ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Enhancing Secured Information in Online Social Network by Text Classifier

A.Maheshwari[1], Prof.B.Sakthivel[2]

M.E, Department of CSE, P.S.V College of Engineering and Technology, Krishnagiri, Tamilnadu, India[1]

Head of the Department, Department of CSE, P.S.V College of Engineering and Technology, Krishnagiri, Tamilnadu, India[2]

ABSTRACT *:* One fundamental issue in today's Online 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 a Machine Learning-based soft classifier automatically labeling messages in support of content-based filtering.

KEYWORDS: Online social networks, information filtering, short text classification, policy-based personalization

1 INTRODUCTION

ONLINE 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 Face book 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 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 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. service that has not been provided so far. Indeed, today OSNs provide very little support to prevent unwanted messages on user walls.

II. RELATED WORK

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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

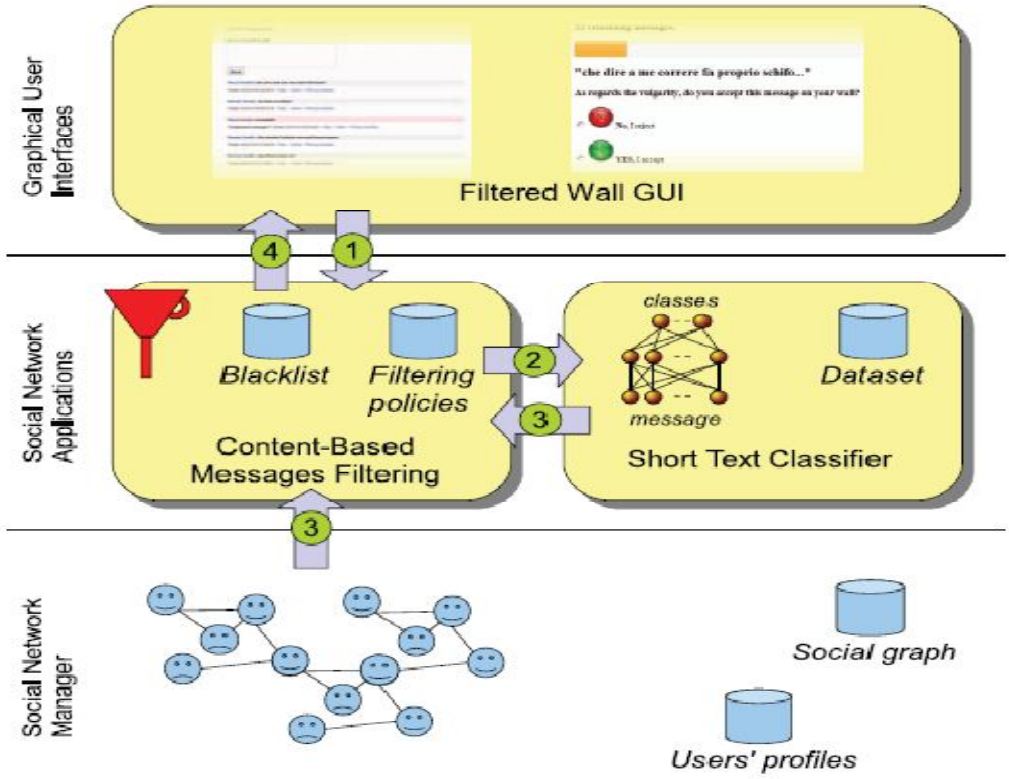
Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

III.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. 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 .

IV.FILTERED 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 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. Messages Filtering (CBMF) and the Short Text Classifier modules. The latter component aims to classify messages

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

according to a set of categories. 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/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 user's 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.

V.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 "non neutral" sentences by the class of interest instead of doing everything in one step.

VI.TEXT REPRESENTATION

The extraction of an appropriate set of features by which representing the text of a given document is a crucial task strongly affecting the performance of the overall classification strategy. These features play a key role in deterministically understanding the semantics of the messages. All proposed features have been analyzed in the experimental evaluation phase in order to determine the combination that is most appropriate for short message classification.

VII. FILTERING RULES AND BLACKLIST MANAGEMENT

In this section, we introduce the rule layer adopted for filtering unwanted messages. We start by describing FRs, then we illustrate the use of BL's. In what follows, we model a social network as a 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)

Filtering Rules

1. A set of attribute constraints of the form an OP av, where an is a user profile attribute name, av and OP are, respectively, a profile attribute value and a comparison operator, compatible with an's domain.
2. A set of relationship constraints of the form ðm; rt; mi Depth; maxTrustÞ, denoting all the OSN users participating with user m in a relationship of type rt, having a depth greater than or equal to minDepth, and a trust value less than or equal to maxTrust.

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 procedure.. 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 data set 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.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

VIII. 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 FR's. 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 pre classified 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.

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 do believe that such a tool should suggest trust value based on users actions, behaviors, and reputation in OSN, which might imply to enhance OSN with audit mechanisms. However, the design of these audit-based tools is complicated by several issues, like the implications an audit system might have on users privacy and/or the limitations on what it is possible to audit in current OSNs.

IX.FUTURE WORK

As the future work and our contribution we enhance the system by creating a instance randomly notifying a message system that should instead be blocked, or detecting modifications to profile attributes that have been made for the only purpose of defeating the filtering system. Automatically user will get a mail notification. . The system exploits a ML soft classifier to enforce customizable content-dependent FR's. 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 I have obtained on the classification procedure prompt us to continue with other work that will aim to improve the quality of classification.

Additionally, I 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.

X.ACKNOWLEDGEMENT

The success and final outcome of this project required a lot of guidance and assistance from many people and I am extremely fortunate to have got this all along the completion of my project work. Whatever I have done is only due to guidance and assistance and I would not forget to thank them. I respect and thank my guide Mr. B. SAKTHIVEL M.E., Head of the Department of Computer Science and Engineering for providing all support and guidance which made me to complete the project on time. . The blessing, help and guidance given by him time to time shall carry me a long way in the journey of life on which I am about to embark .Lastly, I thank almighty, my parents, husband and children for their constant encouragement without which this assignment would not be possible.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

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.
[11] P.E. Baclace, "Competitive Agents for Information Filtering," Comm. 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," Proc. Sixth IEEE Conf. Artificial Intelligence Applications (CAIA '90), pp. 320-326, 1990.
[13] G. Amati and F. Crestani, "Probabilistic Learning for Selective Dissemination of Information," Information Processing and Management, vol. 35, no. 5, pp. 633-654, 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," Proc. 24th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval, pp. 294-302, 2001.