

International Journal of Advance Research in Computer Science and Management Studies

Research Article / Survey Paper / Case Study

Available online at: www.ijarcsms.com

Novel Approach to Filter Unwanted Messages in On-line Social Networks

Sana Patka¹

Department of Computer Science and Engineering,
Priyadarshini Institute of Engineering & Technology
Nagpur University, Maharashtra – India

S. D. Zade²

Department of Computer Science and Engineering,
Priyadarshini Institute of Engineering & Technology
Nagpur University, Maharashtra – India

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

Keywords: *Filtering Rules, Online Social Network, Content Based Filtering, Machine Learning.*

I. INTRODUCTION

Social networking sites share a variety of technical features that allow individuals to: construct a public profile & can keep some contents of profile private visible to selected friends, articulate a list of other users that they share a connection and can see their list of connections within the system. The basic of these are visible profiles containing a list of "friends" those are also users of the site. A profile is generated from obtaining answers of many common questions, such as age, location, interests, etc. Many sites allow users to upload images or pictures, and can even add multimedia content to modify the view of the profile. Others, e.g., Facebook, allow users to improve their profile by providing facility to insert Applications. Many sites permit the users to post their status or blogs, search for other peoples with similar interests and share contacts. Generally User profiles consist an area to comments to friends and other users. Social networks typically have controls that allow users to choose who can view their profile & contact them, include them to the list of contacts, and so on, all these will be done to protect user privacy, but up to now OSNs provide little support to give users the ability to control the messages posted on their own private space to avoid unwanted content to be displayed.

Our aim is therefore to propose an automated system, called Filtered Wall (FW), able to filter unwanted messages from OSN user walls for this purpose we will going to use Machine Learning (ML) based classification, the ability of a machine to improve its performance based on previous results[2]. In OSNs, information filtering can also be used for a sensitive issues. Because the fact that in OSNs there is the possibility of posting or commenting other posts on particular public/private areas, called in general walls. However, no content-based preferences are supported and therefore it is not possible to prevent undesired messages, such as political or vulgar ones. So Information filtering can be used to give users the ability to automatically control the messages written on their own walls, by filtering out unwanted or undesired messages[10],[14]. According to us this is an important OSN service that has not been provided so far. Even today OSNs provide very less support to prevent unwanted messages on user walls. As the learning model is concerned, we will be going to use neural learning which

is one of the most efficient solutions in text classification. Other than the classification facilities, the system will also provides a powerful rule layer that will exploit a flexible language to specify Filtering Rules (FRs), with the help of which users can state what contents should not be displayed on their walls. FRs can support a number of different filtering criteria that can be combined and customized according to the user needs. In addition, the system provides the support for user defined BlackLists (BLs) contains list of users, which will provide the facility of temporary prevention to post any kind of messages on a user wall.

II. RELATED WORK

Our work has a association both with the content-based filtering & also with the field of policy-based system for OSNs and web contents

A. Content based filtering

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 [8]. While electronic mail was the original domain of early work on information filtering, many related papers have addressed different domains including newswire articles, Internet “news” articles, and broader network resources[13],[11]. Documents processed in content-based filtering are mostly textual in nature and this makes content-based filtering close to text classification. The task of filtering can be modeled, in fact, as a case of classification, dividing all incoming documents into non relevant and relevant 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.

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. Upto now Short text classification has received little attention in the scientific community[4]. Recent work highlights difficulties in defining many robust features, essentially because of the fact that the description of the short text is concise with many spellings mistakes, non standard terms and noise. Zelikovitz and Hirsh attempt to improve the classification of short text strings developing a semi supervised learning strategy based on a combination of labeled training data plus a secondary corpus of unlabeled but similar large documents. This solution is not applicable in our domain in which short messages are not summary and even not a part of large semantically related documents.

A different approach is proposed by Bobicev and Sokolova that circumvent the problem of error-prone feature construction by adopting a statistical learning method that can perform reasonably good without having feature engineering. However, Prediction by Partial Mapping method, generates a language model that is used in probabilistic text classifiers which are hard classifiers in nature and do not easily integrate soft classifier that is multi-membership paradigms. In our scenario, for defining flexible policy-based personalization strategies [5] we consider gradual membership to classes a key feature.

B. Policy-based personalization of OSN contents

Policy based personalization has been applied to number of contexts. According to user defined policies it adapts a service in specific context. The policy based system which focuses on Twitter. It define & allocates a category to every tweet and display only those tweet to the user which are of interest. In this scenario, policy based personalization represent the ability of the user to filter wall messages according to filtering criteria specified by user. In contrast, Golbeck and Kuter[15] propose an application, given a name FilmTrust, that makes the use of OSN trust relationships and provenance information to personalize access to the website. However these kind of systems doesn't provide a policy layer for filtering by which the user can exploit the result of the classification process to decide how and to which extent filtering out the information which are not wanted. In contrast, our filtering policy language allows the setting of FRs according to a different criterias, that will not consider only the output or 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 [7]. 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 social networking service which gives its subscribers the ability to: 1) rate resources according to four criteria: truthfulness, trader or vendor reliability, privacy, and safety of child 2) specify preferences determining whether the browser should block access to a given or specified resource, or should simply give a warning message on the basis of the specified rating. In spite of the presence of some similarities, the method adopted by MyWOT is somehow different from ours. Particularly it supports filtering criteria which are less flexible than the ones of Filtered Wall since they are only based on the four criteria mentioned above. Moreover, there is no automatic filtering mechanism is provided to the end user.

III. RESEARCH METHODOLOGY

In Accordance to the research some prime methodologies are Short Text Classifier (STC) in which a Text representation, Machine learning based classification is there and another is the Content Based Message Filtering (CBMF).

A. 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 are short. Within 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 uniform set of supervised examples. Our goal of study is designing and evaluating various representation techniques in combination with a neural learning strategy to categorize short texts semantically. Keeping ML point of view in mind, 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[12]. 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 as neutral and non neutral crisp set. In the second level soft classifier acts on the crisp set of non-neutral short texts and, for each of them, it “simply” generates estimated appropriateness or “gradual membership” for each of the conceived classes, without considering any “hard” decision on any of them. Such type of a grades is then used by the subsequent phases of the filtering process [1][3].

B. Classification depends or based on machine learning

We consider short text categorization as a hierarchical two-level process of classification. In the first-level, the classifier will perform a binary hard categorization that labels messages or text as Neutral and Non-Neutral. The first-level filtering action facilitates the subsequent second-level task in which a classification is performed which is finer-grained. The second-level classifier performs a task of soft-partition of Non-neutral text or messages assigning a given message a gradual membership to each of the non neutral classes. Among the number of multi-class ML models for text classification, for the experimented behaviour we adopt Neural Network with respect to other state of the art classifiers [6], [9].

C. Content based message filtering

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 [8]. While electronic mail was the original domain of early work on information filtering, many related papers have addressed different domains including newswire articles, Internet “news” articles, and broader network resources[13],[11]. Documents processed in content-based filtering are mostly textual in nature and this makes content-based filtering close to text classification. The task of filtering can be modeled, in fact, as a case of classification, dividing or

partitioning all incoming documents into relevant and non relevant 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.

IV. PROPOSED SYSTEM

So our aim is to build a system that allows OSN users to have a direct control on the messages posted on their walls. Avoid messages from undesired creators (blacklist user), independent from their contents and allow users to state constraints on messages (filtering rules). Following is the flowchart of proposed system.

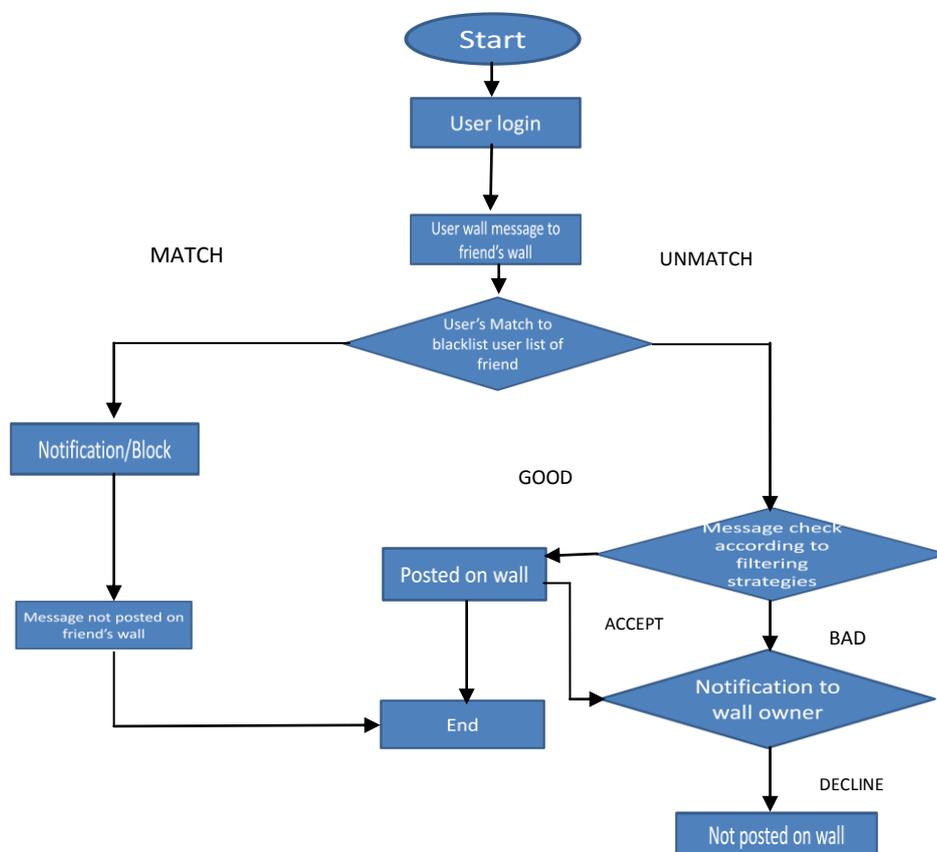


Figure 1

Machine Learning

Machine Learning is “The ability of a machine to improve its performance based on previous results.” We are using basic schema of a Artificial Neural Networks for machine learning, also known as “Artificial neural nets” or ANN in short, are a computational tool used on the interconnection of the neuron in the nervous systems of the human brain and that of other organisms. Biological Neural Nets (BNN) is the naturally occurring similar to the ANN. Artificial neural networks are very different from biological networks, although many characteristics of biological systems are faithfully reproduced in the artificial systems. Artificial neural nets are a kind of non-linear processing system that is ideally suited for a many range of tasks, especially for those tasks where there is no existing algorithm for completion of task. ANN can also be used to trained & to solve certain problems using a teaching method and sample data. In this way, similar type of constructed ANN can be used to perform different tasks depending on the training received. When proper training is provided, ANN can be capable of generalization, the ability to identify similarities among different input patterns that have been corrupted by noise.

Filtering rule & blacklist management

The admin will create a bad word dictionary which consist of number of badwords according to the survey of swear words on the internet then the admin will create some rules which will be applied on each text message that will be posted on user

wall, if any kind of word present in the messages matches with badword dictionary it means the badword is detected & notification will be sent to the wall owner then it will depend on the wall owner to accept or deny the message.

A further component of our system is a BL mechanism to avoid messages from unrecognized creators, which are not depending on their contents. Black list are directly managed by the system admin or by the system, which should be able to determine which users to be inserted in the BL and decide when users retention in the BlackList will finished. To increase flexibility, such information is given to the system through a set of rules called BL rules. For example If any user1 who will sent bad message more than 5 times repeatedly to a particular user2 then the user1 will be automatically blocked & then after the duration of 3 months the blocked user will automatically unblocked.

For checking the severity of the word how bad the word is, it is categorized as badword & illegal word for this purpose we have used a fuzzy logic for ranking the words Fuzzy logic is a form of many-valued logic; it deals with reasoning that is approximate rather than fixed and exact. Compared to traditional binary sets (where variables may take on true or false values), variables in fuzzy logic may have a truth value that can ranges between 0 and 1. The admin will set a rank for each word i.e illegal word is having more rank than a badword as soon as the badword will detected its severity will be checked if the calculated value of word is equal to the rank given to the illegal word or more than that the detected badword will be consider as illegal word.

The following screen shot is showing the notification that the message is having a badword & it will be on the wall owner to accept or deny

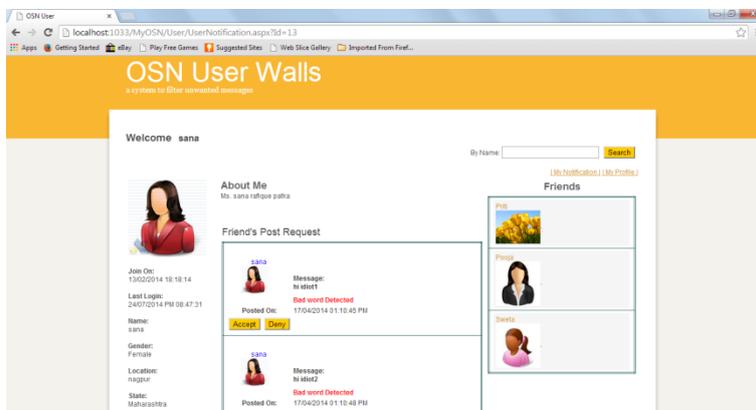


Figure 2

The figure 3 will show that the user has accepted the message having badword & then that message is displayed on user wall but the badword will be displayed in the form of asterisk i.e *****.

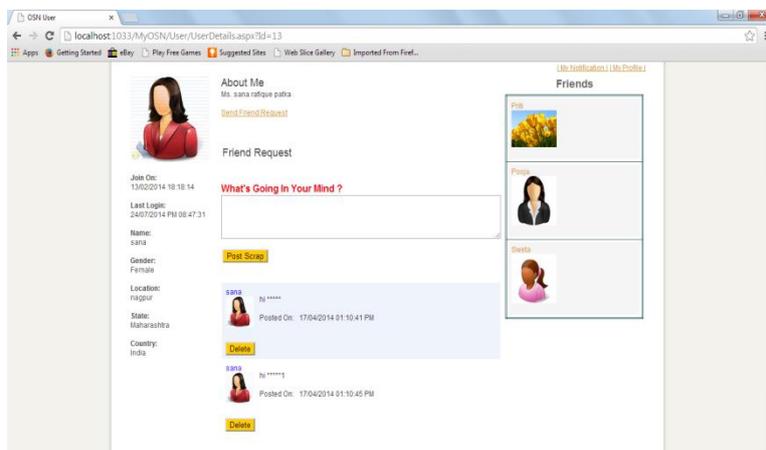


Figure 3

V. CONCLUSION

Thus this paper describes different approaches to filter unwanted messages in online social network. Additionally the flexibility of a system can be enhanced through filtering rules and blacklist management and also we studied strategies and techniques limiting the inferences that a user can do on the enforced filtering rules with the aim of bypassing the filtering system. With the help of blacklist rules created users those are sending unwanted messages continuously will be blocked automatically after some attempts.

References

1. Marco Vanetti, Elisabetta Binaghi, Elena Ferrari, Barbara Carminati, Moreno Carullo, "A System to Filter Unwanted Messages from OSN User Walls", IEEE Transactions on Knowledge and Data Engineering vol:25 year 2013.
2. 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.
3. 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.
4. C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval. Cambridge, UK: Cambridge University Press, 2008.
5. G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Information Processing and Management, vol. 24, no. 5, pp. 513–523, 1988.
6. J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing units," Neural Computation, vol. 1, p. 281, 1989.
7. M. J. D. Powell, "Radial basis functions for multivariable interpolation: a review," pp. 143–167, 1987.
8. 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.
9. 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.SS
10. D. D. Lewis, "An evaluation of phrasal and clustered representations on a text categorization task," in Proceedings of 15th ACM International Conference on Research and Development in Information Retrieval (SIGIR-92), N. J. Belkin, P. Ingwersen, and A. M. Pejtersen, Eds. ACM Press, New York, US, 1992, pp. 37–50.
11. V. Bobicev and M. Sokolova, "An effective and robust method for short text classification," in AAAI, D. Fox and C. P. Gomes, Eds. AAAI Press, 2008, pp. 1444–1445.
12. 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 2010, 2010, pp. 841–842.
13. S. Pollock, "A rule-based message filtering system," ACM Transactions on Office Information Systems, vol. 6, no. 3, pp. 232–254, 1988.
14. P. E. Baclace, "Competitive agents for information filtering," Communications of the ACM, vol. 35, no. 12, p. 50, 1992.
15. 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, 2006, vol. 4145, pp. 101–108.

AUTHOR(S) PROFILE



Sana Patka received the Bachelor of engineering degree in Computer Science & Engineering from Tulsiramji gaikwad patil college of engineering & Technology in 2011 and Diploma in computer engineering from Anjuman Polytechnic in 2008, Currently pursuing M.tech in computer Science & engineering from Priyadarshini Institute of engineering & tech respectively.