AN ARRANGEMENT TO REFINE UNDESIRABLE INFORMATION FROM OSN USER WALLS

V. Basha[1], D. Bheekya[2]
[1]*M.Tech, CSE Dept, MLRIT, Hyderabad*
[2]*M.Tech, Asst. Professor, MLRIT, Hyderabad*

***Abstract*— Users should have control on the content displayed in the Online Social Network (OSN) to avoid unwanted data seen from other users. Now a day's OSNs administers insufficient support to this requirement. To fill the gap, in this paper, we propose a system allowing OSN users to have a direct control on the messages posted on their user accounts. This is conclude through a adjustable 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.**

***Index Terms*—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. 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. This is possible thank to the use of a Machine Learning (ML) text categorization procedure [12] able to automatically assign with each message a set of categories based on its content. 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.

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 [4] 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 discriminant features. The solutions investigated in this paper are an extension of those adopted in a previous work by us [5] 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 [4]. 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 in performing the learning phase creates the premise for an adequate use in OSN domains, as well as facilitates the experimental evaluation tasks.

FRs can support a variety of different filtering criteria that can be combined and customized according to the user needs. More precisely, FRs exploit user profiles, user relationships as well as the output of the ML categorization process to state the filtering criteria to be enforced. In addition, the system provides the support for user-defined Black Lists (BLs), that is, lists of users that are temporarily prevented to post any kind of 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. The current paper substantially extends [5] 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 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.

II. PROBLEM STATEMENT

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, Face book 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 [4] 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 [5] 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 [4]. 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.

III. SYSTEM DEVELOPMENT

The implementation stage involves careful planning, investigation of the existing system and it's constraints on implementation, designing of methods to achieve changeover and evaluation of changeover methods.

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

B. 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 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.

C. Blacklists

A further component of our system is a BL mechanism to avoid messages from undesired creators, independent from their contents. BLs is 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 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.

IV. RELATED WORK

However, the problem of applying content-based filtering on the varied contents exchanged by users of social networks has received up to now little attention in the scientific community. One of the few examples in this direction is the work by Boykin and Roychowdhury [3] that proposes an automated anti-spam tool that, exploiting the properties of social networks, can recognize unsolicited commercial e-mail, spam and messages associated with people the user knows. However, it is important to note that the strategy just mentioned does not exploit ML content-based techniques. 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 differ applications. However difficulties arise in finding an appropriate set of features by which to represent short, grammatically ill formed sentences and in providing a consistent training set of manually classified text.

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 [4]. Work more related to our proposals is 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., [5, 19, 1, 9]), 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 blacklists and their management are not considered by any of these access control models.

Filtered Wall Conceptual Architecture

The aim of this paper is to develop a method that allows OSN users to easily filter undesired messages, according to content based criteria. In particular, we are interested in defining an automated language-independent system providing a flexible and customizable way to filter and then control incoming messages.

Before illustrating the architecture of the proposed system, we briefly introduce the basic model underlying OSNs. In general, the standard way to model a social network is as directed graph, where each node corresponds to a network user and edges denote relationships between two different users. In particular each edge is labeled by the type of the established relationship (e.g., friend of, colleague of, parent of) and, possibly, the corresponding trust level, which represents how much a given user considers trustworthy with respect to that specific kind of relationship the user

with whom he/she is establishing it. Therefore, there exists a direct relationship of a given type RT and trust value X between two users, if there is an edge connecting them having the labels RT and X. Moreover, two users are in an indirect relationship of a given type RT if there is a path of more than one edge connecting them, such that all the edges in the path have label RT [11].

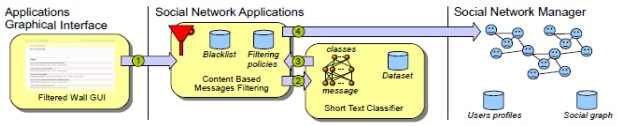
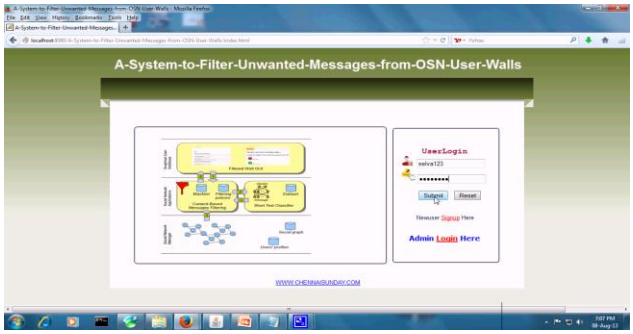


Fig. 1. Filtered Wall Conceptual Architecture.

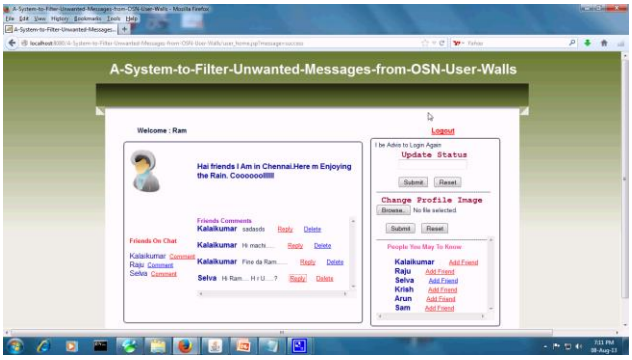
In general, the architecture in support of OSN services is a three-tier structure. The first layer commonly aims to provide the basic OSN functionalities (i.e., profile and relationship management). Additionally, some OSNs provide an additional layer allowing the support of external Social Network Applications (SNA)1. Finally, the supported SNA may require an additional layer for their needed graphical user interfaces (GUIs).

According to this reference layered architecture, the proposed system has to be placed in the second and third layers (Figure 1), as it can be considered as a SNA. In particular, users interact with the system by means of a GUI setting up their filtering rules, according to which messages have to be filtered out (see Sect. 5 for more details). Moreover, the GUI provides users with a FW that is a wall where only messages that are authorized according to their filtering rules are published.

V. RESULTS



First of all, in OSNs like in everyday life, the same message may have different meanings and relevance based on who writes it.



A further component of our system is a BL mechanism to avoid messages from undesired creators, independent from their contents. BLs is directly managed by the system.

VI. CONCLUSION

In this paper, we have presented a system to filter undesired messages from OSN walls. The system exploits a ML soft classifier to enforce customizable content-dependent FRs. Moreover, the flexibility of the system in terms of filtering options is enhanced through the management of BLs.

The present batch learning strategy, based on the preliminary collection of the entire set of labeled data from experts, allowed an accurate experimental evaluation but needs to be evolved to include new operational requirements. In future work, we plan to address this problem by investigating the use of on-line learning paradigms able to include label feedbacks from users. Additionally, we plan to enhance our system with a more sophisticated approach to decide when a user should be inserted into a BL. The development of a GUI and a set of related tools to make easier BL and FR specification is also a direction we plan to investigate, since usability is a key requirement for such kind of applications. In particular, we aim at investigating a tool able to automatically recommend trust values for those contacts user does not personally known.

To overcome this problem, a promising trend is to exploit data mining techniques to infer the best privacy preferences to suggest to OSN users, on the basis of the available social network data [14]. As future work, we intend to exploit similar techniques to infer BL rules and FRs. Additionally, we plan to study strategies and techniques limiting the inferences that a user can do on the enforced filtering rules with the aim of bypassing the filtering system, such as for instance randomly notifying a message that should instead be blocked, or detecting modifications to profile attributes that have been made for the only purpose of defeating the filtering system.

REFERENCES

[1] 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.

[2] 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.

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

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

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

[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, no. 12, pp. 51–60, 1992.

[9] P. S. Jacobs and L. F. Rau, "Scisor: Extracting information from online news," Communications of the ACM, vol. 33, no. 11, pp. 88–97, 1990.

[10] S. Pollock, "A rule-based message filtering system," ACM Transactions on Office Information Systems, vol. 6, no. 3, pp. 232–254, 1988.

[11] P. E. Baclace, "Competitive agents for information filtering," Communications of the ACM, vol. 35, no. 12, p. 50, 1992.

[12] 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. IEEE Computer Society Press, Los Alamitos, US, 1990, pp. 320–326.

[13] G. Amati and F. Crestani, "Probabilistic learning for selective dissemination of information," Information Processing and Management, vol. 35, no. 5, 1999.

[14] 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.

[15] 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.

AUTHOR DETAILS:

First Author: ***V. Basha*** received B.Tech from SSJ College of Engineering, Hyderabad, in the year 2012. He is currently M.Tech student in Computer Science and Engineering Department from Marri Laxman Reddy Institute of Technology. And his research interested areas are in the field of Cloud Computing, Mobile Computing, Networking and Information Security.

Second Author: ***D. Bheekya*** working as an Asst. Professor in Marri Laxman Reddy Institute of Technology. He has completed his M.Tech CSE and he has 6 years of teaching experience. His research interested areas are Data Mining, Network Security and Cloud Computing.