Two sides of the same coin?" Communications of the ACM, vol. 35, no. 12, pp. 29–38, 1992.

[4] P. J. Hayes, P. M. Andersen, I. B. Nirenburg, and L. M. Schmandt, "Tcs: a shell for content-based text categorization," in Proceedings of 6th IEEE Conference on Artificial Intelligence Applications (CAIA- 90). IEEE Computer Society Press, Los Alamitos, US, 1990, pp. 320–326.

[5] G. Amati and F. Crestani, "Probabilistic learning for selective dissemination of information," Information Processing and Manage- ment, vol. 35, no. 5, pp. 633–654, 1999.

[6] A. 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.

[7] M. J. Pazzani and D. Billsus, "Learning and revising user profiles: The identification of interesting web sites," Machine Learning, vol. 27, no. 3, pp. 313–331, 1997.

[8] 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. New York: ACM Press, 2000, , pp. 195–204.

[9] Y. Zhang and J. Callan, "Maximum likelihood estimation for filtering thresholds," in Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2001, pp. 294–302.

[10] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demir- bas, "Short text classification in twitter to improve information filtering," in Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, 2010, pp. 841–842.

[11] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demir- bas, "Short text classification in twitter to improve information filtering," in Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR 2010, 2010, pp. 841–842.

[12] J. Golbeck, "Combining provenance with trust in social networks for semantic web content filtering," in Provenance and Annotation of Data, ser. Lecture Notes