FILTER UNWANTED MESSAGES ON OSN USER WALL

PANKAJA.KAMBLE, RAHUL. S. DUDHANE, AMOL O. GHARPANDE

1Student, CE Department, Nagpur University
2Student, CE Department, Nagpur University, Nagpur University
3Asst.Prof., CE Department,

SD College of Engineering, Wardha, Maharashtra, India
Email Id: pankajkmbl8@gmail.com Email id: rahuldudhane91@gmail.com Email id: amol_gharpane@hotmail.com

Abstract-Social Network Online Social Network is to give users the ability to control the messages posted on their own private space to avoid that unwanted content is displayed. To fill the gap, 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 classifier automatically labelling messages in support of content-based filtering.

Keywords-Online Social Network, Content-based filtering, Machine learning, unwanted content.

I. INTRODUCTION

Daily and continuous communications imply the exchange of several types of content, including free text, image, audio and video data. According to Facebook statistics, average user creates 90 pieces of content each month, whereas more than 30 billion pieces of content 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.

There are instruments which 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.

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. 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, 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.

II. RELATED WORK

This paper presents an overview of the field of recommender systems and describes the current generation of recommendation methods that are usually classified into the following three main categories: content-based, collaborative, and hybrid recommendation approaches. This paper also describes various limitations of current recommendation methods and discusses possible extensions that can improve recommendation capabilities and make recommender systems applicable to an even broader range of applications. These extensions include, among others, an improvement of understanding of users and items, incorporation of the contextual information into the recommendation process, support for multi criteria ratings, and a provision of more flexible and less intrusive types of recommendations.

As the Web continues to grow, it has become increasingly difficult to search for relevant information using traditional search engines. 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. 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 basic problem that we are going to see in using these sites is “Lack of Privacy”. Till today, Social Networks Sites provide little support to this requirement. To sort out this problem, in this project we are proposing a system which will provide the indirect control to the users of these sites. This proposed model can be achieved through a modern rule based system, that allows administrators to customize the filtering criteria to be applied to their walls, and a Machine Learning based soft classifier automatically labelling messages in support of content based filtering.

Protecting personal information privacy has become a controversial issue among online social network providers and users. Most social network providers have developed several techniques to decrease threats and risks to the users’ privacy. These risks include the misuse of personal information which may lead to illegal acts such as identity theft. This study aims to measure the awareness of users on protecting their personal information privacy, as well as the suitability of the privacy systems which they use to modify privacy settings. Survey results show high percentage of the use of smart phones for web services but the current privacy settings for online social networks need to be improved to support different type of mobile phones screens. Because most users use their mobile phones for Internet services, privacy settings that are compatible with mobile phones need to be developed. The method of selecting privacy settings should also be simplified to provide users with a clear picture of the data that will be shared with others. Results of this study can be used to develop a new privacy system which will help users control their personal information easily from different devices, including mobile Internet devices and computers.

Nowadays, online social networking websites are commonly used. These websites are called as social sites. Social networking websites works similar to an online community of internet users. Depending on the different social websites many of these online community members share common interests in hobbies, religion, politics and education as well as alternative lifestyles. Once you are become an authorized user to access to a social networking website you will be able to socialize. Socialization includes retrieving, accessing or reading the profile pages of other members those are involved in social sites and possibly even contacting them. But the sensitive issue in online social networks is that users are not able to control the messages posted on their walls so to fill the gap in this paper we are proposing some approaches which will allow OSN users to have a direct control on the messages posted on their walls. This can be done with the help of rule based system and Machine Learning classification based on content filtering.

III. OBJECTIVE OF THE WORK

- We 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.

IV. WORK PLAN

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 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. 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 from whom 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 to performing the learning phase creates the premise for an adequate use in OSN domains, as well as facilitates the experimental evaluation tasks.

- Complete Word Discovery Algorithm Pseudo Code

A complete word is a complete substring of the collated text of the input text message, defined in the following way: Let T be a sequence of elements \((t_1, t_2, t_3 \ldots t_n)\). S is a complete substring of T when S occurs in k distinct positions p_1, p_2, p_3 \ldots p_k in T.

In other words, a complete word cannot be extended by adding preceding or trailing elements, because at least one of these elements is different from the rest. Here are the whole word extraction phases as pseudo-code.

Step 1 - Discover right-complete word.
Step 2 - Discover left-complete word.
Step 3 - Sort the left-complete word automatically.
Step 4 - Combine the left- and right-complete words into a set of complete words.
The architecture in support of OSN services is a three-tier structure (Figure No. A). 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). The supported SNAs may in turn require an additional layer for their needed Graphical User Interfaces (GUIs). According to this 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 i.e. 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 (STC) modules. The latter component aims to classify messages according to a set of categories. 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 process.

- **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 user walls 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.

- **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 is 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.

- **Text Categorization**
  The assignment of natural language texts to one or more predefined categories based on their content – is an important component in many information organization and management tasks. We compare the effectiveness of five different automatic learning algorithms for text categorization in terms of learning speed, real time classification speed, and classification accuracy. We also examine training set size, and alternative document representations. Very accurate text classifiers can be learned automatically from training examples. Linear Support Vector Machines (SVMs) are particularly promising because they are very accurate, quick to train, and quick to evaluate.

V. **ADVANTAGES**

- A system to automatically filter unwanted messages from OSN user walls on the basis of both message content and the message creator relationship and characteristics.
- The substantially extends for what concerns both the rule layer and the classification modules.

CONCLUSION

Existing system is used to filter undesired messages from OSNs wall using customizable filtering rules (FR) enhancing through Black lists (BLs). In present system we are more focus on an investigation of two interdependent tasks in depth. This system approach decides when user should be inserted into a black list.

REFERENCES


Filter Unwanted Messages On OSN User Wall

89


[5] “Content-Based Filtering in On-line social Network” M. Vanetti, E. Binaghi, B. Carminati, M. Carullo and E. Ferrari, Department of Computer Science and Communication University of Insubria 21100 Varese, Italy fmarco.vanetti, elisabetta.binaghi, barbar.carullo, elena.ferraring @uninsubria.it


★★★