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

An Analysis of Unwanted Messages Filtering Methods from OSN User Walls

Mr. K.Arulmurugan[1], Mr.P.Ranjithkumar[2]

M.E II year, Department of CSE, Sri Subramanya College of Engineering and Technology, Palani, Dindigul, Tamilnadu, India-624 615

Assistant Professor, Department of CSE, Sri Subramanya College of Engineering and Technology, Palani, Dindigul, Tamilnadu, India-624 615

ABSTRACT-One major problem in today's Online Social Networks (OSNs) is to give users skill to regulate the messages posted on their own personal space to avoid that unauthorized data is displayed. OSNs give small support to these needs. In this thesis, i propose a system permit OSN users to have a straight control on the messages posted on their walls. it is achieved through a flexible rule-based system, that permit users to customize the filtering criteria to be put to their walls, and a Machine Learning-based soft classifier automatically labelling messages in endure of content-based filtering. first conduct a set of large-scale measurements with a collection of accounts observe the difference among human, bot, and cyborg in terms of tweeting behavior, tweet content, and account properties. Our experimental evaluation demonstrates the efficacy of the proposed classification system and also we use pattern matching and text classification algorithm for accurate results. In computer science, pattern matching is the act of checking a perceived sequence of tokens for the presence of the constituents of some pattern. The patterns generally have the form of either sequences or tree structures. Uses of pattern matching include outputting the locations of a pattern within a token sequence, to output some component of the matched pattern, and to substitute the matching pattern with some other token sequence.

KEYWRODS: pattern matching, text classification, short text classifier, filtered wall, online social network.

I. 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 Facebook 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 (e.g., [1], [2], [3]).

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, 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.

II. RELATED WORK

In the thesis, first conduct a series of measurements to characterize the differences among human, bot, and cyborg in terms of tweeting behavior, tweet content, and account properties. By crawling Twitter, we collect over 500,000 users and more than 40 million tweets posted by them. Then, perform a detailed data analysis, and find a set of useful features to classify users into the three classes. Based on the measurement results, we propose an automated classification system that consists of four major components.

1. The entropy component uses tweeting interval as a measure of behavior complexity, and detects the periodic and regular timing that is an indicator of automation.
2. The spam detection component uses tweet content tocheck whether text patterns contain spam or not.
3. The account properties component employs useful account properties, such as tweeting device makeup, URL ration, to detect deviations from normal.
4. The decision maker is based on Random Forest, andit uses the combination of the features generated bythe above three components to categorize anunknown user as human, bot, or cyborg.

A.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 [3], [1]. 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 [3]. 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 nonrelevant categories. More complex filtering systems include multilabel 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 preclassified 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 [6], [1]. 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. In [4], a detailed comparison analysis has been conducted confirming superiority of Boosting-based classifiers [1], Neural Networks and Support Vector Machines over other popular methods, such as Rocchio [2] and Naı¨ve Bayesian [4]. 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.

B.Policy-Based Personalization Of Osn Contents

Recently, there have been some proposals exploiting classification mechanisms for personalizing access in OSNs. For instance, in a classification method has been proposed to categorize short text messages in order to avoid overwhelming users of microblogging services by raw data. The system described in focuses on Twitter and associates a set of categories with each tweet describing its content. The user can then view only certain types of tweets based on his/her interests. In contrast, Golbeck and Kuter propose an application, called FilmTrust that exploits OSN trust relationships and provenance information to personalize access to the website. However, such systems do not provide a filtering policy layer by which the user can exploit the result of the classification process to decide how and to which extent filtering out unwanted information. In contrast, our filtering policy language allows the setting of FRs according to a variety of criteria, that do not consider only the results of the classification process but also the relationships of the wall owner with other OSN users as well as information on the user profile. Moreover, our system is complemented by a flexible mechanism for BL management that provides a further opportunity of customization to the filtering procedure. The only social networking service we are aware of providing filtering abilities to its users is MyWOT,3 a social networking service which gives its subscribers the ability to: 1) rate resources with respect to four criteria: trustworthiness, vendor reliability, privacy, and child safety; 2) specify preferences determining whether the browser should block access to a given resource, or should simply return a warning message on the basis of the specified rating.

Despite the existence of some similarities, the approach adopted by MyWOT is quite different from ours. In particular, it supports filtering criteria which are far less flexible than the ones of Filtered Wall since they are only based on the four above-mentioned criteria. Moreover, no automatic classification mechanism is provided to the end user. Our work is also inspired by the many access control models and related policy languages and enforcement mechanisms that have been proposed so far for OSNs (see [2] for a survey), since filtering shares several similarities with access control. Actually, content filtering can be considered as an extension of access control, since it can be used both to protect objects from unauthorized subjects, and subjects from inappropriate objects. In the field of OSNs, the majority of access control models proposed so far enforce topology-based access control, according to which access control requirements are expressed in terms of relationships that the requester should have with the resource owner. We use a similar idea to identify the users to which a FR applies. However, our filtering policy language extends the languages proposed for access control policy specification in OSNs to cope with the extended requirements of the filtering domain.

FILTERED WALL ARCHITECTURE

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

SHORT TEXT CLASSIFIER

Established techniques used for text classification work well on data sets with large documents such as newswires corpora [4], 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 "nonneutral" 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 Nonneutral labels. The second-level soft classifier acts on the crisp set of nonneutral 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.

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. Different sets of features for text categorization have been proposed in the literature [4]; however, the most appropriate feature set and feature representation for short text messages have not yet been sufficiently investigated. Proceeding from these considerations and on the basis of our experience [5], [2],we consider three types of features, BoW, Document properties (Dp) and Contextual Features (CF). The first two types of features, already used in [5], are endogenous, that is, they are entirely derived from the information contained within the text of the message.

III. MODELS AND ASSUMPTIONS

A.Creating user account:

User accounts make it so that several people can easily access their account. Each person can have a separate user account with unique settings and preferences, such as a web background and color theme. User accounts also control the files and programs you can access and what types of changes you can make to the web account. In this a new user can create their user accounts with their mailed, user name, password and photo. After creating account the user can login and also comment in to the pages. We can select the followers so that we can get updates of their comments in our block. The follower's page shows all the users of twitter with their photos and also previous comments of the user.

B.Classifying Twitter:

To develop an automatic classification system, we need a ground-truth set that contains known samples of human, bot, and cyborg. Among collected data, we randomly choose different samples and classify them by manually checking their user logs and homepages. In classification, the data set contains tweets posted by the sampled users in their account lifetime, which we canextract useful features for classification, such as tweeting behaviors and text patterns.

C.Data Collection:

User API will collect the information of active users, increasing the diversity of the user pool. The crawler calls the timeline API to collect the authors of the tweets included in the timeline. Since the Twitter timeline frequently updates, the crawler can repeatedly call the timeline API. During the same time our proposed approach will identify the illicit user account properties.

Twitter API functions support detailed user information query, ranging from profile, follower, and friend lists to posted tweets. In the above crawl, for each user visited, we call API functions to collect abundant information related with user classification. Most information is returned in the format of XML or JSON. We develop some toolkits to extract useful information from the above well-organized data structures. A common strategy shared by bots is following a large number of users (either targeted with purpose or randomly chosen), and expecting some of them will follow back. Following back is considered as a form of etiquette on Twitter.

D.Detecting Account:

In this module, we are portrait the differences among human, bot, and cyborg in terms of tweeting behavior, tweet content, and account properties. The data analysis is to measure the accurate of the tweet. This will validate the tweet content and their account details such as Ip address, Credential and Date time stamp of the received tweet. All details are captured in the dataset to validate the account properties for the user.

IV. OUR PROPOSED SCHEME

I 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 [4] 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 (STC) are concentrated in the extraction and selection of a set of characterizing and discriminant 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 in performing the learning phase creates the premise for an adequate use in OSN domains, as well as facilitates the experimental evaluation tasks. The data analysis is to measure the accurate of the tweet. This will validate the tweet content and their account details such as Ip address, Credential and Date time stamp of the received tweet. All details are captured in the dataset to validate the account properties for the user

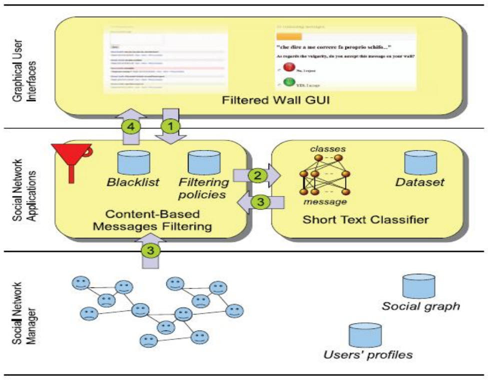

Graphical User Interfaces
"che dire a me correre fa proprio schifo..."
As regards the vulgarity, do you accept this message on your wall?
No, I reject
YES, I accept
Filtered Wall GUI
4
1
Social Network Applications
Blacklist
Filtering policies
Content-Based Messages Filtering
2
3
classes
message
Dataset
Short Text Classifier
3
Social Network Manager
Social graph
Users' profiles

Fig 1 The proposed framework for analysis of unwanted messages filtering methods from osn user walls

V. DETAILS OF THE PROPOSED FRAMEWORK

MACHINE LEARNING-BASED CLASSIFICATION

I 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 Nonneutral. 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 Nonneutral messages assigning a given message a gradual membership to each of the nonneutral classes. Among the variety of multiclass ML models well suited for text classification, we choose the RBFN model [3] 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 [4], and are demonstrated to have the universal approximation property [5]. As outlined in [3], 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 multi membership to classes.

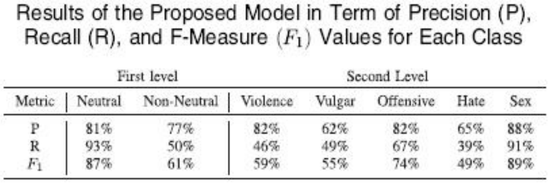
IMPLEMENTATION OUTLINE

Implementation is the most crucial stage in achieving a successful system and giving the user's confidence that the new system is workable and effective. Implementation of a modified application to replace an existing one. This type of conversation is relatively easy to handle, provide there are no major changes in the system.

Each program is tested individually at the time of development using the data and has verified that this program linked together in the way specified in the programs specification, the computer system and its environment is tested to the satisfaction of the user. The system that has been developed is accepted and proved to be satisfactory for the user and so the system is going to be implemented very soon. A simple operating procedure is included so that the user can understand the different functions clearly and quickly.

Performance Measures

In order to provide an overall assessment of how effectively the system applies a FR, we look again at Table 1. This table allows us to estimate the Precision and Recall of our FRs, since values reported in Table 2 have been computed for FRs with content specification component set to (C, 0:5), where $C \in \Omega.$ Let us suppose that the system applies a given rule on a certain message. As such, Precision reported in Table 1 is the probability that the decision taken on the considered message (that is, blocking it or not) is actually the correct one. In contrast, Recall has to be interpreted as the probability that, given a rule that must be applied over a certain message, the rule is really enforced. Let us now discuss, with some examples, the results presented in Table 1, which reports Precision and Recall values.



Results of the Proposed Model in Term of Precision (P), Recall (R), and F-Measure (F_1) Values for Each Class

	First level		Second Level				
Metric	Neutral	Non-Neutral	Violence	Vulgar	Offensive	Hate	Sex
P	81%	77%	82%	62%	82%	65%	88%
R	93%	50%	46%	49%	67%	39%	91%
F_1	87%	61%	59%	55%	74%	49%	89%

Table 1 The proposed model analysis

The second column of table 1 represents the Precision and the Recall value computed for FRs with (Neutral, 0:5) content constraint. In contrast, the fifth column stores the Precision and the Recall value computed for FRs with (V ulgar, 0.5) constraint. Results achieved by the content-based specification component, on the first-level classification, can be considered good enough and reasonably aligned with those obtained by well-known information filtering techniques. Results obtained for the content-based specification component on the second level are slightly less brilliant than those obtained for the first, but we should interpret this in view of the intrinsic difficulties in assigning to messages a semantically most specific category. However, the analysis of the features reported in Table 1 shows that the introduction of contextual information (CF) significantly improves the ability of the classifier to correctly distinguish between nonneutral classes.

VI. CONCLUSION

In this thesis, I 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 I 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 preclassified 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, I plan to address this problem by investigating the use of online learning paradigms able to include label feedbacks from users. Additionally, I plan to enhance our system with a more sophisticated approach to decide when a user should be inserted into a BL.

VII. FUTURE ENHANCEMENT

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.

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