Bayesian Decision Framework for an Efficient Spam Filtering In Social Network

T.Priyanka M.sc.,M.B.A.[1]
[1] *Mphil.Scholar, Department of Computer Science, KG College of Arts and Science,Coimbatore. Tamil nadu, India.*

Abstract— Internet email is one of the most popular communication methods in business and personal lives. However, spam is causing a major problem in email systems. The wide growth of unwanted emails has prompted the growth of numerous spam filter techniques. Many Filtering techniques had been used for identifying spam emails, treats spam filtering as a binary classification problem. i.e. the in-coming email is either spam or non-spam. Current work proposes three-way decision approach to Social network Aided Personalized and effective spam filter (SOAP) based on Bayesian decision theory. Three-way decision approach based on Bayesian decision theory is introduced to SOAP for classification of spam details. In each node, components such as social interest-based spam filtering, adaptive trust management and social closeness-based spam filtering is integrated in the Bayesian filter to classify the spam and non spam details. The key advantage of the proposed is that the unresolved cases must be re-examined by collecting additional information from the components of the node. Experimental results shows that the current approach minimizes the error rate of classifying a legitimate email to spam, and offers better spam weighted and precision accuracy.

Keywords— Social network , SOAP, Bayesian theory, three-way decision, Spam filter

I. INTRODUCTION

In our business and personal lives, Internet email is considered as one of the most popular communication methods. Yet, spam is becoming a major problem in email systems. At present, 120 billion spam emails are sent per day, with a projected cost of $338 billion by 2013. These kinds of emails mess with both end users and email service providers. A primary method to prevent spam is to make it unbeneficial to send spam emails, thus destroying the spammers' fundamental business model [1].

Spam filter should rely on attack-resilient, personalized, and user-friendly. *In order to achieve high accuracy,* the *personalized and* attack-*resilient* features are important. A more accurate filter generates less *false positives* and *false negatives*. Generally, *false positives* are legal emails that are incorrectly regarded as spam emails. And then *false negatives* are spam emails that are not actually detected. There are two main types of spam filter attacks as follows:

- Poison attacks and
- Impersonation attacks.

In a poison attack, a lot of legitimate words are stayed to spam emails, consequently decreasing its probability of being detected as spam. In an impersonation attack, a spammer impersonates the identities of ordinary users by forging their IDs or compromising their computers.

Personalization means that a correct spam filter must consider the social Previous spam filtering methods can be largely divided into two categories: identity-based and content-based.

In the identity-based category, spam filters spots spam based on the identities of senders of emails. Generally, a user preserves a white list and a blacklist for email addresses [2]. Social interaction-based spam filters [3] develop friends of friend (FoF) associations between email reporters to produce white lists and blacklists. As these spam filters do not regard as email content, they are resilient to poison attacks.

In the content-based category, emails are parsed and scored based on patterns and keywords that are characteristic in spam. Machine learning approaches [4] (including the Bayesian filter) train spam filters with a combination of both legitimate emails and spam which identify their type, which are used to repeatedly classify future emails into two classes. Still, these approaches deal several problems. First, the spam filters are usually installed in an email server to bring together all the training samples; therefore, they are not personalized so as to increase the efficiency and accuracy of training. Second, the spam filters are susceptible to poison attacks. Third, the spam filters are not considered as user friendly; they need a large extent of users' effort to manually differentiate spam from legitimate emails for training.

In the current research, a three-way decision approach has been introduced to spam filtering based on Bayesian decision theory, namely, to *accept*, *reject*, or *further-exam* an incoming email which is being received. The emails waiting for *further-exam* have to be clarified by gathering extra information. For example, misclassifying a legitimate email to spam is frequently measured more expensive than misclassifying a spam email to legitimate. In the present wok, based on naive Bayesian classification, the conditional probability is interpreted. The major advantage of three-way decision making is that it permits the possibility of rejection, which is refusing to make a decision. The undecided belongings must be considered for re-examination. A loss function is described to suit how expensive each action is, and the concluding decision is to pick the action for which the overall cost is minimum.

II. RELATED WORK

In [5] Boykin *et al* presented spam filtering by considering a graph in which vertices symbolize email addresses and direct edges symbolize email interactions. Spam Emails are identified as valid or unknown based on the general clustering coefficient of the subcomponent of graph. This is rooted in the rationale that the normal nodes social communication network has a greater clustering coefficient than that of a spam node.

In [6] Hameed et al presented LENS, which explores the FoF network by accumulating trusted users from exterior of the FoF networks to alleviate spam away from social circles. Only emails to a recipient that have been acknowledged by the trusted nodes can be sent into the network.

In [7] DeBarr *et al. presented a spam filter by* evaluating the use of social network analysis process to advance the performance of a content filtering model. They defined to detect spam by estimating the degree centrality of message transmit agents and the average path length among senders and receivers. They maintained that the messages from a immoral mail spread or messages with abnormal path lengths that diverge from the average are more probable to be spam.

In [8] Lam *et al* presented a learning approach uses user interaction features for spam sender detection which is extracted from social networks built from email replace logs. Legitimacy scores are owned to senders based on their probability of being a legitimate sender.

In [9] Tran *et al* presented an email client called Social Email, which offers social circumstance to messages by means of a social network's fundamental social graph. This not only provides each email beneficiary control over who can message him/her, but also offers the recipient with a consideration of where the message socially initiated from. On the other hand, if a spammer negotiates a legitimate user's computer, the spammer can simply attack the user's associates in the social network, which is considered by high clustering and short paths. In addition, such social interaction-based methods are not adequately efficient in compacting with legitimate emails from sender's exterior of the social network of the receiver.

In [10] James et al presented an email scoring method based on an email network improved with reputation ratings. An email is measured spam if the reputation scores of the email sender are very low. Unlike social network based methods, SOAP meets on personal interests in conjunction with social relationship nearness for spam detection.

III. PROPOSED WORK

In this work, three-way decision approach is proposed with Social network Aided Personalized and effective spam filter (SOAP) for detecting spam to obtain the three requirements. SOAP includes a social network which contains the social relationships and user (dis)interests into the email network. In this spam filter, users register their emails in client of SOAP and are optimistic to supply their social information, for example religions, occupations, and social relationships, to avoid spam. Each node joins to its social friends in the nearby stored friendlist. SOAP has been used by each node in a network to collect information and check spam autonomously. Through the collected information, spam filtering can be done by using Bayesian decision framework that is to accept, reject, or further-exam an incoming email. The emails of further-exam must be simplified by gathering further information. The major advantage of three-way decision making is that it permits the likelihood of rejection, i.e., of refusing to take a decision. The undecided subjects must be sent for re-examination. Experimental results provides better result when compare with existing work.The system architecture of the proposed framework is shown in fig 1:

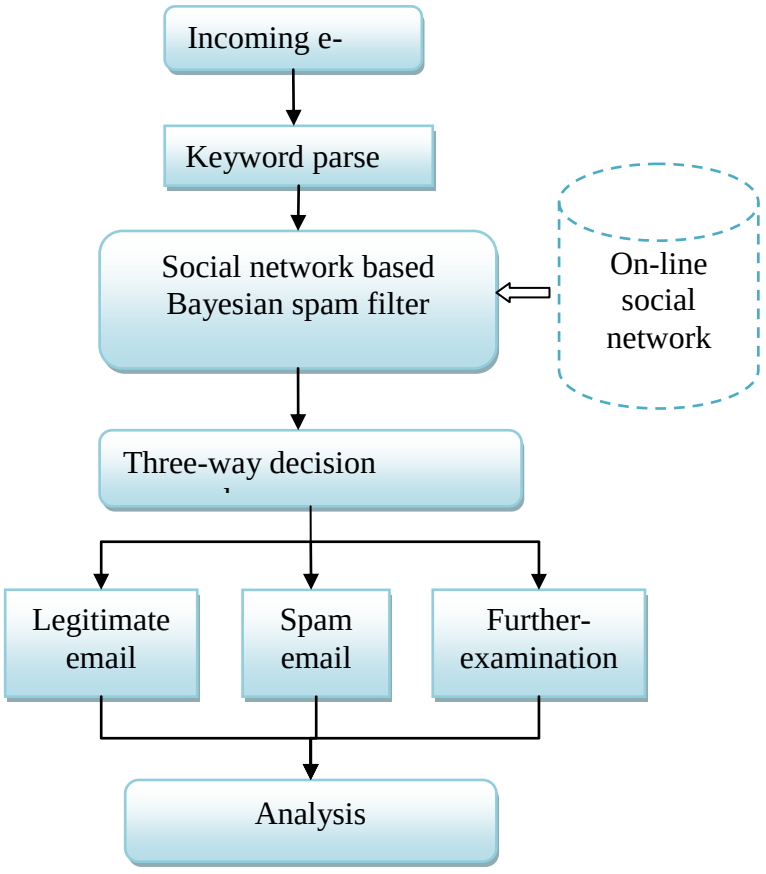

Figure 1: Proposed system achitecture

IV. METHODOLOGY

A. SOCIAL NETWORK AIDED PERSONALIZED AND EFFECTIVE SPAM FILTER (SOAP)

The proposed system considers spam filtering based on following three requirements:

- Social interest-based spam filtering
- Adaptive trust management and
- Social closeness-based spam filtering

1) *Social interest-based spam filtering:* The social interest-based spam filtering component goal is to create SOAP to increase the spam detection accuracy. SOAP increases and decreases the probability of the e-mail keywords to be spam by matching the keywords in an email with the email receiver's social interests and disinterests respectively. In order to infer each user's dis-interests, SOAP uses a rule-based inference system. This inference system has three parameters: profiles, inference engines and inference rules. Parameter of *profile* is defined as a database holding all helpful facts parsed from the user' profile in the social network holding interests, occupations etc. The *inference engine parameter* establishes the rules in the circumstance of the current profile, and picks the most suitable rules for the inference. These facts are built into a fact database. The *inference rules parameter* encloses all the rules that are used for the inference of dis-interests. Those rules can be rational reasoning based on non-monotonic logic.

User dis-interests are derived from user profiles in the social network; the accuracy of spam filtering depends on user interests could be affected if some users do not give accurate or complete profiles.

SOAP alters the weights of keywords in an email consistent with social closeness involving the email receiver and sender. Subsequently, SOAP adjusts the weights along with the email receiver's social dis-interests.

For a spam keyword within the email receiver's interests, its weight is adjusted by:

$P(S|w_{interest}) := P(S|w_{interest}).e^{-\rho I}$

where $w_{interest}$ is the spam keyword in interests and ρI is a scale parameter.

When a spam keyword matches the disinterests of the email receiver, the weight of the keyword is adjusted by

$P(S|w_{disinterest}) := P(S|w_{disinterest}).e^{-\rho D}$

Where $w_{disinterest}$ is the spam keyword in the email receiver's disinterests ρD is a scale parameter..

2) *Adaptive trust management* :Generally spammer masquerade the identities of benign computers by forging their IDs or satisfying them to send spam or describe fake relationships with neighbors to the source node at some stage in the social closeness calculation process. By reason of the small-world characteristics of social networks, that is nodes are greatly clustered, impersonation can extend spam enormously fast. With the purpose of combat impersonation attacks, SOAP incorporates an adaptive trust management component. Purposely, a node follows fast behavior alternations of close-relationship nodes. It utilizes the additive-increase or multiplicative-decrease algorithm to adjust node trust. To update node closeness Node trust is then used for the detection of false negatives caused by impersonation attacks.

The goal is to provide a balance among acceptance of trustable emails and responsiveness to false negatives. Let us assume that, node i originally guess node j with high closeness is truthful until it receives a spam email from node j.

Then trust value of node j regarded by node I is denoted as t(i,j). The maximum trust value $t_{max} = 1$ and $t \leqslant t_{max}$. t is originally set to t_{max}. Once node i receives a spam email from sender j, node i changes the trust value of j by

$t(i,j) := a.\ t(i,j)\ (0 < a < 1)$

Once a node i obtain a legitimate email from sender j, then

$t(i,j) := t(i,j) + b\ (0 < b < 1)$

Thus, SOAP can sensitively regulate node trust value to rapidly react to zombies, thus dropping false negatives.

3) *Social closeness-based spam filtering:*When a person receives an email from other socially related close person, the email has a low probability of being spam except the email sender's machine is in an impersonation attack. Consequently, the social closeness between individuals can be used to get better the accuracy of spam detection.

In this section, a spam filtering method is used in spam detection to experimentally measure social closeness between two persons. Spam filtering relies on nodes' social relationships, namely friendship and kinship, to decide node closeness values. This filtering method sets dissimilar weights for different social relationships to measure node closeness. For instance, the closeness of a kinship relationship generally weights more than a business relationship. The weight of a relationship between node u and v is denoted by c(u,v). In the following section the discussion has been done on how to calculate the closeness of adjacent nodes and non-adjacent nodes in a social network.

Node Closeness:

In a social network, additional relationships between two adjacent nodes compose them closer. Consequently

$C(u,v) = \sum_{i=1}^{n} c_i(u,v)$

where n is the number of relationships between nodes u and v, and c_i (u,v) is the relationship weight of the i^{th} relationship. The closeness of non-adjacent nodes can be estimated with the support of relationship transitivity, in which relationship

closeness can be approved along the nodes which depends upon the closeness value among any two adjacent nodes. In case, if node *A* is B's father and *E* is B's best friend, then *A* is unlikely to send spam to *E*. The closeness transitivity must detain three properties so as to correctly reflect the social relationship.

B. THREE-WAY DECISION APPROACH BASED ON BAYESIAN DECISION THEORY

Another In this section, the details obtained from the above SOAP methods are analysed by means of Three-way decision approach based on Bayesian decision theory. In this, an incoming email is classified as legitimate if the subsequent odds ratio goes beyond a definite threshold value. A pair of threshold values is employed to create a three-way decision of an incoming email. The first threshold value estimates the probability required for a re-examination, and the second value estimates the probability essential to reject an email.

With regards, a set of emails to be estimated which considers a set of two states $\Omega = \{c, c^c\}$representing that an email is in C is considered as legitimate or not in C in considered as spam respectively. The incoming emails is classified into regions three different regions namely, the positive region, Boundary region and negative region. Positive region is deoted as POS(C) contains emails being legitimate, the boundary region is denoted as BND(C) contains emails that need further-exam, and the negative region is denoted as NEG(C) contains emails that are spam.

The loss function of these regions is denoted as follows.

	C(P)	c^c(N)
α_P	$\lambda_{PP}=\lambda(\alpha_P\|C)$	$\lambda_{PN}=\lambda(\alpha_P\|c^c)$
α_B	$\lambda_{BP}=\lambda(\alpha_B\|C)$	$\lambda_{BN}=\lambda(\alpha_B\|c^c)$
α_N	$\lambda_{NP}=\lambda(\alpha_N\|C)$	$\lambda_{NN}=\lambda(\alpha_N\|c^c)$

Where, λ_{PP} , λ_{BP} *and* λ_{NP} indicates the losses acquired for selecting actions α_P, α_B and α_N respectively, when an email belongs to *C*, and λ_{PN}, λ_{BN} *and* λ_{NN} indicates the losses obtained for selecting these actions when the email does not belong to *C*.

The normal losses connected with selecting different actions for emails with description x can be expressed as:

$R(\alpha_P|x) = \lambda_{PP}\ Pr(C|x) + \lambda_{PN}\ Pr(c^c|x)$

$R(\alpha_B|x) = \lambda_{BP}\ Pr(C|x) + \lambda_{BN}\ Pr(c^c|x)$

$R(\alpha_N|x) = \lambda_{NP}\ Pr(C|x) + \lambda_{NN}\ Pr(c^c|x)$

The minimum –risk detection of the above rules can be suggested by Bayesian decision theory is as follows:

(P) If $R(\alpha_P|x) \le R(\alpha_B|x)$ and $R(\alpha_P|x) \le R(\alpha_N|x)$, decide x∈ POS(C)

(B) If $R(\alpha_B|x) \le R(\alpha_P|x)$ and $R(\alpha_B|x) \le R(\alpha_N|x)$, decide x∈ BND(C)

(C) If $R(\alpha_N|x) \le R(\alpha_P|x)$ and $R(\alpha_N|x) \le R(\alpha_B|x)$, decide x∈ NEG(C) .

V. EXPERIMENTAL RESULTS

In this section the performance measure of the proposed system is evaluated with the existing spam filtering technique.

A social network is built based on data crawled from Facebook. Two users with no social relationship in Clemson University from social network as seed nodes and constructed a friend graph using the breadth first search through each node's friend list. The user whose personal information cannot be obtained is eliminated. At last, a connected social network with 32344 users was recognized for proposed system. Personal information such as interest and religion of each node was viewed and stored in the node.

The proposed system uses the performance measures of false positive, false negative rate and detection accuracy for detecting the spam filtering efficiency. The *false positive rate* (*FP*) is defined as the proportion of negatives cases that were incorrectly classified as positive. *False negative rate* (*FN*) is defined as the proportion of positives cases that were incorrectly classified as negative. The accuracy rate is defined as the ratio between the number of successfully classified emails and the number of all received emails.

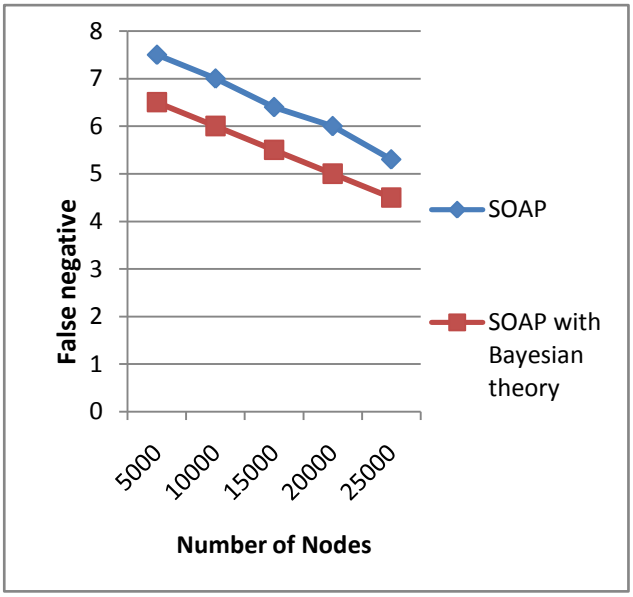

Figure 2: False negative comparision graph

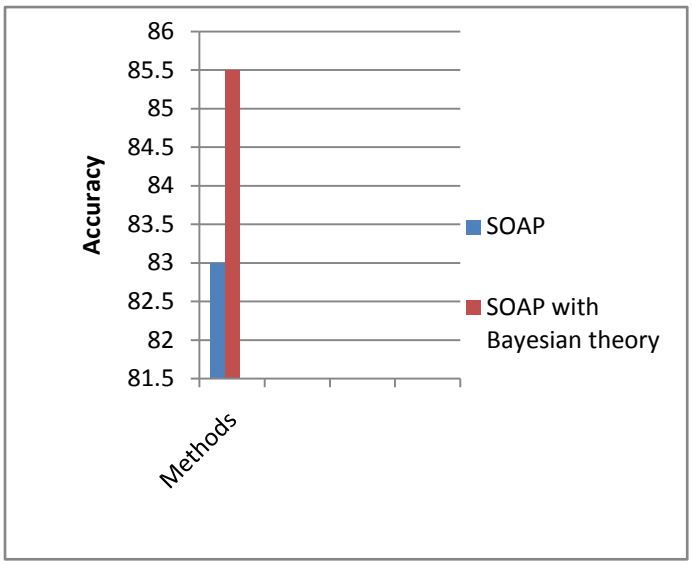

Figure 3: False positive comparision graph

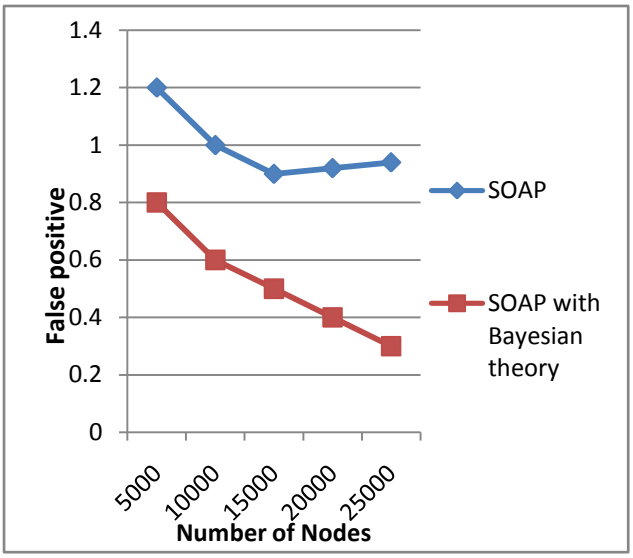

Figure 4: Accuracy comparision graph

Thus the above graph in figure 2, 3 and 4 shows that proposed system of SOAP with basyesian theory lesser False positive and false negative rate nad provides higher accuracy when compared with existing method of SOAP.

VI. CONCLUSION

In the present work, Three-way decision approach based on Bayesian decision theory is introduced to SOAP for classification of spam details. Social network Aided Personalized and effective spam filter (SOAP) is proposed to meet the requirements of different spam's obtained from three components such as social interest-based spam filtering, adaptive trust management and social closeness-based spam filtering. These requirements are included in the proposed Three-way decision approach based on Bayesian decision theory to identify the incoming mail as spam or not. Unlike traditional spam filtering approach, proposed method adds a third action to allow users make further examinations for undecided situations. The advantage of the present work is that it provides a more reasonable feedback to users for using their emails, thus minimizes the misclassification rate. A pair of threshold values is used is used for estimated. The first threshold value estimates the idea necessary for a re-examination, and the second value estimates the idea to reject an email. Experimental result provides better result in real world dataset when compare with the existing spam filtering approaches.

.

REFERENCES

[1] P. O. Boykin and V. Roychowdhury. Personal Email Networks: An Effective Anti-Spam Tool. IEEE Computer, 2004

[2] DNS Real-time Black List. http://www.dnsrbl.com/index.html.

[3] O. Boykin and V. Roychowdhury. Personal Email Networks: An Effective Anti-spam Tool. *IEEE Computer*, 2004.

[4] M. Uemura and T. Tabata. Design and Evaluation of a Bayesian filter- based Image Spam Filtering Method. In *Proc. of ISA*, 2008

[5] P. Oscar Boykin and Vwani P. Roychowdhury. Leveraging Social Networks to Fight Spam. *IEEE Computer*, 2005.

[6] S. Hameed, X. Fu, P. Hui, and N. Sastry. LENS: Leveraging social networking and trust to prevent spam transmission. In *Proc. Of FIST*, 2011.

[7] D. DeBarr and H. Wechsler. Using Social Network Analysis for Spam Detection. In *Proc. of SBP*, 2010

[8] H. Lam and D. Yeung. A Learning Approach to Spam Detection based on Social Networks. In *Proc. of CEAS*, 2007.

[9] T. Tran, J. Rowe, and S. F. Wu. Social Email: A Framework and Application for More Socially-Aware Communications. In *Proc. Of SocInfo*, 2010

[10] J. James and J. Hendler. Reputation Network Analysis for Email Filtering. In *Proc. of CEAS*, 2004