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. 3, Issue 12, December 2015

A Feedback and Threshold Based Filtering System for OSN-Online Social Networks

Shareen S. Anthony, Prof. Nitin A.Shelke

M.E. Final Year CSE, Dept. of CSE, GHRCEM, Amravati, India

Assistant Professor, Dept. of CSE, GHRCEM, Amravati, India

ABSTRACT: Now-a-days Online Social Networks (OSNs) has become the most important medium to share post and distribute a huge amount of human life information. Information filtering (text filtering) is also used in OSNs for more different and responsive function. This is because of the fact that in OSNs there are chances or possibility of posting or commenting other posts on particular public/private regions, called in general walls. Just because of this fact information filtering can be used to give users the ability to automatically control the messages written on their own walls, by filtering out undesired messages. A very little support is provided by OSNs to prevent undesired or unwanted messages on user walls. The main aim is to propose and experimentally evaluate an automated system, called Filtered Wall (FW), which can be able to filter out undesired or unwanted messages from OSN walls. Moreover, this work will also consider the feedback of the sink, so if any user doesn't have any objection on such messages from his/her friend then OSN must not interfere in it. But to built such filter a sender user must be trust worthy and hence trust value of each user in OSN must be calculated and based on user's trust value the filter should work. This work presents an approach to calculate trust value of user based on the feedback given to them by receiver user for each post.

KEYWORDS: Online Social Networks, Machine Learning, Filtering Rules, Content-based filtering, Filtering system, Threshold and Trust value, FB, FR, etc.

I. INTRODUCTION

A significant role is played by information and communication technology in today's networked society. An online interaction between users, who are aware of issues related to security applications and their impact on personal privacy are greatly affected by this. It became the necessity to develop more mechanism related to security for various communication technologies which mainly includes online social networks (OSNs). A very little support is provided by OSNs to prevent undesired or unwanted messages on user walls. Due to the lack of filtering tools or classification, the user of OSN receives all the messages posted by other users he or she follows. In other cases, the user usually receives a noisy stream of updates. In this paper, a threshold and feedback based information Filtering system is introduced. The main aim of the system is to focus on only one kind of feeds as: Make list of manually selected group of users on OSN. Make list of feeds which generally aims to focus on specified topics; however it still produces noisy due to irrelevant messages. Therefore, we propose an online filtering system, which tends to extracts such topics in a list, hence filtering irrelevant messages which produce noise [1].

Information filtering can also be used for a different, more sensitive, purpose in OSNs. This is owing to the fact that in OSNs there are chances to post or comment other posts on particular public/private areas, called in general walls. Information filtering is therefore used in the proposed system to give users the ability to a control the messages written on their own walls automatically, by just filtering out unwanted or undesired messages. Hence the purpose of the present work is to propose an automated system able to filter the undesired messages. Machine Learning (ML) text categorization techniques [2] are exploited to assign automatically with each short text message a set of categories based on its content. Selection and extraction of discriminate features are the main efforts in building a robust short text classifier.

Unfortunately, all the content in OSNs is generated by users and is not necessarily legitimate. The posted messages could be spam. So it is necessary to restrict that unwanted message. Today OSNs provide very little support to prevent unwanted messages on user walls. 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.

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. 3, Issue 12, December 2015

Despite the efforts in the fields mentioned above, other important issues have been explored include user privacy, trustworthiness and context-aware recommendation. One of user concerns to use recommender systems freely and comfortably is user privacy. Users are usually reluctant to disclose their private information such as purchase, reading, browsing records. However, most current filtering algorithms need to obtain user private information for further analysis and recommendation services. Some work has studied on how to protect user privacy in recommender systems. Current filtering techniques assume that user ratings are trustable and treat all users equally. However, some may argue that the opinions of experts should be more emphasized than that of novices.

II. PROPOSED SYSTEM

The main goal of the proposed work is to design an online message filtering system that is categorized or arranged at the service provider side of OSN. Once categorized at the service provider side, it inspects each and every message before delivering it to the intended recipients and takes the decision immediately on whether the inspected message should be dropped or not.

A. *Filtering Rules*

1) Input

Filtering Rules are customizable by the user. User can have authority to decide what contents should be blocked or displayed on his wall by using Filtering rules. For specify a Filtering rules user profile as well as user social relationship will be considered.

FR= {Trustier, SOUs, Rule, TuV}

FR is dependent on following factors

- Trustier
- Set of Users (SOUs)
- Rule
- Action

Trustier is a person who defines the rules.

SOUs denote the set of OSN user.

Rule is a Boolean expression defined on content.

2) Process

FR= {SOUs, Rule==category (Violence, Vulgar, offensive, Hate, Sexual), TuV}

- FR
- SOUs
- Rule
- TuV

Here,

FR Block Messages at basic level.

SOUs Denotes set of users

Rule Category of specified contents in message.

TuV is the trust value of sender.

In processing, after giving input message, the system will compare the text with the different categories which are prevented. If message found in that prevented type of category then message will display to the user that "can't send this type of messages", and still the user wants to send the message he/she can continue with sending the message. After getting the message the Trustier will give the Feedback (FB) to the sender and the sender will gain the TuV accordingly.

Process denotes the action which is to be performed by the system on the messages matching Rule and created by users identified by SOUs.

E.g. FR== {Friends, Rule==category (Vulgar, Sexual), TuV>50}

i.e. Trustier will accept the message from friends but message should not contain vulgar or sexual words. Message containing such words will affect the TuV of sender.

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. 3, Issue 12, December 2015

3) Output
PFM= {Rule, M||Y}

- PFM Percentages of filtered message in a year or month.

In general, more than a filtering rule can apply to the same user. A message is therefore published only if it is not blocked by any of the filtering rules that apply to the message creator.

B) *Trust Value Calculations*

The trust value of any user in OSN is dependent on the feedback they gain by the user to whom they sent a message. Feedback from the user must also be trust worthy. That's why the FB can be categorized into following:-

- Positive with content (PC) - Good FB, message is acceptable with objectionable content. This will increase the TuV of sender.
- Positive without content (PWC) - Good FB, message is acceptable as this message does not contain objectionable content. This will increase the TuV of sender.
- Negative with content (NC) - Bad FB, such messages must not be sent again, which are against the Rule. This will decrease the TuV of sender.
- Negative without content (NWC) - Bad FB, message doesn't contain any objectionable content but the Trustier is giving negative FB. Such type of FB from Trustier will affect the TuV of its own, and the TuV of sender will remain same.

So, based on above categories the TuV will be calculated as follows:-
FB as 1 and 2 TuV= TuV+ abs [(PC+PWC)/ (NC+NWC)]
FB as 3 TuV= TuV-[1+ (NC+NWC) / (PC+PWC)] for [(NC+NWC) / (PC+PWC)] <1
Otherwise, send system generated message to sender, FB Negative with content exceeds limit of Threshold Value (ThV) and deduct 5 points from TuV, so TuV=TuV-5.
FB as 4 TuV= TuV of sender, but TuV= TuV-[1+ (NC+NWC) / (PC+PWC)] for Trustier.

C) *Blacklists*

BLs is directly managed by the system.
This should be able to determine the users to be inserted into the blacklist and decide when to retain user back from the blacklist. To enhance flexibility, such information is given to the system through a set of rules, hereafter called BL rules.
1) BL rules
INPUT = {Sender, FB, TuV, ThV} Where

- Sender is the OSN user who is sending the message;
- FB is the FeedBack gain by the sender after sending the message
- TuV is the new Trust Value calculated as formulas specified in A.3.
- ThV is the Threshold Value.

BL Rules:
ThV= PC+ PWC when, PC+PWC=NC+NWC.
For sender, when 5 points are deducted by system, which means sender cross the ThV put sender into BL for a specific duration.
For Trustier, after giving feedback, check ThV, if true, put Trustier in BL for specific duration.

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. 3, Issue 12, December 2015

Following Figure 1 shows the DFD for Mathematical Model.

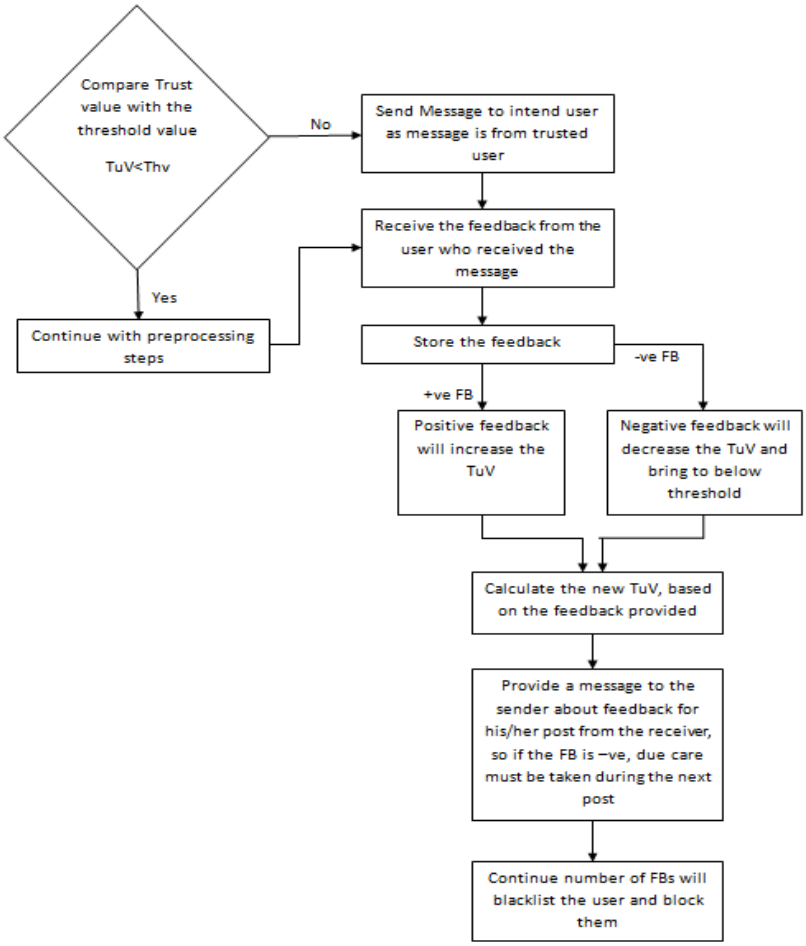

Fig.1. DFD Mathematical Model

This functionality will increase the performance of overall system by applying the pre-processing only on the message from un-trusted user. Moreover, this will give rights to the user in OSN whose who want to send and accept messages containing unwanted words legitimately.

III. SIMULATION RESULTS

A) Computational Analysis

The following measuring criteria are used to evaluate the strength and accuracy of the stemmer.
1) Index Compression Factor (ICF): The index compression factor represents percent by which a collection of distinct words is reduced by stemming. Higher the number of words stemmed, greater the strength of the stemmer. This is calculated as:

$$ICF = \frac{(N-S)}{S} \times 100$$

Where,
N – Number of distinct words before stemming.
S – Number of distinct stems after stemming.

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. 3, Issue 12, December 2015

2) Word Stemmed Factor (WSF): It is the percentage of words that have been stemmed by the stemming process out of the total words in a sample. Larger the number of words stemmed, greater the strength of the stemmer. Minimum threshold for this factor should be 50%.

$$WSF = \frac{WS}{TW} \times 100$$

Where,
WS – Number of words stemmed.
TW – Number of words in a sample.
3) Correctly Stemmed words Factor (CSWF): It is the percentage of words that have been stemmed correctly out of the number of words stemmed. Higher the percentage of this factor, higher will be the accuracy of the stemmer. Minimum threshold for this factor should be 50%.

$$CSWF = \frac{CSW}{WS} \times 100$$

Where,
CSW – Number of correctly stemmed words. WS – Total number of words stemmed.

4) Average Words Conflation Factor (AWCF): This indicates the average number of variant words of different conflation group that are stemmed correctly to the root words. To calculate AWCF, we first compute the number of distinct words after conflation as:

$$NWC = S - CW$$

Where,
CW – Number of correct words not stemmed. Thus, Word Conflation Factor is obtained as:

$$AWCF = \frac{CSW - NWC}{CSW} \times 100$$

Higher the percentage of AWCF, higher will be the accuracy of the stemmer.
For example, if the corpus have variant words of conflation group, such as, “accepts”, accepted”, “accepting”, “acceptance”, “acceptable”, “acceptances”, acceptation”, and the stemming algorithm transformed 5 words among them correctly to the root word, either changing others to improper stem, or living unchanged. Then the word conflation factor for this class is 5/7*100 i.e. 71%. Thus the average words conflation factor (AWCF) gives the average percentage of variant words of different conflation groups that are transformed to the correct root word.

To evaluate the performance of the stemmer described in this paper, we have applied implemented stemming algorithm to the sample vocabulary downloaded from the web site [17]. It contains 29714 distinct words, arranged into “conflation groups”. Some of them are incorrect words. For example, there are 287 incorrect words in the sample of 1858 words which begin with alphabet ‘a’. Table 1 shows the index compressions factor applied to all vocabulary of 29714 words. To measure the strength and accuracy of stemmer, we considered a sample of 1858 words containing ‘a’ alphabet words and analyze the result using the measuring criteria specified in section previous section. The result of is noticeable aggressive stemmers referred in this paper is shown in Table 2.

Table1. Index Compression Factor

Analysis of stemmers (Alphabet words)	Porter1 Stemmer	Porter2 Stemmer
Total Words (TW)	1858	1858
Number of Distinct words before stemming (N)	1571	1571
Number of Distinct words after stemming (S)	756	727
Index Compression Factor (ICF)	51.88	53.72
Number of words Stemmed(WS)	1248	1237
Words Stemmed Factor (WSF)	67.17	66.58

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. 3, Issue 12, December 2015

Correctly Stemmed words (CSW)	399	430
Incorrectly stemmed words (ISW)	849	807
Correctly Stemmed words Factor (CSF)	31.97	34.76
Correct Words not stemmed (CW)	323	334
No. of Distinct words after conflation (NWC)	433	393

B) Experimental Analysis

The developed application is presented to some users and they rated 1 to 10 marks out of 10, the application based on the following parameters:-

- GUI- Graphical User Interface (GUI)
- Navigation- Site Navigation (N)
- Filtering Efficiency- Ability to Filter the Unwanted words (FE)
- Feedback- How Useful the Feedback System Is? (F)
- Overall Performance (OP)

Table2. User Ratings

Parameters → / Users ↓	GUI	N	FE	F	OP	Average
User1	8	8	10	9	9	8.8
User2	7	7	9	8	7	7.6
User3	7	9	10	7	8	8.2
User4	8	6	8	9	7	7.6
User5	9	8	9	10	9	9

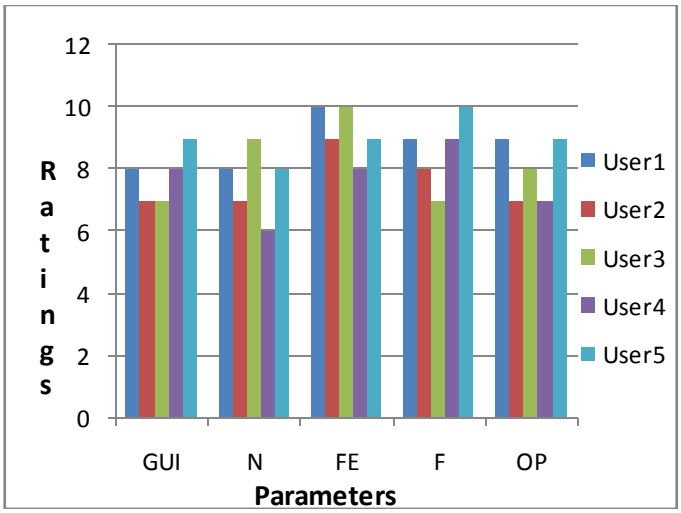

Fig.2. Graph for User Ratings

IV. CONCLUSION

In this paper, we describe our work to provide unwanted message filtering for social networks. A system to filter out undesired messages from OSN user wall is presented in this paper. Messages from OSN walls.ML soft classifier is exploited by the system for enforcing customizable content-dependent FRs. Further, the system's flexibility in terms of filtering options is greatly enhanced through the management of BLs. It's remarkable to note that proposed system in

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. 3, Issue 12, December 2015

this paper focuses on just the set of functionalities that are must to provide a sophisticated tool for OSN message filtering. Additionally, we studied strategies and techniques limiting the inferences that a user can do on the enforced filtering rules with the goal to bypass 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. Miss. Shareen Anthony, Mr. Nitin Shelke, "A Survey of Filtering System for OSN (Online Social Networks)", IJCST, 2014.
2. G. Amati and F. Crestani, "Probabilistic learning for selective dissemination of information," Information Processing and Management, vol. 35, no. 5, pp. 633–654, 1999.
3. H. Sch¨utze, D. A. Hull, and J. O. Pedersen, "A comparison of classifiers and document representations for the routing problem," in Proceedings of the 18th Annual ACM/SIGIR Conference on Resea. Springer Verlag, 1995, pp. 229–237.
4. G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval, Information Processing and Management, vol. 24, no. 5, pp. 513–523, 1988.
5. A. K. Jain, R. P.W. Duin, and J. Mao, "Statistical pattern recognition: A review," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 4–37, 2000.
6. Cleverdon, "Optimizing convenient online access to bibliographic databases," Information Services and Use, vol. 4, no. 1, pp. 37–47, 1984.
7. J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," Biometrics, vol. 33, no. 1, pp. 159– 174, March 1977.
8. J. Nin, B. Carminati, E. Ferrari, and V. Torra, "Computing reputation for collaborative private networks," in Proceedings of the 2009 33rdAnnual IEEE International Computer Software and Applications Conference - Volume 01. Washington, DC, USA: IEEE Computer Society, 2009, pp. 246–253.
9. K. Strater and H. Richter, "Examining privacy and disclosure in a social networking community," in Proceedings of the 3rd symposium on Usable privacy and security (SOUPS 2007). New York, NY, USA: ACM, 2007, pp. 157–158.
10. 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.
11. E. D. Wiener, J. O. Pedersen, and A. S. Weigend, "A neural network approach to topic spotting," in Proceedings of 4th Annual Symposium on Document Analysis and Information Retrieval (SDAIR-95), Las Vegas, US, 1995, pp. 317–332.
12. 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.
13. S. Dumais, J. Platt, D. Heckerman, and M. Sahami, "Inductive learning algorithms and representations for text categorization," in Proceedings of Seventh International Conference on Information and Knowledge Management (CIKM98), 1998, pp. 148–155.
14. Y. Zhang and J. Callan, "Maximum likelihood estimation for filtering thresholds," in Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 2001, pp. 294–302.
15. L. Fang and K. LeFevre, "Privacy wizards for social networking sites," in Proceedings of the 19th international conference on World Wide Web (WWW 2010). New York, NY, USA: ACM, 2010, pp. 351–360.