International Journal of Engineering Research-Online
A Peer Reviewed International Journal
Articles available online http://www.ijoer.in

Vol.3., Issue.1, 2015

RESEARCH ARTICLE

ISSN: 2321-7758

Filtering Unwanted Messages from OSN User Walls using text classifier algorithm

N.PRIYA DHARSINI*[1], M.KRISHNA SUDHA[2]

[1]M.Phil Research Scholar, Department of Computer Science, Sri Vasavi College, Erode

[2]Head, Department of Information Technology, Sri Vasavi College, Erode

Article Received: 05/01/2015 | Article Revised on: 12/01/2015 | Article Accepted on:13/01/2015

MECHANICAL ENGINEERING
www.ijoer.in
ENGINEERS
MAKE A WORLD OF DIFFERENCE
International Journal of Engineering Research-Online

ABSTRACT

One fundamental issue in today On-line 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.

KEYWORDS—On-line Social Networks, Information Filtering, Short Text Classification, Policy-based Personalization

INTRODUCTION

Information and communication technology plays a significant role in today's networked society. There is a need to develop more secured mechanisms for different communication technologies, particularly online social networks. Online social networks provide very little support to prevent unwanted messages on user walls. With the lack of classification or filtering tools, the user receives all messages posted by the users he follows. In this work, an information Filtering system is introduced. In online social networks, information filtering can also be used for a different, more sensitive, purpose. This is due to the fact that in online social networks there is the possibility of posting or commenting other posts on particular public/private areas, called 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.

It exploits machine learning 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 original set of aspects, derived from endogenous assets of short texts, is inflamed here including exogenous information associated to the context from which the messages begin. The aim of the present work is to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter out unwanted messages from social network user walls.

2. PROBLEM FORMULATION

Objectives

Even though the Social Networks today, have the restrictions on the users who can post and comment on any user's wall, they do not have any restrictions on what they post. So, some people will use the indecent and vulgar words in commenting on the public posts.

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 classification strategies.

Problem statement

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. 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.

On classification of short text messages integrate messages with meta-information from other information sources such as Wikipedia and Word Net. Sankaranarayanan et al introduced TweetStand to classify tweets as news and non-news. Automatic text classification and hidden topic extraction approaches perform well when there is meta-information or the context of the short text is extended with knowledge extracted using large collections.

A naive approach would ask the user to manually configure her privacy settings for all friends. While this approach may produce perfect accuracy if carried to completion, it also places burden on the user.

Existing system

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.

Drawbacks of existing system

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.

Proposed system

The aim of the present work is therefore to propose and experimentally evaluate an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls. We exploit Machine Learning (ML) text categorization techniques to automatically assign with each short text 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 discriminate features. The solutions investigated in this paper 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 reclassified data.

Advantages of proposed system

A system to automatically filter unwanted messages from OSN user walls on the basis of both message content and the message creator relationships.

The current paper substantially extends for what concerns both the rule layer and the classification module.

Major differences include, a different semantics for filtering rules to better fit the considered domain, an online setup assistant (OSA) to help users in FR specification.

The extension of the set of features considered in the classification process, a more deep performance evaluation study and an update of the prototype implementation to reflect the changes made to the classification techniques.

ANALYSIS

The problems above mentioned are solved with help of some algorithms and techniques.

- Short Text Classifier Techniques
- Machine Learning Techniques

Short text classifier

On datasets with large documents such as newswires corpora, established techniques used for text classification work well but suffer when the documents in the corpus are short. In this context, critical aspects are the definition of a set of characterizing and discriminate features allowing the representation of underlying concepts and the collection of a complete and consistent set of supervised algorithms.

Text representation using endogenous knowledge has a good general applicability, though in operational settings it is appropriate to use also exogenous knowledge. It introduce contextual features (CF) modelling information that characterize the environment where the user is posting. These features play important role in deterministically understanding the semantics of the messages. According to Vector Space Model (VSM) for text representation, a text document *dj* is represented as a vector of binary or real weights *dj* = *w1j*, . . . , *w*|*T*|*j*, where *T* indicates the set of terms that occur at least once in at least one document of the collection *Tr*, and *wkj* € [0; 1] denotes how much term *tk* contributes to the semantics of document *dj*. In the BoW representation, terms are identified with words. For non-binary weighting, the weight *wkj* of term *tk* in document *dj* is computed according to the standard term frequency inverse document frequency (tf-idf) weighting function, defined as Where # (*tk*, *dj*) indicates the number of times *tk* occurs in *dj*, and *#Tr* (*tk*) indicates the document frequency of term *tk*, i.e., the number of documents in *Tr* in which *tk* occurs. DP features are heuristically calculated; their definition stems from intuitive considerations, domain specific criteria and in some cases required trial and error procedures.

Correct words

It represents the amount of terms *tk* € *T* ∩ *K*, where *tk* is a term of the considered document *dj* and *K* is a set of known words for the domain language. This value is normalized by #(t $|T|$ k=1 k, *dj*).

Bad words

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

Capital words

It represents 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. 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 computed 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 6/24.

Exclamation marks

It is computed as the percentage of exclamation marks over the total number of punctuation characters in the message. Referring to the aforesaid document, the value is 2/5.

Question marks

It is computed as the percentage of question marks over the total number of punctuations characters in the message. Referring to the aforesaid document, the value is 2/5.

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.

A quantitative evaluation of the agreement among experts is then developed to make transparent the level of inconsistency under which the classification process has taken place.

Let be the set of classes to which each message can belong to. Each element of the supervised collected set of messages

D = f(mi; ~yi); : : : ; (mjDj; ~yjDj)g is composed of the text mi and the supervised label ~yi 2 f0; 1gjj describing the belongingness to each of the defined classes. The set D is then split into two partitions, namely the training set TrSD and the test set TeSD.

Let M1 and M2 be the first and second level classifier, respectively, and ~y1 be the belongingness to the Neutral class. The learning and generalization phase works as follows:

1). Formula`s used for text classifier algorithm

The two sets TrSD and TeSD are then

TrSD = f(~xi; ~yi); : : : ; (~xjTrSDj; ~yjTrSDj)g

TeSD = f(~xi; ~yi); : : : ; (~xjTeSDj; ~yjTeSDj)g.

2) TrS1 = f(~xj ; ~yj) 2 TrS __ (~xj ; yj); yj = ~yj1g is created for M1.

3)TrS2 = f(~xj ; ~yj) 2 TrS __ (~xj ; ~y0 j); ~y0jk = ~yjk+1; k = 2; : : : ; jjg is created for M2.

4) M1 is trained with TrS1 with the aim to recognize whether or not a message is non-neutral. The performance of the model M1 is then evaluated using the test set TeS1.

5) M2 is trained with the non-neutral TrS2 messages with the aim of computing gradual membership to the non-neutral classes. The performance of the model M2 is then evaluated using the test set TeS2.

6) To summarize, the hierarchical system is composed of M1 and M2, where the overall computed function f : Rn ! Rjj.

CONCLUSION

A system to filter unwanted message in OSN wall is presented. The first step of the project is to classify the content using several rule. Next step is to filter the undesired rules. Finally Blacklist rule is implemented. So that owner of the user can insert the user who posts undesired messages. Better privacy is given to the OSN wall using our system. In this system to prevent the indecent messages from the Social Networking site walls has been presented. The usage of sort text classification has given higher results to the system to trace the messages and the users to distinguish between the good and bad messages and the authorized and unauthorized users in the Social Networking User Profiles automatically.

Thus the short text classification technique plays a vital role in this project in order to generate the blacklist of the bad words and the unauthorized users. The user has to update his privacy setting in his account in order to add this method to prevent the obscenity in his public profile. In this context, a statistical analysis has been conducted to provide the usage of the good and bad words by the persons in the sites. Overall, the obscenity of the users has been prevented.

SCOPE FOR FUTURE ENHANCEMENT

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. Plan to address this problem by investigating the use of on-line learning paradigms able to include label feedbacks from users.

Additionally, plan to enhance our system with a more sophisticated approach to decide when a user should be inserted into a BL. The development of a GUI and a set of related tools to make easier BL and FR specification is also a direction plan to investigate, since usability is a key requirement for such kind of applications. In particular, aim at investigating a tool able to automatically recommend trust values for those contacts user does not personally known.

As future work, intend to exploit similar techniques to infer BL rules and FRs. Additionally plan to study strategies and techniques. Limiting the inferences that a user can do on the enforced filtering rules with the aim of by passing 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

REFERENCE

[1]. P. Bonatti and D. Olmedilla, “Driving and monitoring provisional trust negotiation with metapolicies,” in In 6th IEEE International Workshop on Policies for Distributed Systems and Networks (POLICY 2005). IEEE Computer Society, 2005, pp. 14–23.

[2]. C. Bizer and R. Cyganiak, “Quality-driven information filtering using the wiqa policy framework,” Web Semantics: Science, Services and Agents on the World Wide Web, vol. 7, pp. 1–10, January 2009.

[3]. D. D. Lewis, Y. Yang, T. G. Rose, and F. Li, “Rcv1: A new benchmark collection for text categorization research,” Journal of Machine Learning Research, 2004.

[4]. M. Carullo, E. Binaghi, and I. Gallo, “An online document clustering technique for short web contents,” Pattern Recognition Letters,vol. 30, pp. 870–876, July 2009.

[5]. M. Carullo, E. Binaghi, I. Gallo, and N. Lamberti, “Clustering of short commercial documents for the web,” in Proceedings of 19th International Conference on Pattern Recognition (ICPR 2008), 2008.

[6]. C. D. Manning, P. Raghavan, and H. Sch¨utze, Introduction to Information Retrieval. Cambridge, UK: Cambridge University Press, 2008.

[7]. B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, and M. Demirbas, “Short text classification in twitter to improve information filtering,” in Proceeding of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR , p. 841–842,2010 .

[8]. J. Golbeck, “Combining provenance with trust in social networks for semantic web content filtering,” in Provenance and Annotation of Data, ser. Lecture Notes in Computer Science, L. Moreau and I. Foster, Eds. Springer Berlin / Heidelberg, , vol. 4145, pp.101–108,2006.

[9]. F. Bonchi and E. Ferrari, Privacy-aware Knowledge Discovery: Novel Applications and New Techniques. Chapman and Hall/CRC Press, 2010.

[10]. J. A. Golbeck, “Computing and applying trust in web-based social networks,” Ph.D. dissertation, PhD thesis, Graduate School of the University of Maryland, College Park, 2005.