ISSN 2395-1621

Content-Based Classification Algorithms For Email Filtering

[#1]Tushar Bhagat, [#2]Ravikiran Tikale, [#3]Kiran Ugalmugale

[1]tusharbhagat313@gmail.com
[2]ravikiran.tikale@gmail.com
[3]onlyukboss@gmail.com

[#123]Department of Computer Engineering

JSPM's
Imperial College Of Engineering & Research,
Wagholi, Pune – 412207

ABSTRACT

Electronic mail is used daily by millions of people to communicate around the globe and is a mission-critical application for many businesses. Over the last decade, unsolicited bulk email has become a major problem for email users. An overwhelming amount of spam is flowing into users‟ mailboxes daily. Not only is spam frustrating for most email users, it strains the IT infrastructure of organizations and costs businesses billions of dollars in lost productivity. The necessity of effective spam filters increases. In this paper, we presented our study on various problems associated with spam and spam filtering methods, techniques. We also filter the upcoming mail from other user, if any unwanted mail coming then we automatically block that mail for security purpose.

ARTICLE INFO

Article History

Received: 24th November 2016

Received in revised form :

24th November 2016

Accepted: 27th November 2016

Published online :

27th November 2016

Keywords: **Spam mail, Spammers, Security, and Content Based Filtering.**

I. INTRODUCTION

The internet has become an integral part of everyday life and email has become a powerful tool for information exchange. Along with the growth of the Internet and e-mail, there has been a dramatic growth in spam in recent years. Spam can originate from any location across the globe where Internet access is available. Despite the development of anti-spam services and technologies, the number of spam messages continues to increase rapidly. In order to address the growing problem, each organization must analyze the tools available to determine how best to counter spam in its environment. Tools, such as the corporate e-mail system, e-mail filtering gateways, contracted anti-spam services, and end-user training, provide an important arsenal for any organization. However, users cannot avoid the very serious problem of attempting to deal with large amounts of spam on a regular basis. If there are no anti spam activities, spam will inundate network systems, kill employee productivity, steal bandwidth, and still be there tomorrow Social networking sites which are available now a days are most famous and easy way for communication, sharing a huge amount of information about the people. Generally daily and continuous communications implies sharing of the numerous types of materials which includes texts, images, audio clips and video clips too. According to recent survey the statistics of Facebook and Twitter users shows there average users sharing approximately 90-95% of data every month .The dynamic characters of data can create premises for employment of web content mining strategy which are aim automatically discover useful information within data itself. They are many technics provide active support in complex and sophisticate tasks involved in OSN (Online Scouting Network) management, such an instance can have Access Control, Information Filtering which has explored for concerns of text documents and web content. However, aim of majority of all these proposals is to provide users a classification mechanism to avoid they are over useless data. In OSN, Information Filtering can use for various sensitive purposes. This is fact that OSN is possibility of posting or commenting others posts on particular public and private areas called as a general wall. The Information filtering can also give users ability to work automatically control messages written on walls by applying filtering unwanted messages. We believe that a key that has not been provided so far. The aim is to represent work of proposed

experimentally evaluating an automated system called as Filter Wall that will be able to filtering unwanted messages from OSN users wall. We have exploited Machine Learning text categorization technique which can assign automatically a short text messages that is set of categories on basis of their contents. The major effort is to building robust Short Text Classifiers that is concentrated in extraction and the selection that set is characterizing and discrete features.

II. RELATED WORK

In Content-Based filtering, user is assumed to operate independently. As a result Content-Based Filtering system select information items base on correlation between content of items and user can also prefer to oppose collaborative filtering system that select items based on correlations between peoples with similar preference. While electronic mail was original domain of early work on information filtering process. Document processed in content-based filtering is mostly textual in nature and can makes content-based filtering close to text classification. This activity of filtering can modelled, as case of single labelled, binary classification and partition of incoming document in the relevant and non-relevant categories.

The various spam filtering techniques:

Rule based filtering:

Evaluate a large number of patterns-mostly regular expressions--against a candidate message. Some matched patterns add to a message's score, while others subtract from it.

Bayesian classifier:

Particular words have particular probabilities of occurring in spam email and in legitimate email. The filter doesn't know these probabilities in advance, and must first be trained so it can build them up.

K nearest neighbors:

If at least t messages in k neighbors of the message m are unsolicited, then m is unsolicited email, otherwise, it is legitimate.

Content based Spam Filtering Techniques:

The neural networks are quite famous to be well adapted for problems of classification. Without being spread out over the model, we will retain in what follows the characteristics which contribute to the design of an anti-spam filter.

III.PROPOSED SYSTEM

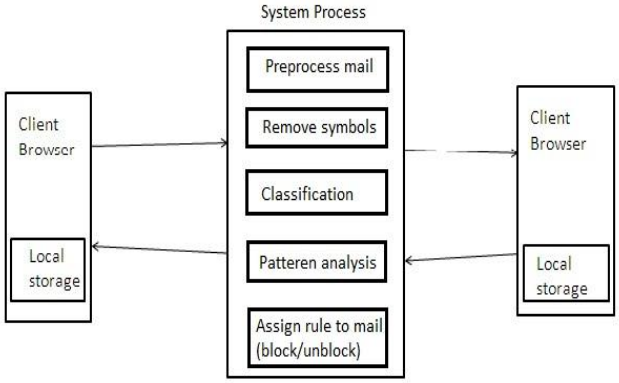

Fig 1: System Architecture

Module:

Client:
Client menace number of user cans browse the emails. He can perform different activity like manage account forward emails etc.

Server:
Server can perform interface between client browser and database. Server is strong parameter in our project, if server is slow down then the forwarding or receiving emails also slow down.

Database:
Database can stored the all reporting data; user can perform in sending or receiving all types of emails.

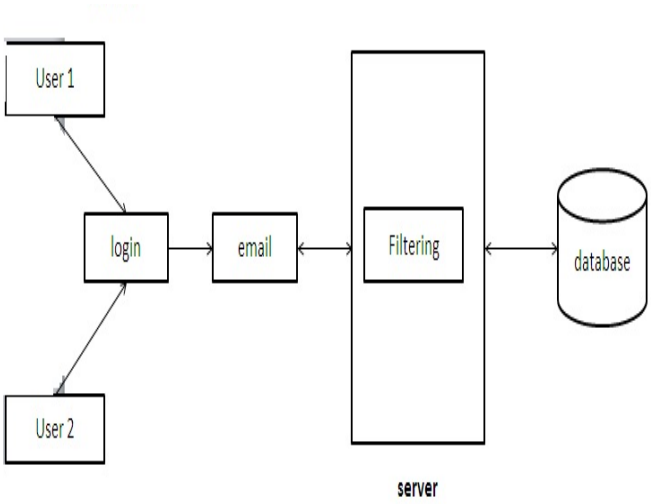

Fig 2: Functionality working

IV. CONCLUSION

In this paper we have basic study of different email filtering technique for classifies on the basis of the perfection, accuracy performance of the algorithms. The better approach of this paper will be the additional features added for classifying the ham or spam mails using advanced email Filtering algorithms.

REFERENCE

[1] Miszalska, I., Zabierowski, W., & Napieralski, A. (2007, February). "Selected Methods of Spam Filtering in Email." In CAD Systems in Microelectronics, 2007. CADSM'07. 9th International Conference-The Experience of Designing and Applications of (pp. 507-513). IEEE.

[2] Scholar, M. (2010). "Supervised learning approach for spam classification analysis using data mining tools." organization, 2(08), 2760-2766.

[3] Youn, S., & McLeod, D. (2007). "A comparative study for email classification." In Advances and Innovations in Systems, Computing Sciences and Software Engineering (pp. 387-391). Springer Netherlands.

[4] Xiao-li, C., Pei-yu, L., Zhen-fang, Z., & Ye, Q. (2009, August). "A method of spam filtering based on weighted support vector machines." In IT in Medicine & Education, 2009. ITIME'09. IEEE International Symposium on (Vol. 1, pp. 947-950). IEEE.

[5] Sculley, D., & Wachman, G. M. (2007, July). "Relaxed online SVMs for spam filtering." In Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval (pp. 415-422).

[6] Chan, T. Y., Ji, J., & Zhao, Q. "Learning to Detect Spam: Naive-Euclidean Approach." International Journal of Signal Processing, 1.

[7] Puniškis, D., Laurutis, R., & Dirmeikis, R. (2006). "An artificial neural nets for spam e–mail recognition." Elektronika ir Elektrotechnika (Electronics and Electrical Engineering), 5(69), 73-76.

[8] Drucker, H., Wu, D., & Vapnik, V. N. (1999). "Support vector machines for spam categorization." Neural Networks, IEEE Transactions on, 10(5), 1048-1054.

[9] Provost, J. (1999). "Naıve-Bayes vs. Rule-Learning in Classification of Email." University of Texas at Austin.

[10] Medlock, B. (2006, July). "An Adaptve, Semi-Structured Language Model Approach to Spam Filtering on a New Corpus."

