Privacy and Blocking of Unwanted Comments OSN User Walls

Vidhya.A[1], Keerthana Sukumar[2], Kalaivani Rajendran[3], Divyameenakshi Velusamy[4]

[1]Assistant Professor Department of Computer Science, Valliammai Engineering College, Chennai, Tamilnadu, India

[2, 3, 4]Department of Computer Science, Valliammai Engineering College, Chennai, Tamilnadu, India

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.

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. 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. 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.

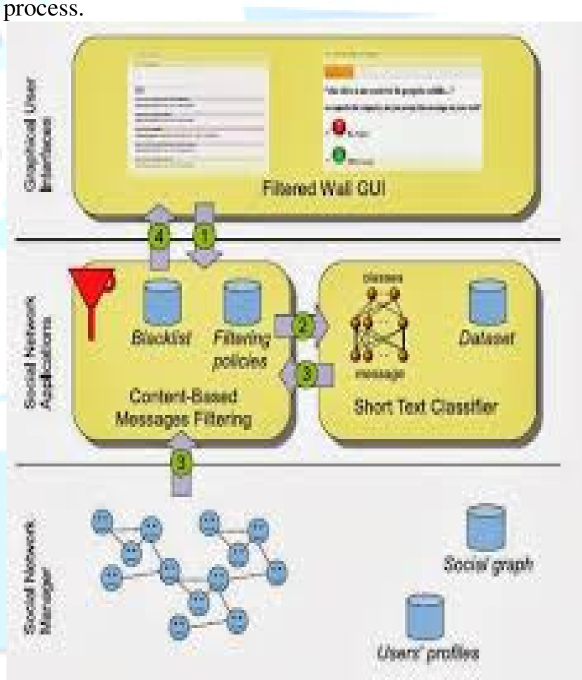
1.1 CONTENTBASED FILTERING

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 as opposed to a collaborative filtering system that chooses items based on the correlation between people with similar preferences 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 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 multilevel 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 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 are 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. A detailed comparison analysis has been conducted confirming superiority of Boosting-based classifiers, Neural Networks and Support Vector Machines over other popular methods, such as Rocchio and Naive Bayesian However, it is worth to note that most of the work related to text filtering by ML has been applied for long-form text and the assessed performance of the text classification methods strictly depends on the nature of textual documents. The application of content-based filtering on messages posted on OSN user walls poses additional challenges given the short length of these messages other than the wide range of topics that can be discussed. Short text classification has received up to now few attentions in the scientific community.

1.2 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).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 sensation 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. This choice is motivated by related work showing advantages in classifying text and/or short texts using a hierarchical strategy. 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.



1.3 Short Text Classifier

Correct Words. It expresses the amount of terms $\mathbf{t_k}$ € T ∩ K, where $\mathbf{t_k}$ is a term of the considered document dj and K is a set of known words for the domain language.

Bad Words. They are computed similarly to the correct words feature, where the set K is a collection of "dirty words" for the domain language.

Capital Words. It expresses the amount of words mostly written with capital letters, calculated as the percentage of words within the message, having more than half of the characters in capital case. The rationale behind this choice lies in the fact that with this definition we intend to characterize the willingness of the author's message to use capital letters excluding accidental use or the use of correct grammar rules. For example, the value of this feature for the document "To be OR Not to BE" is 0.5 since the words "OR" "Not" and "BE" are considered as capitalized ("To" is not uppercase since the number of capital characters should be strictly greater than the characters count).

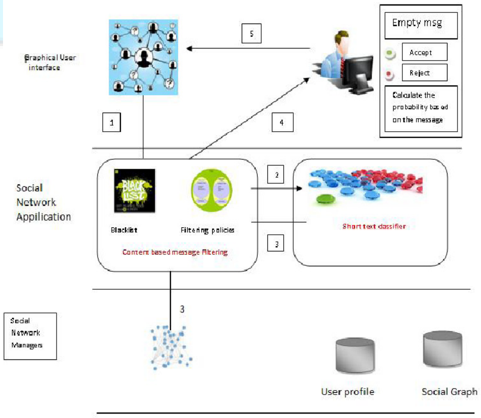
Punctuations Characters It is calculated as the percentage of the punctuation characters over the total number of characters in the message. For example, the value of the feature for the document"Hello!!! how're u doing?" is 5=24. . Exclamation marks. It is calculated as the percentage of exclamation marks over the total number of punctuation characters in the message. Referring to the aforementioned document, the value is 3=5. .

Question Marks It is calculated as the percentage of question marks over the total number of punctuations characters in the message. Referring to the aforementioned document, the value is 1=5. Regarding features based on the exogenous knowledge, CF, instead of being calculated on the body of the message, they are conceived as the VSM representation of the text that characterizes the environment where messages are posted (topics of the discussion, name of the group or any other relevant text surrounding the messages). CFs are not very dissimilar from Bow features describing the nature of data. Therefore, all the formal definitions introduced for the Bow features also apply to CFs.

2. MACHINE LEARNING BASED CLASSIFICATION

We address short text categorization as a hierarchical two level classification process. The first-level classifier performs a binary hard categorization that labels messages as Neutral and non neutral. The first-level filtering task facilitates the subsequent second-level task in which a finer-grained classification is performed. The second-level classifier performs a soft-partition of Non neutral messages assigning a given message a gradual membership to each of the non neutral classes. Among the variety of multiclass ML models well suited for text classification, we choose the RBFN model for the experimented competitive behavior with respect to other state-of-the-art classifiers. RFBNs have a single hidden layer of processing units with local, restricted activation domain: a Gaussian function is commonly used, but any other locally tunable function can be used. They were introduced as a neural network evolution of exact interpolation and are demonstrated to have the universal approximation property.

As outlined in RBFN main advantages are that classification function is nonlinear, the model may produce confidence values and it may be robust to outliers; drawbacks are the potential sensitivity to input parameters, and potential overtraining sensitivity. The first-level classifier is then structured as a regular RBFN. In the second level of the classification stage, we introduce a modification of the standard use of RBFN. Its regular use in classification includes a hard decision on the output values: according to the winner-take-all rule, a given input pattern is assigned with the class corresponding to the winner output neuron which has the highest value. In our approach, we consider all values of the output neurons as a result of the classification task and we interpret them as gradual estimation of multimember ship to classes.The collection of preclassified messages presents some critical aspects greatly affecting the performance of the overall classification strategy. To work well, a ML-based classifier needs to be trained with a set of sufficiently complete and consistent pre classified data. The difficulty of satisfying this constraint is essentially related to the subjective character of the interpretation process with which an expert decides whether to classify a document under a given category. In order to limit the effects of this phenomenon, known in literature under the name of interindexer inconsistency, our strategy contemplates the organization of "tuning sessions" aimed at establishing a consensus among experts through discussion of the most controversial interpretation of messages.



3. RELATED WORK

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. 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 propagation neural 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-specific search engine development and other Web applications such as Web content management.

Recommender systems improve access to relevant products and information by making personalized suggestions based on previous examples of a user's likes and dislikes. Most existing recommender systems use collaborative filtering methods that base recommendations on other users' preferences. By contrast, content-based methods use information about an item itself to make suggestions. This approach has the advantage of being able to recommend previously unrated items to users with unique interests and to provide explanations for its recommendations. We describe a content-based book recommending system that utilizes information extraction and a machine-learning algorithm for text categorization. Initial experimental results demonstrate that this approach can produce accurate recommendations Existing recommender systems almost exclusively utilize a form of computerized matchmaking called *collaborative filtering* or *social filtering.*

4. TEXT CLASSIFICATION

Text classification is the study of classifying textual document into predefined categories. The topic has been extensively studied at SIGIR conferences and evaluated on standard testbeds. There are a number of major approaches. It uses the joint probabilities of words and categories to estimate the probability that a given document belongs to each category. Documents with a probability above a certain threshold are considered relevant to that category. The k-nearest neighbor method is another popular approach to text classification. The categories of these neighbors are then used to decide the category of the given document. A threshold is also used for each category. Neural network programs, designed to model the human neural system and learn patterns by modifying the weights among nodes based on learning examples, also have been applied to text classification. Term frequencies or TFIDF of the terms are used as the input to the network. Based on learning examples, the network can be trained to predict the category of a document. It has been shown that SVM achieved the best performance among different classifiers on the Reuters-21578 data set. In addition to general text documents, classification of Web pages also has been studied. Web pages are often noisy, but they provide additional information about each document. For example, terms marked with different HTML tags (such as titles or headings) can be assigned a higher weight than regular text Terms from neighborhood Web pages also have been used in attempt to improve classification performance. However, it turns out to worsen performance because there are often too many neighbor terms and too many cross linkages between different classes Use of other information about neighborhood Web pages has been proposed. It has been shown that using such additional information improves classification results.

5. FILTERING RULES AND BLACKLIST MANAGEMENT

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:

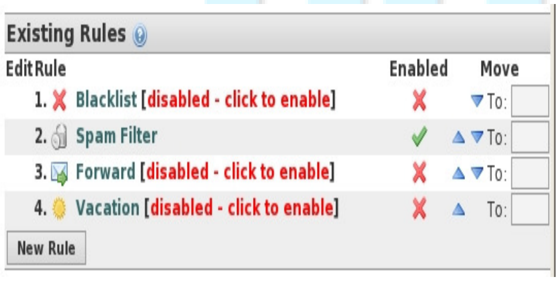
Definition 1 (Creator specification). A creator specification creatorSpec implicitly denotes a set of OSN users. It can have one of the following forms, possibly combined:

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; in-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 in-depth, and a trust value less than or equal to maxTrust.

.

Definition (Filtering rule). A filtering rule FR is a tuple (Author, creator Spec, content Spec, action), where

author is the user who specifies the rule;
creator Spec is a creator specification
content Spec 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(fb block; notify) denotes the action to be performed by the system on the messages matching content Spec and created by users identified by creator Spec.



Existing Rules

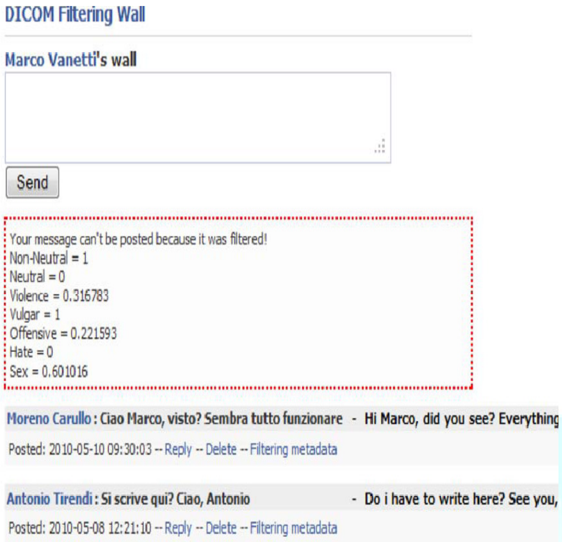
Edit Rule	Enabled	Move
1. Blacklist [disabled - click to enable]	✗	▼ To:
2. Spam Filter	✓	▲ ▼ To:
3. Forward [disabled - click to enable]	✗	▲ ▼ To:
4. Vacation [disabled - click to enable]	✗	▲ To:

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 SNMP; 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.

6. 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.



DICOM Filtering Wall

Marco Vanetti's wall

Your message can't be posted because it was filtered!
Non-Neutral = 1
Neutral = 0
Violence = 0.316783
Vulgar = 1
Offensive = 0.221593
Hate = 0
Sex = 0.601016

Moreno Carullo : Ciao Marco, visto? Sembra tutto funzionare - Hi Marco, did you see? Everything
Posted: 2010-05-10 09:30:03 -- Reply -- Delete -- Filtering metadata

Antonio Tirendi : Si scrive qui? Ciao, Antonio - Do i have to write here? See you,
Posted: 2010-05-08 12:21:10 -- Reply -- Delete -- Filtering metadata

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.

7. FUTURE ENHANCEMENT

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.

REFERENCES

[1] 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.

[2] 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.

[3] 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

[4] R.E. Schapire and Y. Singer, "Boostexter: A Boosting-Based System for Text Categorization," Machine Learning, vol. 39, nos. 2/3, pp. 135-168, 2000.

[5] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," Proc. European Conf. Machine Learning, pp. 137-142, 1998.

[6] A. Uszok, J.M. Bradshaw, M. Johnson, R. Jeffers, A. Tate, J. Dalton, and S. Aitken, "Kaos Policy Management for Semantic Web Services," IEEE Intelligent Systems, vol. 19, no. 4, pp. 32-41, July/Aug. 2004.

[7] L. Kagal, M. Paolucci, N. Srinivasan, G. Denker, T. Finin, and K. Sycara, "Authorization and Privacy for Semantic Web Services,"IEEE Intelligent Systems, vol. 19, no. 4, pp. 50-56, July 2004.