International Journal of Research Studies in Science, Engineering and Technology
Volume 1, Issue 7, October 2014, PP 38-43
ISSN 2349-4751 (Print) & ISSN 2349-476X (Online)

A System to Block Unwanted Messages from OSN User Walls using Filtering Rules

Nair Pratap Premchandran
P.G Scholar, Dept of CSE, VVIT
Chevella, Hyderabad, India

M.Sravanthi
Assistant Professor, Dept of CSE, VVIT,
Chevella, Hyderabad, India

Abstract: *Now days On-line Social Networks (OSNs) are one of the most popular interactive medium to communicate, share, and disseminate a considerable amount of human life information. This project represents a system enforcing filtering of unwanted messages coming from the user based on its content. Our system gives ability to OSN users to have a direct control on the messages posted on their walls. Up to now, OSNs provide little support to prevent unwanted messages on user walls. There is no content-based preferences are supported and therefore it is not possible to prevent unwanted 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. One fundamental issue in this system is blocking of user for lifetime. We overcome this Problem by using Proposed System. In this paper, we propose a system that performs blocking of user for particular time limit and also send notification, E-Mail to that who has posted unwanted message on wall. Along with that we are using Self Organizing Neural Network (SONN) with Redial Based Function (RBF) for classification of text. In this we use the back propagation technique of neural network.*

Keywords: *On-line Social Networks, Content Filtering, Filtering rules, Blacklists, Machine learning text categorization.*

1. INTRODUCTION

On-line 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 Facebook statistics, 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 and 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. 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. 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. It exploits 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 discriminant features. The solutions investigated in this project 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 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, the use of neural learning, which is today recognized as one of the most efficient solutions in text classification. In particular, it emphasis 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. We insert the neural model within a hierarchical two level classification strategy.

2. RELATED WORK

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 [6]. 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 as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences [7], [8]. While electronic mail was the original domain of early work on information filtering, subsequent papers have addressed diversified domains including newswire articles, Internet "news" articles, and broader network resources [9], [10], [11].

The main contribution of this paper is the design of a system providing customizable content-based message filtering for OSNs, based on ML techniques. 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. Therefore, in what follows, we survey the literature in both these fields. A. Documents processed in content-based filtering are mostly textual in nature and this makes content-based filtering close to text classification. The activity of filtering can be modeled, in fact, as a case of single label, binary classification, partitioning incoming documents into relevant and non relevant categories. More complex filtering systems include multi-label text categorization automatically labeling messages into partial thematic categories. Content-based filtering is mainly based on the use of the ML paradigm according to which a classifier is automatically induced by learning from a set of pre-classified examples. A remarkable variety of related work has recently appeared, which differ for the adopted feature extraction methods, model learning, and collection of samples, [1]. The feature extraction procedure maps text into a compact representation of its content and is uniformly applied to training and generalization phases. Several experiments prove that Bag of Words (BoW) approaches yield good performance and prevail in general over more sophisticated text representation that may have superior semantics but lower statistical quality. As far as the learning model is concerned, there is a number of major approaches in content-based filtering and text classification in general showing mutual advantages and disadvantages in function of application dependent issues.

3. FILTERED WALL ARCHITECTURE

The architecture in support of OSN services is a three-tier structure (Figure 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. 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.

1) After entering the private wall of one of his/her contacts, the user tries to post a message.

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.

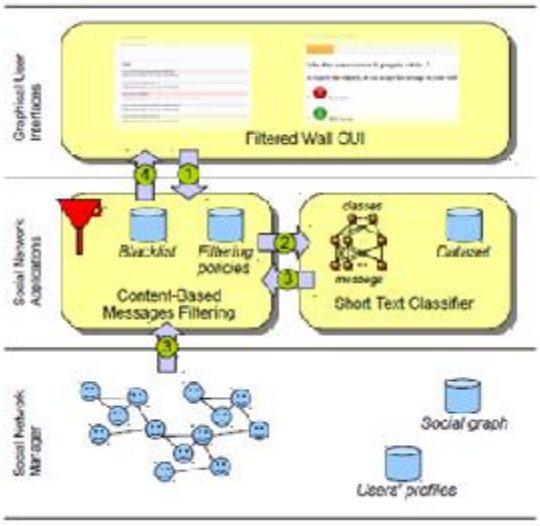

Fig 1. *Filtered Wall Conceptual*

The Filtered wall architecture in support of OSN services is a three-tier structure (see Fig. 1). The first layer, called Social Network Manager (SNM), it provides 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)

3.1 Short Text Classifier

Established techniques used for text classification work well on datasets 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 sematically 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. This choice is motivated by related work showing advantages in classifying text and/or short texts using a hierarchical strategy [1]. The first level task is conceived as a hard classification in which short texts are labeled with crisp Neutral and Non-Neutral labels. The second level soft classifier acts on the crisp set of non-neutral short texts and, for each of them, it "simply" produces estimated 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 by the subsequent phases of the filtering process.

3.2 Filtering Rules

FRs allows 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. 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 [11]. All these options are formalized by the notion of creator specification, defined as follows:

Definition 1 (Creator specification):

A creator specification creator Spec denotes a set of OSN users. It can have one of the following forms, possibly combined:

A set of attribute constraints

A set of relationship constraints of the form

Definition 2 (Filtering rule):

A filtering rule FR is a tuple (author, creatorSpec, contentSpec, action), where author is the user who specifies the rule;

creatorSpec is a creator specification, specified according to Definition 1;

contentSpec is a Boolean expression defined on content constraints of the form (C, ml), where C is a class of the first or second level and ml is the minimum membership level threshold required for class C to make the constraint satisfied;

action:{block; notify} denotes the action to be performed by the system on the messages matching contentSpec and created by users identified by creatorSpec.

Definition 3 (BL rule):

A BL rule is a tuple (author, creator Spec, creator Behavior, T), where

author is the OSN user who specifies the rule, i.e., the wall owner;

creatorSpec is a creator specification, specified according to Definition 1;

creator Behavior consists of two components RF Blocked and min Banned.

4. MACHINE LEARNING BASED CLASSIFICATION

In this section, we illustrate the performance evaluation study. We have carried out the classification and filtering modules. We start by describing the dataset. A. Problem and Dataset Description. The analysis of related work has highlighted the lack of a publicly available benchmark for comparing different approaches to content based classification of OSN short texts. To cope with this lack, we have built and made available a dataset D of messages taken from Facebook. The dataset, called WmSnSec, is available online at 1266 messages from publicly accessible Italian groups have been selected and extracted by means of an automated procedure that removes undesired spam messages and, for each message, stores the message body and the name of the group from which it originates. The messages come from the group's web page section, where any registered user can post a new message or reply to messages already posted by other users.

4.1 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. All these options are formalized by the notion of creator specification, defined as follows.

4.2 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 (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 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 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.

4.3 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 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.

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 FRs. The two measures can be 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.

5. 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 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.

REFERENCES

[1] Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo,"A System to Filter Unwanted Messages from OSN User Walls Using Filtering Rules", IEEE Transactions

On Knowledge And Data Engineering Vol:25 Year 2013

[2] A. Adomavicius, Gand Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions", IEEE Transaction on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734–749, 2005.

[3] 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.

[4] R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," in Proceedings of the Fifth ACM Conference on Digital Libraries. New York: ACM Press, 2000, pp. 195–204.

[5] F. Sebastiani, "Machine learning in automated text categorization" ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.

[6] 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.

[7] 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.

[8] P. J. Denning, "Electronic junk," Communications of the ACM, vol. 25, no. 3, pp. 163–165, 1982.

[9] 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.

[10] P. S. Jacobs and L. F. Rau, "Scisor: Extracting information from online news," Communications of the ACM, vol. 33, no. 11, pp. 88–97, 1990.

[11] S. Pollock, "A rule-based message filtering system," ACM Transactions on Office Information Systems, vol. 6, no. 3, pp. 232–254, 1988.

[12] P. E. Baclace, "Competitive agents for information filtering," Communications of the ACM, vol. 35, no. 12, p. 50, 1992.s