Privacy-Preserving Protocol for unwanted Message With Cooperative Firewall Optimization

Bhanu Chandar P[1], G.VaraPrasad[2]

[1]Dept. of CSE, Nova College of Engineering & Technology, Jangareddy Gudem, A.P, India

[2]Associate Professor, Nova College of Engineering & Technology, Jangareddy Gudem, A.P, India

Abstract: The aim of the present work is to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from Online Social Networks user walls. Today, the impact of Online Social Networks utility goes beyond personal communications. For instance, even intelligence communities may use the technology for effectively carrying out their missions.

The ability to control the messages posted on their own private space to avoid that unwanted content is displayed. 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. Up to now, Online Social Networks provide little support to this requirement. To fill the gap, in this paper, we propose a system allowing Online Social Networks 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 a Machine Learning-based soft classifier automatically labeling messages in support of content-based filtering. 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 Online Social Networks user walls

Keywords: Firewall Optimization Networks, information filtering, short text classification, policy-based personalization.

Introduction:

OSN modeling approach is presented to fill the gap. This model is

based on an innovative game theoretical approach and it is analyzed both from a theoretical and simulation-oriented view. Firewall Optimization Networks (FONs),(Online Social Networks) 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 statistics1 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.

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 commenting other 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 Online Social Networks 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.

Scope

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 to automatically assign with each short text message a set of categories based on its content.

Motivation:

The major efforts in building a robust short text classifier (STC) 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 from which we inherit the learning model and the elicitation procedure for generating preclassified 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 in performing the learning phase creates the premise for an adequate use in OSN domains, as well as facilitates the experimental evaluation tasks.

I. METHODS

Besides classification facilities, the system provides a powerful rule layer exploiting a flexible language 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. 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. OSN

modeling approach is presented to fill the gap. This model is based on an innovative game theoretical approach and it is analyzed both from a theoretical and simulation-oriented view. The game theoretic model is implemented in order to analyze several attack scenarios. As the results show, there are several

scenarios where OSN services are very vulnerable and hence more protection mechanisms should be provided in order to

secure the data contained across these networks.The experiments we have carried out show the effectiveness of the developed filtering techniques. In particular, the overall strategy was experimentally evaluated numerically assessing the performances of the ML short classification stage and subsequently proving the effectiveness of the system in applying FRs. Finally, we have provided a prototype implementation of our system having Facebook as target OSN, even if our system can be easily applied to other OSNs as well. 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.

(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.

II. EXISTING APPROACH

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 commenting other 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, Face book allows users to state who is allowed to insert messages in their walls. 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.

III. OUR PROPOSED APPROACH

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 to automatically assign with each short text message a set of categories based on its content.

We insert the neural model within a hierarchical two level classification strategy. In the first level, the RBFN categorizes short messages as Neutral and Non-neutral; in the second stage, Non-neutral messages are classified producing gradual estimates of appropriateness to each of the considered category. Besides classification facilities, the system provides a powerful rule layer exploiting a flexible language 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. 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.

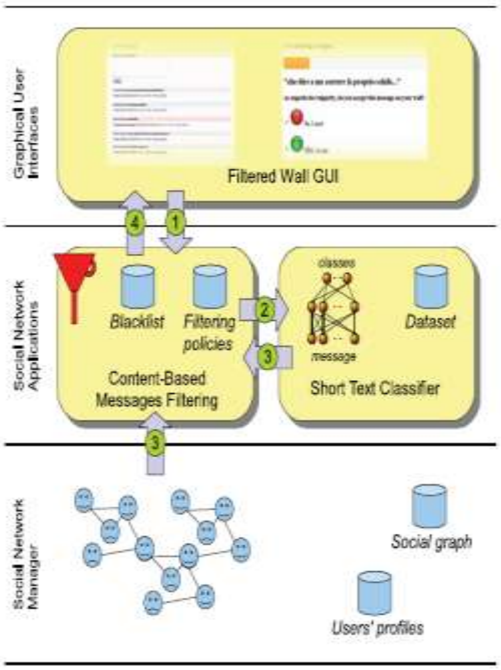

Fig. 1. Filtered wall conceptual architecture and the flow messages follow, from writing to publication.

Fig. 1 Filter Wall Architecture

The architecture in support of OSN services is a three-tier structure (Fig. 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 externalSocial 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 modules. The latter component aims to classify messages according to a set of categories. The strategy underlying this module is described in Section 4. 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. As graphically depicted in Fig. 1, the path followed by a message, from its writing to the possible final publication

can be summarized as follows:

1. After entering the private wall of one of his/hercontacts, the user tries to post a message, which isintercepted by FW.

2. A ML-based text classifier extracts metadata from the content of the message.

3. FWuses 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. In what follows, we explain in more detail some of the above-mentioned steps.

SHORT TEXT CLASSIFIER

Established techniques used for text classification work well on data sets with large documents such as newswires corpora, but suffer when the documents in the corpus are short. In this context, critical aspects are the definition of a set of characterizing and discriminant features allowing the representation of underlying concepts and the collection of a complete and consistent set of supervised examples.

Our study is aimed at designing and evaluating various representation techniques in combination with a neural learning strategy to semantically categorize short texts.From a ML point of view, we approach the task by defining a hierarchical two-level strategy assuming that it is better to identify and eliminate "neutral" sentences, then classify "nonneutral" sentences by the class of interest instead of doing everything in one step. This choice is motivated by related work showing advantages in classifying text and/or short texts using a hierarchical strategy. The first-level

task is conceived as a hard classification in which short texts are labeled with crisp Neutral and Nonneutral labels. The second-level soft classifier acts on the crisp set of nonneutral short texts and, for each of them, it "simply" produc esestimated appropriateness or "gradual membership" for each of the conceived classes, without taking any "hard"decision on any of them. Such a list of grades is then used

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.

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 SNMP; 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.

IV. EXPERIMENT RESULTS

Comparison Analysis

The lack of benchmarks for OSN short text classification makes problematic the development of a reliable comparative analysis. However, an indirect comparison of our method can be done with work that show similarities orcomplementary aspects with our solution. A study that responds to these characteristics is proposed in, where a classification of incoming tweets into five categories is described. Similarly to our approach, messages are very short and represented in the learning framework with both internal, content-based and contextual properties. By trial and error, we found a quite good parameter configuration for the RBFN learning model. The best value for the M parameter, that determines the number of Basis Function, is heuristically addressed to N=2, where N is thenumber of input patterns from the data set. The value used for the spread _, which usually depends on the data, is 32 for both networks M1 and M2 To calculate Correct words and Bad words Dp features, we used two specific Italian word-lists, one of these is the CoLFIS corpus. The cardinalities of TrSD and TeSD, subsets of D with TrSD \ TeSD ¼ ;, were chosen so that TrSD is twice larger than TeSD. Network M1 has been evaluated using the OA and the K value. Precision, Recall, and F-Measure were used for the M2 network because, in this particular case, each pattern can be assigned to one or more classes.

V. 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.

This work is the first step of a wider project. The early encouraging results we have obtained on the classification procedure prompt us to continue with other work that will aim to improve the quality of classification. In particular, future plans contemplate a deeper investigation on two interdependent tasks. The first concerns the extraction and/ or selection of contextual features that have been shown to have a high discriminative power. The second task involves the learning phase. Since the underlying domain is dynamically changing, the collection of preclassified data may not be representative in the longer term. 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 online 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.

Future Enhancements

We plan to study strategies and techniques limiting the inferences that a user can do on the enforced filtering rules with the aim of by passing 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.

VI. REFERENCES

[1] A. Adomavicius and G. Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions,” IEEE Trans. Knowledge and Data Eng., vol. 17, no. 6, pp. 734-749, June 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,” Proc. Fifth ACM Conf. Digital Libraries, pp. 195-204, 2000.

[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,” Proc. ECML/PKDD Workshop Privacy and Security Issues in Data Mining and Machine Learning (PSDML ’10), 2010.

[6] N.J. Belkin and W.B. Croft, "Information Filtering and Information Retrieval: Two Sides of the Same Coin?" Comm. ACM, vol. 35, no. 12, pp. 29-38, 1992.

[7] P.J. Denning, "Electronic Junk," Comm. 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," Comm. ACM, vol. 35, no. 12, pp. 51-60, 1992.

[9] P.S. Jacobs and L.F. Rau, "Scisor: Extracting Information from On- Line News," Comm. ACM, vol. 33, no. 11, pp. 88-97, 1990.

[10] S. Pollock, "A Rule-Based Message Filtering System," ACM Trans. Office Information Systems, vol. 6, no. 3, pp. 232-254, 1988.

M.Tech Student Mrs P.Bhanu Chandar CSE Branch Nova College of Engineering & Technology Vegavaram, Jangareddygudem.