A System Approach to Avoid Unwanted Messages from User Walls

Dipali D. Vidhate[1], Ajay. P. Thakare[2]

[1]Student, Information Technology, Sipna COET / Sant Gadage Baba Amravati University, India

[2]HOD, Electronic & Telecommunication, Sipna COET / Sant Gadage Baba Amravati University, India

Abstract: ***In recent years, Online Social Networks have become an important part of daily life for many. One fundamental issue in today user wall(s) 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 user walls provide little support to this requirement. To fill the gap, I propose a system allowing user wall users to have a direct control on the messages posted on their walls. This is achieved through a flexible rule-based system, that allows users to customize the filtering criteria to be applied to their walls, and Machine Learning based soft classifier automatically labeling messages in support of content-based filtering.***

Keywords: On-line Social Networks, Information Filtering, Short Text Classification, Policy-based Personalization

1. Introduction

In the last years, On-line Social Networks 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, and 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.

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. This is possible thank to the use of a Machine Learning (ML) text categorization procedure able [4] to automatically assign with each message a set of categories based on its content. We believe that the proposed strategy is a key service for social networks in that in today social networks users have little control on the messages displayed on their 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. For instance, it is not possible to prevent political or vulgar messages. In contrast, by means of the proposed mechanism, a user can specify what contents should not be displayed on his/her wall, by specifying a set of filtering rules. Filtering rules are very flexible in terms of the filtering requirements they can support, in that they allow to specify filtering conditions based on user profiles, user relationships as well as the output of the ML categorization process. In addition, the system provides the support for user defined blacklists, that is, list of users that are temporarily prevented to post messages on a user wall.

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. Major differences include a different semantics for filtering rules to better fit the considered domain, an online setup assistant 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.

2. Literature Review & Related Work

In 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? Works more related to their 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, 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. They use a similar idea to identify the users to which a filtering rule applies. However, the overall goal of their proposal is completely different, since they mainly deal with filtering of unwanted contents rather than with access control. As such, one of the key ingredients of their 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 exploits the content of the resources to enforce access control. They believe that this is a fundamental difference. Moreover, the notion of blacklists

and their management are not considered by any of these access control models.

Content-based filtering has been widely investigated by exploiting ML techniques as well as other strategies. However, the problem of applying content-based filtering on the varied contents exchanged by users of social networks has received up to now few attentions in the scientific community. The advantages of using ML filtering strategies over ad-hoc knowledge engineering approaches are a very good effectiveness, flexibility to changes in the domain and portability in different applications. This system providing customizable content-based approach to avoid unwanted messages from user wall, 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 user walls? However, their 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 Policy-based personalization

Recently, there have been some proposals exploiting classification mechanisms for personalizing access in user walls. For instance, in 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 in focuses on Twitter 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, an application, called Film Trust, 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.

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 user walls, 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 user walls, 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.

2.2 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[6]. 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[7], [8]. 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. More complex filtering systems include multi-label text categorization automatically labeling messages into partial thematic categories.

3. Analysis of Problem

The analysis of related work has highlighted the lack of a publicly available benchmark for comparing different approaches to content based classification of user walls short texts. To cope with this lack, we have built and made available a dataset D of messages taken from Facebook. Many messages from publicly accessible Italian groups have been selected and extracted by means of an automated procedure that removes undesired spam messages and, for each message, stores the message body and the name of the group from which it originates. The messages come from the group's web page section, where any registered user can post a new message or reply to messages already posted by other users. The group of experts has been chosen in an attempt to ensure high heterogeneity concerning sex, age, employment, education and religion. In order to create a consensus concerning the meaning of the Neutral class and general criteria in assigning multi-class membership we invited experts to participate to a dedicated tuning session.

We are aware of the fact that the extreme diversity of OSNs content and the continuing evolution of communication styles create the need of using several datasets as a reference benchmark. We hope that our dataset will pave the way for a quantitative and more precise analysis of user wall short text classification methods.

4. Limitation of Existing System

- However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages. No matter user who propose 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.

5. 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 user walls. We exploit Machine Learning (ML) text categorization techniques 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 by us from whom 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. 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.

6. 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 relationship and characteristics.
- The substantially extends for what concerns both the rule layer and the classification modules.

4.1 Complete word discovery algorithm pseudo code

A complete word is a complete substring of the collated text of the input text message, defined in the following way: Let T be a sequence of elements (t1, t2, t3 . . . tn). S is a complete substring of T when S occurs in k distinct positions p1, p2, p3 . . . pk in T.

In other words, a complete word cannot be extended by adding preceding or trailing elements, because at least one of these elements is different from the rest. Here are the whole word extraction phases as pseudo-code.

Step 1- discover right-complete word;
Step 2- discover left-complete word;
Step 3- sort the left-complete word alphabetically;
Step 4- combine the left- and right-complete words into a set of complete words.

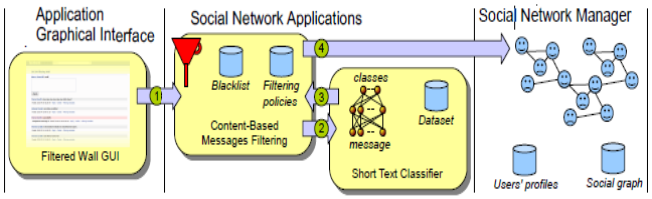

Figure 1: Filtered wall Conceptual Architecture

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. The strategy underlying this module is described in Section IV. 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 (see Section V for more details).

- **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 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. All these options are formalized by the notion of creator specification, defined as follows.

- **Online setup assistant for FRs thresholds:**

As mentioned in the previous section, 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 discussed in Section VI-A. 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 computing 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.

- **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 is 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.

- **Improve Short Text Classification**

We describe a method for improving the classification of short text strings using a combination of labeled training data plus a secondary corpus of unlabeled but related longer documents. We show that such unlabeled background knowledge can greatly decrease error rates, particularly if the number of examples or the size of the strings in the training set is small. This is particularly useful when labeling text is a labor-intensive job and when there is a large amount of information available about a particular problem on the World Wide Web. Our approach views the task as one of information integration using WHIRL, a tool that combines database functionalities with techniques from the information-retrieval literature.

- **Text Categorization**

The assignment of natural language texts to one or more predefined categories based on their content – is an important component in many information organization and management tasks. We compare the effectiveness of five different automatic learning algorithms for text categorization in terms of learning speed, real time classification speed, and classification accuracy. We also examine training set size, and alternative document representations. Very accurate text classifiers can be learned automatically from training examples. Linear Support Vector Machines (SVMs) are particularly promising because they are very accurate, quick to train, and quick to evaluate.

7. Conclusion

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. The future implication of this work is we exploit Machine Learning (ML) text categorization techniques to automatically assign with each short text message a set of categories based on its content.

References

[1] "A System to Filter Unwanted Messages from OSN User Walls" Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo, Department of Computer Science and Communication University of Insubria 21100 Varese, Italy IEEE Transactions On Knowledge And Data Engineering Vol:25 Year 2013
[2] "Content-Based Filtering in On-line Social Networks" M. Vanetti, E. Binaghi, B. Carminati, M. Carullo and E. Ferrari, Department of Computer Science and Communication University of Insubria 21100 Varese, Italy fmarco.vanetti, elisabetta.binaghi, barbara.carminati, moreno.carullo, elena.ferrarig @uninsubria.it
[3] 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.
[4] F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.
[5] 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.
[6] 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.
[7] P. J. Denning, "Electronic junk," Communications of the ACM, vol. 25, no. 3, pp. 163–165, 1982.
[8] 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.

Author Profile

Dipali D. Vidhate is Student of ME (IT) in Sipna College of Engineering & Technology, Amravati, India

Ajay P. Thakare is working as HOD of EXTC Department in Sipna College of Engineering & Technology, Amravati. India