

Filtering Unwanted Messages from (OSN) User Wall's Using MLT

Prof.Sarika.N.Zaware¹, Anjiri Ambadkar², Nishigandha Bhor³, Shiva Mamidi⁴,
Chetan Patil⁵

¹Department of Computer Engineering, AISSMS IOIT, Savitribai Phule

Pune University, Pune, India

ABSTRACT

The most common issue in today's online social networking sites are walls of user get affected and spammed by a spammer. Spam being unwanted messages, Vulgar messages etc. By observing all such issues in mind we developed software which helps to OSN user to protect their wall from spammer and software provides a direct control while posting comments on wall. we achieve this by obtaining the dataset of user details and putting these dataset through Binary TF and Normalization, also using machine learning concepts which automatically identify and labels the comment as spam, Vulgar or legit.

KEYWORDS

Machine Learning Techniques (MLT), Online Social Network (OSN), Term Frequency (TF), Spam, Filtering Wall, Cosine Similarity, Normalisation, and Dataset.

1. INTRODUCTION

All of we know that today's Online social Networking sites are strong and most popular communication modem for staying connected with people. Also we know that famous entities or public figures made their social networking fan pages in order to have a direct interaction with their fans. Uses OSN's are exchange of data, leisure, share and spread of information such that many kind of interactions are done by Online Social Networks. Data can be images, audio, video, texts these types of data are spread over the network in all over

the world in large amounts, a statistics on facebook reveals that regular user roughly shares around 85-90 pieces of content(i.e. link, gossips, stories, news, advertisements, images, audio, video, etc.) in every month. The Large amount of data creates a scope for web content mining strategy for locating useful information automatically within whole data by which the OSN management is supported. Now, filtering messages gives the user ability to filer messages automatically and give restrictions to unwanted posts. Earlier OSN provides a little support to user walls for preventing spam messages. Lets we take e.g. various sites such as Google+, Facebook, Twitter allows the form a groups and those group members are allowed to post their messages on the user wall (i.e. friends, family, friends of friends, acquaintances and any predefined groups) But in existing system no content base filtering are there, so unwanted posts could not be filtered such as bad meanings of political statements, advertisements, vulgar messages.

Our goal for developing the filtering wall application is that prevent OSN users walls from unwanted messages, malicious spam, undesired events or consequences. So user should not have to worry about its account and wall for unwanted messages. We develop this software by using some robust terms for classification that are Binary TF, Normalization, and Cosine similarity and jacard algorithms.

2. SYSTEM MODEL

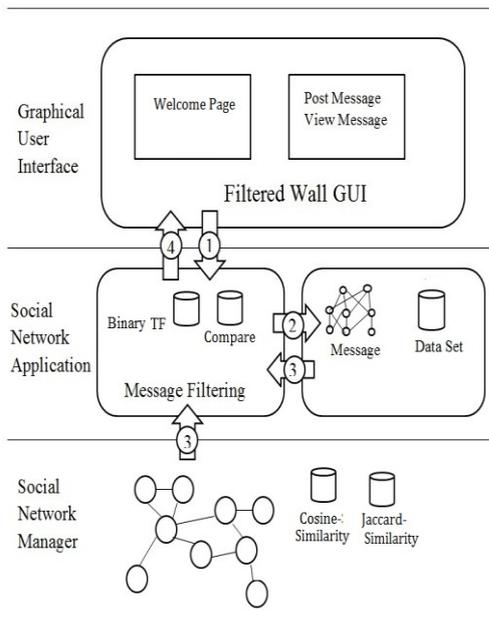


Figure 1. Conceptual System Architecture of Filtered Wall

The proposed system architecture is a three-tier architecture. Tier first is the OSN's Manager; manager manages the relationships and overall profiles of user. Tier second supports applications of other social networks. GUI is most important part for interacting with each other's over a network so third tier provides support of GUI to second tier. Our developed system is resides in second and third tier of the architecture. Again GUI helps us setup and managing TF and Normalisation is used to obtaining filtered messages. Similarity algorithms Cosine similarity and Jacard algorithms are helping us in message category identification to identify that the message is valid or invalid. When using our system GUI provides filtered wall to OSN user means such wall where the unwanted, malicious and spam messages are filtered and only those messages are published which are valid and legitimate messages according to our systems second and third tier rules.

The key elements of our system are TF-Term Frequency and Normalisation. Also Similarity algorithms Cosine similarity and Jacard [Explained in 4]

The flow of posted comment in our system fig 1.is as follows:

a) When user login into his/her own account and try to post a comment on wall. Then whatever the user posted that will be first intercepted by our filtering wall system.

b) For that comment dataset is extracted from our database.

c) Using TF and Normalisation filtering wall firstly start comparing the comment with our dataset and social graphs and profiles.

d) Also result of step (c) will be again tested with Similarity algorithms Cosine similarity and Jacard.

e) Now, step (d) will give the final result. If that the message/comment is spam the message will be removed and if valid then it will be posted on user wall.

3. PREVIOUS WORK

We designed software that provides content base filtering, which can be customized by user according to user needs. We have implemented some machine learning techniques here. We provides both content base filtering and policy base filtering hence we are showing literature survey for both as follows Email related filtration:-

In early years anti-spam system is proposed which based on the uncertain learning strategy. This strategy is integrated by commission collaboration mechanism. Newest developing systems are able to handle dual-way spam filtering, i.e. for both outgoing and incoming and spam. But the result of performance test of 6months on email server has very low filtering output.

Using GAD Clustering Online social network filtering:- All of we know that various online social networking sites are there such as Facebook, Twitter, YouTube, Flickr and LinkedIn. In all over the world people uses these social networking sites for to be in touch and share and spread information quickly and effectively. But many times these social networking sites are get spammed and hacked by malicious hackers and unauthorised users, hence to provide restrictions to posts and avoid spammers is the core importance in today's social networking sites. For those social networking sites online spam detection system using GAD clustering algorithm are available. GAD is a large scale clustering for improves the scalability and real-time detection challenges by integrating learning algorithms.

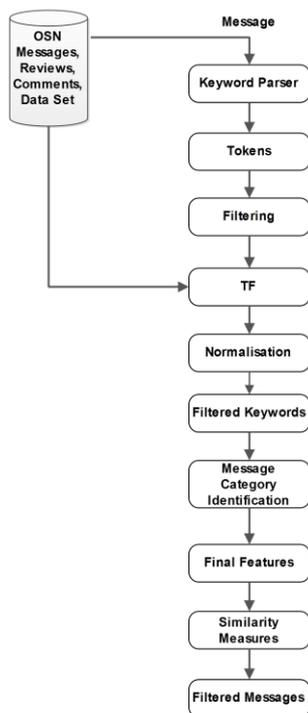


Figure 2. Flow graph of developed system.

When user needs to authenticating original message the SMTP protocol is used. In such cases the content-based filtering are used very often and the techniques such as Blacklists and Whitelists are simple filtering technique are very prevalent. So conclusion is when trying to retrieve the

behaviour of email sources then user has to deal with the various delivery errors and difficulties at that time. We got detail information and result of filtering approach through proposed novel spam filtering approach.

Also, on sites such as Facebook, twitter, Google+ is not supporting content based filtering. Now various sites are just gives us access of block/unblock users depends on which legitimate or not and we can just give access to post on our wall for selected users in our list means **who can post on my wall?** Such types of preventive access are present there on social networking sites. But the problem with that access are that if the person is blacklisted then these blacklisted person cant able to post again new message on user wall. But there should be possibility that if person post a spam message on time, next time he might be post a good or useful message, so removing/blocking a person completely is not a feasible solution. To overcome these drawbacks we developed a system which directly deals with the posts.

4.DEVELOPED METHODOLOGY

By referring the paper [7], we got the concept of filtering OSN walls and taking the concept of that paper we developed software which is Filtering Unwanted Messages from OSN User walls using MLT. The Flow graph of developed system is gives the actual flow of comments in our system. So stepwise explanation of our system is given here:-

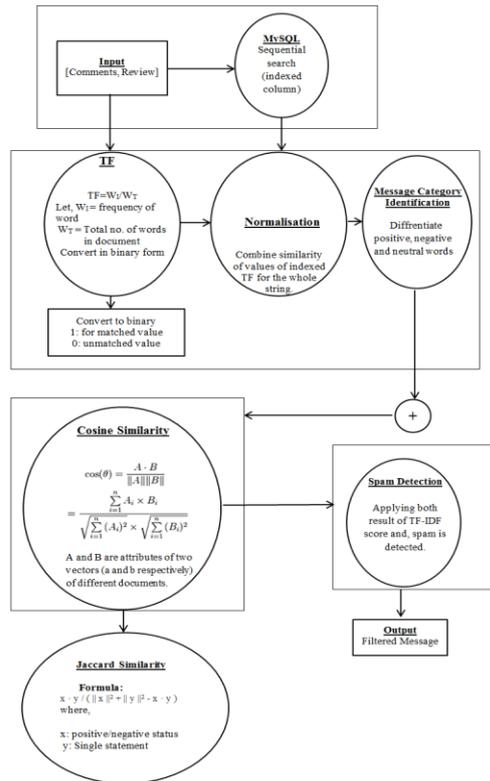


Figure 3. Mathematical Model.

Step 1: First Dataset is present in system which contains all comments and reviews posted on the user wall.

Step2: when user enters comment it will go to the keyword parser. The job of keyword parser is accepting the input string.

Step3: Result of Step2 is passed to Tokens block here, tokenization of statements will perform.

Step4: Filtering block is here, in that commonly occurring stop words (like is, and, the, are, was and so on) are compared with datasets data and filtered out. so that only the words which occurs less are remains in final dataset.

Step5 & 6: In this block the Binary TF and Normalisation is calculated for comment by comparing with dataset.

a) Binary TF means comparing the Words in Comment with the dataset values like as:

1. If the word is match with positive dataset then binary value of that word =1
2. If the word is match with negative dataset then binary value of that word =0

b) Normalisation performs the comparison of Binary TF values with the original Comment the posted by user. Depends on the value it will decide priority of that comment. And

1. If 1's are present more than 0's=Positive string

1. If 0's are present more than 1's=Negative string

Step7: In these Step of Filtered keyword depends on the value of step 5 & 6 we get keywords which are filtered

Step8: In this block we identify the category of messages which is positive or negative.

Step9: Final features means overall rating for that comment from step1 to step8 are added in these block.

Step10: Here, by using Machine Learning techniques similarity algorithm: 1.cosine similarity and 2.Jacard we major the comment is valid or not. If it is valid then it will goes to next step as a result of these step else the comment is restricted and not allowed to post.

a)We can calculate value of jacard algorithm using this Formula:

$$x*y/(x)+(y)-(x*y)$$

Where, x is: comment entered by user

And y: is the positiveNorm or negativeNorm resultant string by Normalisation.

b) We can calculate value of Cosine similarity algorithm using this Formula:

$$x * y / (x^{1/2} * y^{1/2})$$

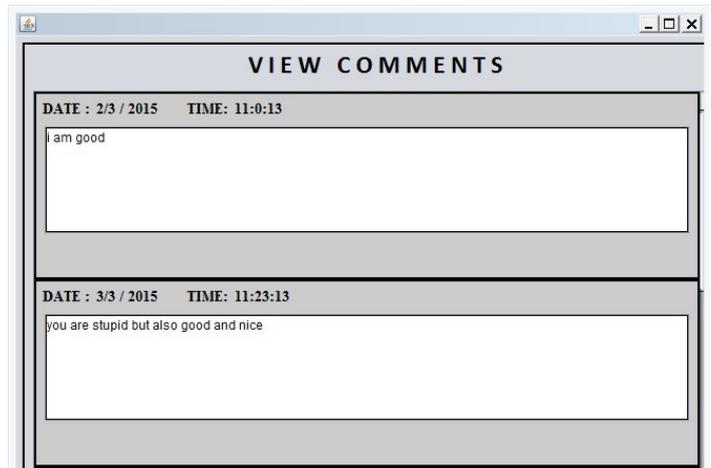
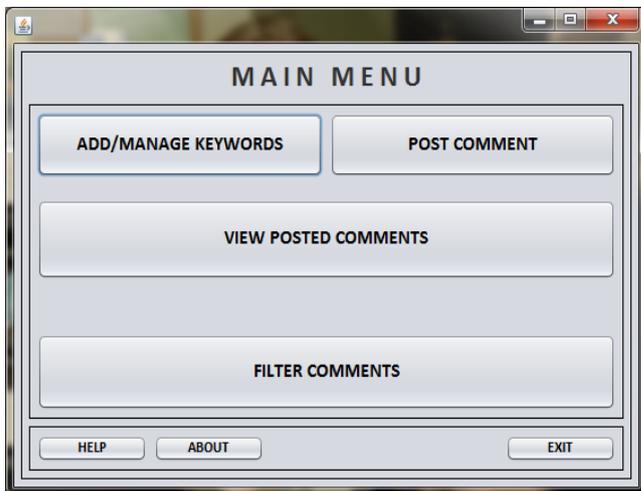
Where, x is: comment entered by user

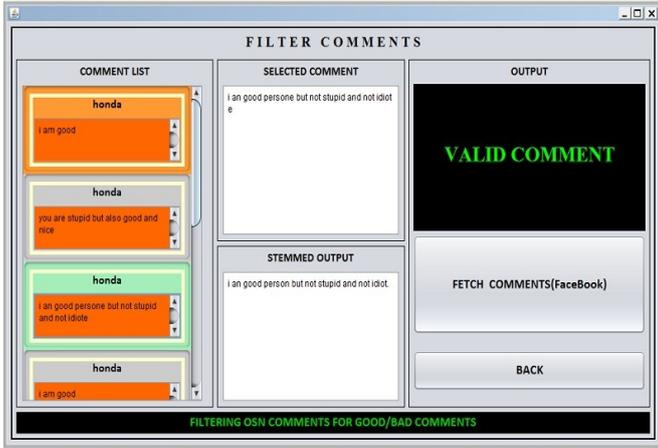
And y: is the positiveNorm or negativeNorm resultant string by Normalisation.

Step11: In these step is the comment is post on user wall successfully then it will be final filtered output i.e. only legitimate comment is posted.

So by referring all those algorithms and going through in that way we developed a system which will use to protect the user wall. And user will enjoy his/her activities on Social Networking Sites.

SNAPSHOTS:





5. CONCLUSION

We developed a system that gives restrictions to unwanted messages pasted on the user wall and filters unwanted and malicious messages from OSN wall. Intension of building such software is keeping user outside of any crime or problems which arises due to issues regarding the unwanted and malicious contents posted on their walls. We use developed a system based on the machine learning algorithms which enables a filtering of data. Systems also use to filter comments which are posted by user to other persons and fan pages wall. This developed system is the first step towards the huge scope of advanced projects in security parameter.

6. FUTURE SCOPES

Now, we developed system for OSN's like that we can develop and use this system for various sites, such as Wikipedia, Google, yahoo etc. Instead of completely we can block some users for specific time and also warn such user for his mistake. This help to block unwanted information from different users. In Today's situation our software is only work for filtering messages/comments in the form of a text, so in future we can extend our project scope to filter images, audio, video format o filtering. The main thing is that we have developed these software only for English language in future we can build it for many languages.

7. REFERENCES

[1] R. J. Mooney and L. Roy, "Content-based book recommending using learning for text categorization," in Proceedings of the Fifth

ACM Conference on Digital Libraries. New York: ACM Press, 2000, pp.195–204.

[2] F. Sebastian, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.

[3]Machine Learning-Mark K Cowan

[4]Towards online spam filtering in social networks, Hongyu GAO, Yan Chen, Kathy Leet

[5]<http://www.BinaryTF.com/>

[6]http://www.site.uottawa.ca/~diana/csi4107/cosine_tf_idf_example.pdf

[7] A System to Filter Unwanted Messages from OSN user walls M. Vanetti, E. Binaghi, B. Carminati, M. Carullo, and E. Ferrari, "Content-based filtering in online social networks" in Preceding of EMCL/PKDD workshop on privacy and security issues in data, Mining and machine learning,2010. Name of authors," Title of the research", Citation Details, year.