

International Journal of Research (IJR) e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015

Available at http://internationaljournalofresearch.org

A Machine Learning Text Categorization Techniques System to Filter Unwanted Words in Social Networks

V. Srinivas

Associate Professor, Department of CSE AURORA'S SCIENTIFIC, TECHNOLOGICAL & RESEARCH ACADEMY, Bandlaguda, Hyderabad.

Vemula Mounika

M.Tech, Computer Science & Engineering AURORA'S SCIENTIFIC, TECHNOLOGICAL & RESEARCH ACADEMY, Bandlaguda, Hyderabad.

Abstract:

Users build explicit networks to represent their social relationships. Users can upload and share information related to their personal lives. The Online Social Networks (OSN) mainly helps an individual to connect with their friends, family and the society online in order to gather and share new experiences with others. Now-a-days, the OSNs are facing the problem of the people posting the indecent messages on any individual's wall which annoys other people on seeing them. In order to filter those unbearable messages a system called Machine Learning is introduced. 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 messages a set of categories based on its content. The major efforts in building a robust short text classifier (STC) are concentrated in the extraction and selection of a set characterizing and discriminating features.

Index Terms: On-line Social Network; Information filtering; short text classification and policy-based personalization

1. INTRODUCTION

Information and communication technology plays a significant role in today's networked society. It has affected the online interaction between users, who are aware of security applications and their implications on personal privacy. There is a need to develop more security mechanisms for different communication technologies, particularly online social networks. OSNs provide very little support to prevent unwanted messages on user walls. With the lack of classification or filtering tools, the user receives all messages posted by the users he follows. In most cases, the user receives a noisy stream of updates. In this paper, an information Filtering system is introduced.

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. In the proposed system 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. 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 categorization (ML) text 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 discriminant features.



International Journal of Research (IJR)

e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015 Available at http://internationaljournalofresearch.org

Web content Mining is used to discover useful and relevant information from a large amount of Data. In OSN's, information filtering can be used for a different purpose. This is due to fact that in OSN's there is the possibility of posting (or) commenting other posts on particular public (or) private areas called Walls. Information filtering is mainly used to give user the ability to control the message written on their own walls by filtering out unwanted messages. Today OSN's provide 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, defined group of friends. Filtered wall is used to filter unwanted messages from OSN user walls. We used Machine Learning text technique categorization to automatically categorize each short text messages based on its content. We base the overall short 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. We use the neural model RBFN categorizes as Neural and Nonneural FR filtering rules by which it can state what contents should not be displayed on their walls. In addition, the system provides the user defined Blacklists that is mainly used to temporarily prevent to post any kind of message on a user wall.



Fig 1: OSN General Architecture

Social Network Manager: To provide the basic OSN functionalities (i.e) Profile and relationship management.

Social Network Applications: To provide external social network applications.

Graphical User Interfaces: User interacts with the system.

Content Based Filtering: It is mainly used to select

information item based on the correlation between the content of the items and the user preferences.

Filtering: It is mainly used to filter the unwanted messages using Blacklists.

2. RELATED WORK

A distinction is made between two types of text filtering systems: content-based and social filtering systems. In content-based systems, filtering is done by exploiting the information extracted from the text of documents. In social filtering systems, documents are filtered based on annotations made by prior readers of the documents. Topic-specific search engines provide an alternative way to support efficient information retrieval on the Web by providing more precise and customized searching in various domains. However, developers of topic-specific search engines need to address two issues: how to locate relevant documents (URLs) on the Web and how to filter out irrelevant documents from a set of documents collected from the Web. This paper reports our research in addressing the second issue. We propose a machine-learning-based approach that combines Web content analysis and Web structure analysis. We represent each Web page by a set of content-based and link-based features, which can be used as the input for various machine learning algorithms. The proposed approach was implemented using both a feed forward/back propagationneural network and a support vector machine. Two experiments were designed and conducted to compare the proposed Web-feature approach with two existing Web page filtering methods - a keyword-based approach and a lexicon-based approach. The experimental results showed that the proposed approach in general performed better than the benchmark approaches, especially when the number of training documents was small. The proposed approaches can be applied in topic-



International Journal of Research (IJR)

e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015 Available at http://internationaljournalofresearch.org

specific search engine development and other Web applications such as Web content management. The advantages of this approach over the knowledge engineering approach (consisting in the manual definition of a classifier by domain experts) are a very good effectiveness, considerable savings in terms of expert labor power, and straightforward portability to different domains.

This survey discusses the main approaches to text categorization that fall within the machine learning paradigm. We will discuss in detail issues pertaining to three different problems, namely document representation, classifier construction, and classifier evaluation. The system allows 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 a user to customize the filtering criteria to be applied to their walls, and a Machine Learning based soft classifier automatically labeling content-based messages in support of filtering.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 who posts them. Even though the Social Networks today, have the restrictions on the users who can post and comment on any user's wall, they do not have any restrictions on what they post. So, some people will use the indecent and vulgar words in commenting on the public posts. 2. 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.

3. IMPLEMENTATION

Machine learning (ML) is used as text categorization techniques to automatically assign

each short text message with in a set of categories based on its content. The major efforts in building a robust Short Text Classifier (STC) concentrate in the extraction and selection of a set characterizing and discriminating features. Here, a database of the categorized words is built and it is used to check the words if it has any indecent words. If the message consists of any vulgar words, then they will be sent to the Blacklists to filter out those words from the message. Finally, the message without the indecent words will be posted in the user's wall on the result of the content-based-filtering technique.The users will create and manage their own "groups" (like the new Face book groups' pages). Each group has a homepage that provides a place for subscribers to post and share (by posting messages, images, etc.) and a block that provides basic information about the group. Users can also enable additional features in their owned page like view friends list and add friends by using friend's requests as well as share their images with selected group's members. The status of their friends' requests should also be updated here Fig.2. shows the filtering process.



Fig.2.Filtering process.

A. Filtering process

In defining the language for FRs specification, we consider three main issues that, in our opinion, affect a message filtering decision. First, in OSNs like in everyday life, the same message may have different meanings and relevance based on who writes it. As a



International Journal of Research (IJR) e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015

Available at http://internationaljournalofresearch.org

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 to creators with creators or 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.

The problem of setting thresholds to filter rules is also addressed, by conceiving and implementing within FW, an Online Setup Assistant (OSA) 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 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.

B. Blacklisting Process

A further component of our system is a Blacklist (BL) mechanism to avoid messages from undesired creators, independent from their contents. BL 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 is given to the system through a set of rules, hereafter called BL rules. Such rules are not defined by the Social Network Management, 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, and at the same time, he will not be able to post in the wall.

Short text categorization is a hierarchical two-level classification process. The first-level classifier does a binary hard classification that labels messages as Neutral and Non-Neutral. The first-level filtering task enables the subsequent second-level task in which a finer-grained classification is performed. The second-level classifier carries out a soft-partition of Nonneutral messages assigning a given message a gradual membership to each of the non-neutral classes. 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 Filtering Rules. The two measures can be



International Journal of Research (IJR)

e-ISSN: 2348-6848, p- ISSN: 2348-795X Volume 2, Issue 08, August 2015 Available at http://internationaljournalofresearch.org

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.

4. CONCLUSION

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. In this paper, a system to prevent the indecent messages from the Social Networking site walls has been presented. The Usage of Machine Learning has given higher results to the system to trace the messages and the users to distinguish between the good and bad messages and the authorized and unauthorized users in the Social Networking User Profiles automatically. Thus the Machine Learning Technique plays a vital role in this paper in order to generate the blacklist of the bad words and the unauthorized users. The user has to update his privacy setting in his account in order to add this method to prevent the obscenity in his public profile. In this context, a statistical analysis has been conducted to provide the usage of the good and bad words by the persons in the sites. Overall, the obscenity of the users has been prevented.

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] 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] 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] C. Cleverdon, "Optimizing Convenient Online Access to BibliographicDatabases," Information Services and Use, vol. 4, no. 1,pp. 37-47, 1984.

[8] P.J. Denning, "Electronic Junk," Comm. ACM, vol. 25, no. 3,pp. 163-165, 1982.

[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.



V. SRINIVAS Completed Master of Technology in Computer Science and Engineering from JNTU Hyderabad. Currently working as an Associate Professor at AURORA'S SCIENTIFIC , TECHNOLOGICAL & RESEARCH ACADEMY, Bandlaguda, Hyderabad.



VEMULA MOUNIKA pursing M.Tech in Computer Science Engineering from AURORA'S SCIENTIFIC, TECHNOLOGICAL & RESEARCH ACADEMY, Bandlaguda, Hyderabad.