

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IJCSMC, Vol. 3, Issue. 8, August 2014, pg.521 – 525

RESEARCH ARTICLE

FILTERING OF UNDESIREDMESSAGES FROM OSN USER SPACE

Poli Reddy¹, B. Bala Krishna²

¹M.Tech 2nd Year, Dept. of CSE, TITS affiliated to JNTU, HYDERABAD, TS, INDIA

²Assistant Professor, Dept. of CSE, TITS affiliated to JNTU, HYDERABAD, TS, INDIA

¹ rpreddy.839@gmail.com; ² balakrishna.bangaru@gmail.com

Abstract - One fundamental issue in today On-line Social Networks (OSNs) 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, in this paper, we propose a system allowing 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 users to customize the filtering criteria to be applied to their walls, and Machine Learning based soft classifier automatically labelling messages in support of content-based filtering.

Keywords— Online social networks, information filtering, short text classification, policy-based personalization

I. Introduction

Online Social Network are today one of the most popular interactive medium to communicate, share and announce a considerable amount of human life information. We share different kinds of content, if including free text, video, audio image data. According to Facebook statistics average user create 90 pieces of content (web links, blog post, notes, news stories, photo album, etc.), whereas more than 30 billion pieces of content are shared each month. The huge and dynamic character creates the room for web content mining techniques aimed to reflect only useful information. They can provide active support to OSN management, such as for access control and information filtering. Information filtering has been greatly explored for what concerns textual documents and mostly web content. However the main aim of all the proposals is to provide user a classification mechanism to avoid useless data.

This is due to the fact that in OSNs there is possibility or commenting on other posts particularly public/private areas, called in general walls. Information gathering can there be used to give users the ability to control on the messages written on their wall, by filtering unwanted messages. And we all will agree that this is the key OSN

service. For example, Facebook allows user to state who is allowed to insert message in their walls(i.e., friends, friends of friends, or defined groups of friends).however no content based preference are supported and therefore it is not possible to prevent undesired messages such as political or vulgar ones, no matter of the user who post them.

Providing this service is not only the matter of using previously defined web content mining techniques for a different application, rather it requires to design additional classification strategies and also wall post comprises of short text and traditional methods have limitations. The aim of the present work is to present or propose a automated system called FW (Filtered Wall), able to filter unwanted messages from OSN wall. We exploit Machine Learning (ML) text categorization techniques to automatically assign with each short text a set of categories based on its content.

We implement a 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.

II. Existing System

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.

III. Proposed System

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.

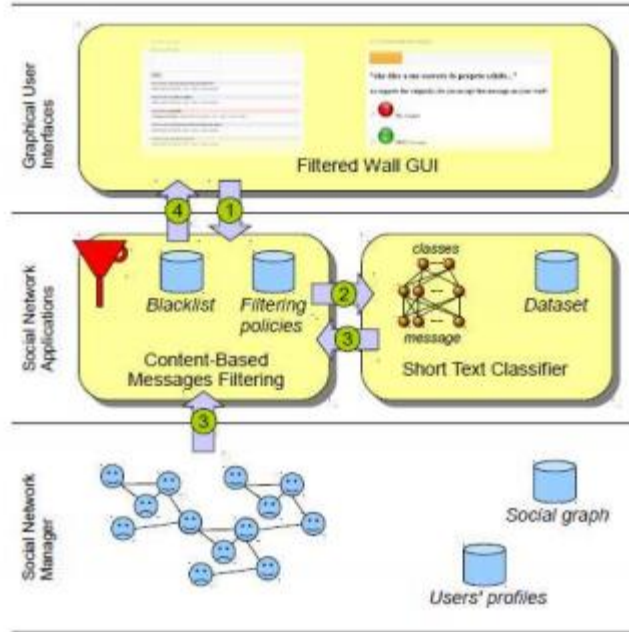


FIGURE 1: FILTERED WALL ARCHITECTURE

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

IV. Implementation

Implementation is the stage of the project when the theoretical design is turned out into a working system. Thus it can be considered to be the most critical stage in achieving a successful new system and in giving the user, confidence that the new system will work and be effective.

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.

V. Related Work

In this section, we introduce the rules adopted for filtering unwanted messages. We model a social network as a directed graph, where each node represents a network user and edges represent relationships between two different users. Each edge is labelled 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 the relationship. We assume that trust levels are rational numbers in the range $[0; 1]$. 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.

A. Filtering Rules

We consider three main issues in defining the language for FRs specification. First is related to the fact that, the same message may have different meanings and relevance based on who writes it. Message 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.

B. Online Setup Assistant for FRs Thresholds

By conceiving and implementing within FW, an Online Setup Assistant (OSA) procedure, we address the problem of setting thresholds to filter rules. OSA presents the user with a set of messages selected from the dataset. For each message, the user expresses 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 permits to calculate customized thresholds representing the user attitude in accepting or rejecting certain contents.

C. Blacklists

BLs are directly managed by the system, and should be able to determine the users to be inserted in the BL and decide user's retention in the BL is finished. Such information are given to the system through a set of rules, called BL rules. We let 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.

CONCLUSION

In this paper, we have presented a system to filter unwanted 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 paper is related to our work which will be the first step of a wider project. On later stage, our work will be around two concerns, first, the extraction and/or selection of contextual features that have been shown to have a high discriminative power and , second, deals with the learning process and its techniques.

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. We believe that such a tool should suggest trust values based on user action, behaviour and reputation in OSN, which might imply to enhance OSN with audit mechanisms. Moreover, we are aware that a usable GUI could not be enough, representing only the first step. And nevertheless many studies have shown that average OSN user have difficulties in understanding the simple privacy settings provided by today OSNs. We decided to deal with this by exploiting data mining techniques to infer the best privacy preferences to suggest to OSN users, on the available social network data. In future, we intend to use similar techniques to infer BLs and FRs . Also 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.

REFERENCES

- [1] 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.
- [2] 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.
- [3] P. E. Baclace, "Competitive agents for information filtering," *Communications of the ACM*, vol. 35, no. 12,p. 50, 1992.
- [4] Boykin, P.O., Roychowdhury, V.P.: Leveraging social networks to fight spam. *IEEE Computer Magazine* 38, 61–67 (2005).

- [5] J. Golbeck, "Combining provenance with trust in social networks for semantic web content filtering," in *Provenance and Annotation Data*, ser. *Lecture Notes in Computer Science*, L. Moreau and I. Foster, Eds. 2006
- [6] Carminati, B., Ferrari, E.: Access control and privacy in web-based social networks. *International Journal of Web Information Systems* 4, 395–415 (2008)
- [7] B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, "Short text classification in twitter to improve information filtering," 2010P. J. Denning, "Electronic junk," *Communications of the ACM*, vol. 25, no. 3, pp. 163–165, 1982.
- [8] D. D. Lewis, Y. Yang, T. G. Rose, and F. Li, "Rcv1: A new benchmark collection for text categorization research," *Journal of Machine Learning Research*, 2004.
- [9] 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.
- [10] S. Pollock, "A rule-based message filtering system," *ACM Transactions on Office Information Systems*, vol. 6, no. 3, pp. 232–254, 1988.