International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-6 Issue-5, June 2017

An Improved Classifier Technique for Spam Filtering

Rahul Maheshwari, Vivek Kapoor, Sandeep Verma

Abstract: -Email spam or junk e-mail (unwanted e-mail "usually of a commercial nature sent out in bulk") is one of the major problems of the today's Internet, bringing financial damage to companies and annoying individual users. There are various approaches developed to stop spam, filtering is an important and popular one. Spam or unsolicited e-mail has become a major problem for companies and private users. This paper explores the problems associated with spam and some different approaches attempting to deal with it. Since spam is a major issue for web world thus the most appealing methods are those that are easy to maintain and prove to have a satisfactory performance. A learning algorithm which uses the Naive Bayesian classifier has shown promising results in separating spam from legitimate mail. There are various initial steps involved in spam classifier like Tokenization, probability estimation and feature selection are processes performed prior to classification and all have a significant influence upon the performance of spam filtering. The main objective of this work is to examine and empirically test the currently known techniques used for each of these processes and to investigate the possibilities for improving the Bayesian classifier performance. There are many different approaches available at present attempting to solve the spam issue. One of the most promising methods for filtering spam with regards to performance and ease of implementation is that of Naive Bayesian classifier. The objective of this paper is to explore the statistical filter called Naive Bayesian classifier and to investigate the possibilities for improving its performance.

Keywords -E-mail classification, Spam, Spam filtering, Machine learning, algorithms.

I. INRODUCTION

Email was widely accepted by the Modern business community as the first broad electronic communication medium and was the first 'e-revolution' in business communication. Email is very simple and easy to understand, email solves two basic problems of communication: logistics and synchronization. Spam is flooding the Internet with many copies of the same message; in an attempt to force the message on people who would not otherwise choose to receive it.Spam is also known as unsolicited junk mail. Most spam is commercial advertising, often for dubious products, get-rich-quick schemes, or quasi-legal services.

Manuscript published on 30 June 2017.
***** Correspondence Author (s)
Rahul Maheshwari, Department of Computer Science and Engineering, Institute of Engineering and Technology, DAVV, Indore (Madhya Pradesh)-452017, India. E-mail: rahul.maheshwari024@gmail.com
Dr. Vivek Kapoor, Department of Information Technology, Institute of Engineering and Technology, DAVV, Indore (Madhya Pradesh)-452017, India. E-mail: vkapoor13@yahoo.com
Sandeep Verma, Department of Information Technology, Institute of Engineering and Technology, DAVV, Indore (Madhya Pradesh)-452017, India. E-mail: vermasandeep42@gmail.com

Spam costs the sender very little to send most of the costs are paid for by the recipient or the carriers rather than by the sender. There is an immediate need to control the steadily growing spam flood. A great deal of on-going research is trying to resolve the problem.

1.1. Types of an E-Mail

An electronic message is "spam" IF: (1) the recipient's personal identity and context are irrelevant because the message is equally applicable to many other potential recipients; AND (2) the recipient has not verifiably granted deliberate, explicit, and still-revocable permission for it to be sent; AND (3) the transmission and reception of the message appears to the recipient to give a disproportionate benefit to the sender.

1.1.1 Marketing Emails

Marketing (or Bulk) emails stimulate your clients and leads. They contain informative / incentive messages. The recipient must agree to receive such emails: opt-in is mandatory. However, the recipient does not make an explicit request for a message in particular. For example: he doesn't subscribe for the "November Newsletter", he rather subscribes to the "Monthly Newsletter". Common examples of marketing emails: Newsletters Flash sales, Sales/promotions announcements.

1.1.2 Notification Emails

Notification email is also known as trigger, alert or auto-responder. This type of mail allows the user to be notified. More generally, this category of email may be used in order to celebrate and/or mark an event. From a marketer's point of view, it can be relevant to encourage the targets to opt in to receive notifications about the services being offered. Think of an email such as "Mr. X is now following you on Twitter". This kind of message is more often opened and it motivates the recipient into checking their account. Common examples of *notification emails*: Getting in touch a few days after registration, Congratulations after a status change (first purchase, subscription...), Birthday email.

1.1.3 Transactional Emails

This is an expected message and its content is information that the client wishes to check or confirm, and not "discover". Transactional email is not intended to optimize the customer relationship but to define it and mark it out. Common examples of **transactional emails**: Welcome message / Account opening, Shipment tracking and order status, Order shipment confirmation.

1.2. Why do people send spam?

Spam is the electronic equivalent of junk mail. The low cost of electronic communications has both benefits and drawbacks. Due to the low cost People send Spam in order to sell products and services or to promote an email scam in very less time span. Some Spam is purely ideological, sent by purveyors of thought.
The bulk of Spam is intended, however, to draw traffic to web sites or to sell sex and money making schemes.

1.3. The Costs of Spam

Many people get annoyed at spam but they don't realize that it's also actually costing them money. We all have to waste a considerable amount of time manually shifting through our mailboxes to sort out what is genuine and what is spam. This time could be spent better by doing something else. Businesses lose billions of dollars as a result of spam. The biggest cost caused by spam will be paid by the ISP's who have to pay increased bandwidth charges as a result of increase network traffic. However, spam also causes problems for many other people because of increased fraud, wasted time, and various other scams. There are several Negative effects of spam email-
1. Decline in Productivity
2. Wasted Storage Space
3. Financial Costs for the Internet Service Provider

II. LITERATURE AND SURVEY

There are several approaches which deal with spam. This section briefly summarizes and describes the spam filtering techniques. Rule based filters are a popular content-based method which applies a set of rules to every incoming email. If there is a match, the e-mail is assigned a score that indicates spaminess or non-spaminess. If the total score exceeds a threshold the e-mail is classified as spam. The rules are generally built up by regular expressions and they come with the software. The rule set must be updated regularly as spam changes, in order for the filtering of spam to be successful. Updates are retrieved via the Internet. The tests results from the comparison of anti-spam programs presented in Holden 2003 show that Spam Assasin finds about 80% of all spam, while statistical filters find close to 99% of all spam.
In Sahami et al. 1998, it is shown that it is possible to achieve remarkable results by using a statistical spam classifier. Since then many statistical filters have appeared. The reason for this is simple; they are easy to implement, have a very good performance and require a little maintenance. Statistical filters require training on both spam and non-spam messages and will gradually become more efficient. They are trained personally on the legitimate and spam e-mails of the user. Hence it is very hard for a spammer to deceive the filter.

In recent years, Spam is beginning to diminish the reliability of e-mail (Hoanca, 2006). Nowadays, Spam filtering is usually tackled by machine learning (ML) algorithms aimed at discriminating between legitimate and Spam messages, providing an auto-mated, adaptive approach, which are the focus of this review. Instead of relying on hand-coded rules, which are prone to the constantly changing nature of Spam messages, ML approaches are capable of extracting knowledge from a set of messages supplied, and using the obtained information in the classification of newly received messages. Wang and Cloete (2005) surveyed some approaches for e-mail classification, including Spam filtering and e-mail categorization. A relatively recent overview of approaches aimed at Spam filtering was presented by Carpinter and Hunt (2006), which focused on more general aspects of the problem. A more recent re-view has been conducted by Blanzieri and Bryl (2008). However, it did not discuss several of the more recent works, such as Case- Based Reasoning models and Artificial Immune Systems. The bias imposed by the commonly used bag-of-words representation and an important difference between naive Bayes models. We also discuss the need to evaluate a filter in a realistic setting, according to some recent corpora available. Emphasis is given to recent works, minimizing the overlap with other reviews (Blanzieri & Bryl, 2008; Carpinter & Hunt, 2006; Wang & Cloete, 2005), although some early works proposing the use of some approaches are also discussed to outline the evolution of their use.

The application of the naive Bayes classifier to Spam filtering was initially proposed by Sahami, Dumais, Heckerman, and Horvitz (1998), who considered the problem in a decision theoretic frame- work given the confidence in the classification of a message. The work of Sahami et al. led to the development and application of other machine learning algorithms to Spam filtering, and, in conjunction with the suggestions later proposed by Graham (2002), Although Graham's work is usually cited as in the same line as that of Sahami et al., an important difference between these two approaches, termed hereafter Graham and Sahami models, respectively. The Sahami model is a straightforward application of the naive Bayes classifier to Spam filtering. It usually requires the application of a feature selection algorithm during training to select the most relevant features,

In contrast, Graham's model does not require the application of a feature selection algorithm during training, when only the terms are retained. Sahami's model performs off-line feature selection, during training, while Graham's model selects the most relevant features online; Graham's model tends to per- form much better in an online fashion.

Since the work of Sahami et al. (1998), various studies focusing on naive Bayes models were conducted. Medlock (2006) proposed ILM (Interpolated Language Model), which considers the structure of an e-mail message, namely the subject and body, in a smoothed word n-gram Bayesian model. Segal, Markowitz, and Arnold (2006) proposed an approximation of Uncertainty Sampling (US), termed Approximated US (AUS), for building a classifier given a pool of unlabelled messages. Instead of considering words, Kim, Chung, and Choi (2007) focused on the URLs (links) in messages, using a naive Bayes model. The filter was periodically updated with the messages that were classified and not fed back until a certain time, besides those that were incorrectly classified.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Ciltik and Gungor (2008) applied a naive Bayes classifier based on word n-grams, using only some of the first words to reduce the classification time. Two classification models were considered: binary (Spam or legitimate) and instance-based (each message represented as a single class). The latter tends to show higher computational cost.

Here, we present a comprehensive review of recent developments in the application of machine learning algorithms to Spam filtering. a major difference between two early naïve Bayes models. Overall, we conclude that while important advancements have been made in the last years, several aspects remain to be explored, especially under more realistic evaluation settings.

III. THE BAYESIAN SPAM FILTER MODEL

A general Naive Bayesian spam filtering can be conceptualized into the model presented in Figure. It consists of four major modules, each responsible for four different processes: Tokenization, probability estimation, features selection and Naive Bayesian classification.

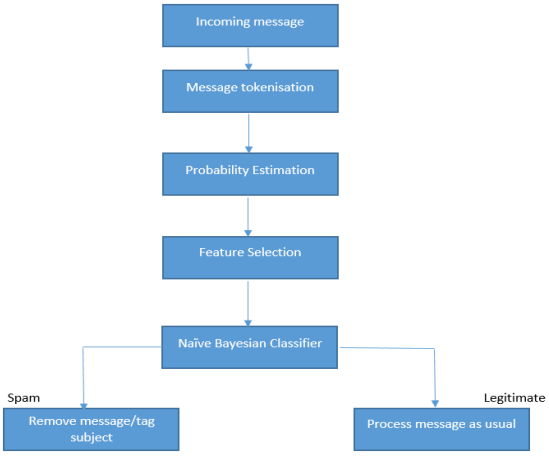


Figure 1 The Bayesian spam filter model

When a message arrives, it is firstly tokenized into a set of features (tokens), F. Every feature is assigned an estimated probability that indicates its spaminess. To reduce the dimensionality of the Feature vector, a feature selection algorithm is applied to output a subset of the features. The Naive Bayesian classifier combines the probabilities of every feature and estimates the probability of the message being spam.

3.1. How the Bayesian Spam Filter Works?

Bayesian filtering is based on the principle that most events are dependent and that the probability of an event occurring in the future can be inferred from the previous occurrences of that event. This technique can be used to classify spam. If some piece of text occurs often in spam but not in legitimate mail, then it would be reasonable to assume that this email is probably spammed. Creating a tailor-made Bayesian word database before mail can be filtered using this method, the user needs to generate a database with words and tokens (such as the $ sign, IP addresses and domains, and so on), collected from a sample of spam mail and valid mail (referred to as 'ham'). Once the ham and spam databases have been created, the word probabilities can be calculated and the filters ready for use. When a new mail arrives, it is broken down into words and the most relevant words – i.e., those that are most significant in identifying whether the mail is spam or not – are singled out. From these words, the Bayesian filter calculates the probability of the new message being spam or not. If the probability is greater than a threshold, say 0.9, and then the message is classified as spam.

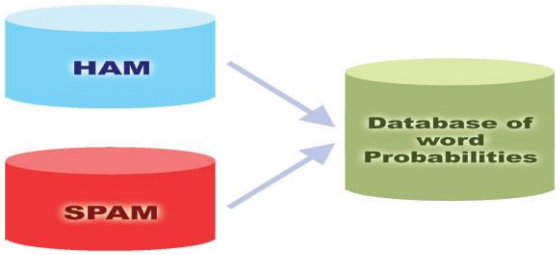


Figure 2 Classification of Spam

IV. WHY BAYESIAN FILTERING IS BETTER?

1. The Bayesian method takes the whole message into account it recognizes keywords that identify spam, but it also recognizes words that denote valid mail. Bayesian filtering is a much more intelligent approach because it examines all aspects of a message, as opposed to keyword checking that classifies a mail as spam on the basis of a single word.
2. A Bayesian filter is constantly self-adapting - By learning from new spam and new valid outbound mails, the Bayesian filter evolves and adapts to new spam techniques. A Bayesian filter is constantly self-adapting - By learning from new spam and new valid outbound mails, the Bayesian filter evolves and adapts to new spam techniques.
3. The Bayesian technique is sensitive to the user. The Bayesian method is multi-lingual and international – A Bayesian anti-spam filter, being adaptive, can be used for any language required.

V. MEASURING THE PERFORMANCE

The meaning of a good classifier can vary depending on the domain in which it is used. For example, in spam classification it is very important not to classify legitimate messages as spam as it can lead to e.g. economic or emotional suffering for the user. To evaluate the performance of a filter, performance indices typical of two distinct areas are commonly used: information retrieval (recall, precision and derived measures) and decision theory (false positives and false negatives).

5.1. Precision and recall

A well employed metric for performance measurement in information retrieval is precision and recall. These measures have been diligently used in the context of spam classification (Sahami et al. 1998). *Recall* is the proportion of relevant items that are retrieved, which in this case is the proportion of spam messages that are actually recognized. For example if 9 out of 10 spam messages are correctly identified as spam, the recall rate is 0.9.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Precision is defined as the proportion of items retrieved that are relevant. In the spam classification context, precision is the proportion of the spam messages classified as spam over the total number of messages classified as spam. Thus if only spam messages are classified as spam then the precision is 1. As soon as a good legitimate message is classified as spam, the precision will drop below 1.It calculates the occurrence of false positives which are good messages classified as spam. When this happens p drops below 1. Such misclassification could be a disaster for the user whereas the only impact of a low recall rate is to receive spam messages in the inbox. Hence it is more important for the precision to be at a high level than the recall rate. The precision and recall reveal little unless used together. Commercial spam filters sometimes claim that they have an incredibly high precision value of 0.9999% without mentioning the related recall rate. This can appear to be very good to the untrained eye. A reasonably good spam classifier should have precision very close to 1 and a recall rate > 0.8.

VI. PROPOSED WORK

Now in these days the email and messaging is a routine work for most of the persons. Sometimes the mails come from the unauthenticated servers or malicious mails can harm the user's privacy and security by fishing emails and their contents. Therefore an adoptive and secure technique is desired which can able to preserve the training feature for future uses. In addition of that system is able to detect the malicious emails with their Spam filtering ability.

Many of the filtering techniques are based on text categorization methods. Thus filtering Spam turns on a classification problem. Roughly, we can distinguish between two methods of machine classification. The first one is done on some rules defined manually. This kind of classification can be used when all classes are static, and their components are easily separated according to some features. The second one is done using machine learning techniques. It is more convenient when the characteristics of discrimination are not well defined. These techniques attempt to generate on a set of samples, quasi or semi automatically a classifier with an acceptable error rate.

6.1. Bayesian classifier

The Naive Bayes classification algorithmic rule is a probabilistic classifier. It is based on probability models [19] that incorporate robust independence assumptions. The independence assumptions usually don't have an effect on reality. So they're thought of as naive. You can derive probability models by using Bayes' theorem (proposed by Thomas Bayes). Based on the nature of the probability model, you will train the Naive Bayes algorithm program in a much supervised learning setting. In straight forward terms, a naive Bayes classifier assumes that the value of a specific feature is unrelated to the presence or absence of the other feature, given the category variable. There are two types of probability as follows:

1. Posterior Probability [P (H/X)]

2. Prior Probability [P (H)]

Where, X is data tuple and H is some hypothesis. According to Baye's Theorem

$$P\left(\frac{H}{X}\right) = \frac{P\left(\frac{X}{H}\right)P(H)}{P(X)}$$

6.2. Neural Network

The implementation of neural network is defined in two phases' first training and second prediction: training method utilizes data and designs the data model. By this data model next phase prediction of values is performed [18].

Training:

1. Prepare two arrays, one is input and hidden unit and the second is output unit.
2. Here first is a two dimensional array W_{ij} is used and output is a one dimensional array Y_i.
3. Original weights are random values put inside the arrays after that the output.

$$x_j = \sum_{i=0} y_i W_{ij}$$

Where, y_i is the activity level of the j^{th} unit in the previous layer and W_{ij} is the weight of the connection between the i^{th} and the j^{th} unit.

4. Next, action level of y_i is estimated by sigmoid function of the total weighted input.

$$y_i = \left[\frac{e^x - e^{-x}}{e^x + e^{-x}}\right]$$

When event of the all output units have been determined, the network calculates the error (E).

$$E = \frac{1}{2}\sum_i (y_i - d_i)^2$$

Where, y_i is the event level of the j^{th} unit in the top layer and d_i is the preferred output of the j_i unit.

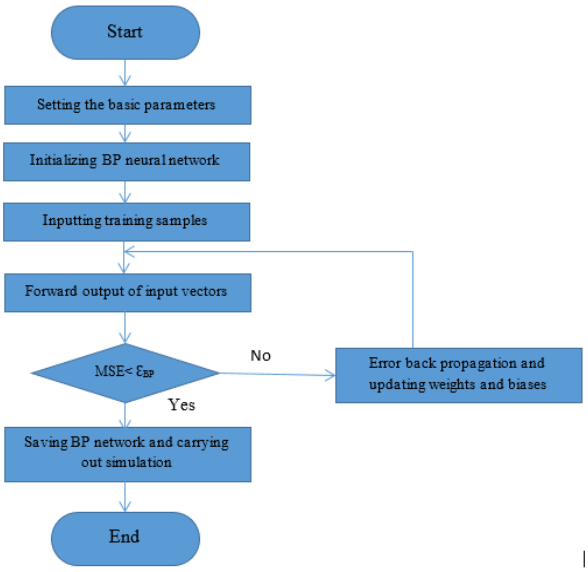


Figure 3 Flow chart of Neural Network

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Calculation of error for the back propagation algorithm is as follows:

Error Derivative (EA_j is the modification among the real and desired target:

$$EA_j = \frac{\partial E}{\partial y_j} = y_j - d_j$$

Error Variations is total input received by an output changed

$$EI_j = \frac{\partial E}{\partial X_j} = \frac{\partial E}{\partial y_j} X \frac{dy_j}{dx_j} = EA_j y_j (1 - y_i)$$

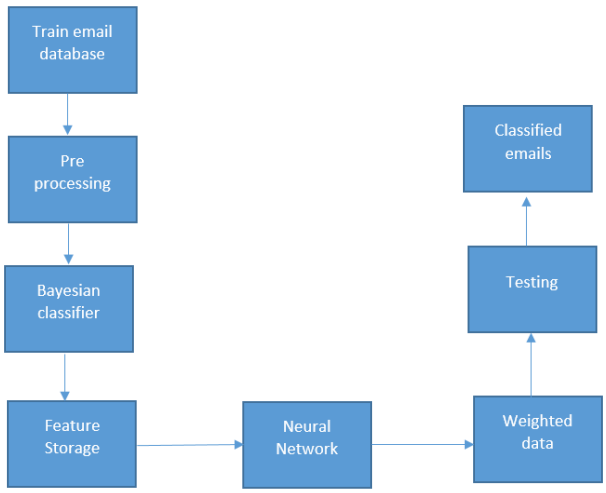
In Error Fluctuations calculation connection into output unit is required:

$$EW_{ij} = \frac{\partial E}{\partial W_{ij}} = \frac{\partial E}{\partial X_j} = \frac{\partial X_j}{\partial W_{ij}} = EI_j y_i$$

Overall Influence of the error:

$$EA_i = \frac{\partial E}{\partial y_i} = \sum_j \frac{\partial E}{\partial x_j} X \frac{\partial x_j}{\partial y_i} = \sum_j EI_j W_{ij}$$

The overview of the proposed systems components discussed in this section.



FLOW CHART

Figure 4 Flow Chart

Training email database -the proposed system is a machine learning technique for classifying emails into four categories. Therefore in order to perform training of the data model a set of pre-labelled data is required for train the classification algorithms.

Pre-processing - the training data is cleaned and the undesired contents from the raw email data are removed. In this process the stop words and the frequent less weighted words are also eliminated from the email contents.

Bay's classifier - in this phase the bay's classifier is applied over the data by which two different probabilities is estimated for each words in data base. For example a word "ICICI bank" the probability to be in spam mail and the same words probability for become in a legitimate mail is estimated as features of the email training.

Feature storage -the extracted features from the bay's classifier is preserved for future use in neural network learning.

Neural network - that is a neural network learning phase where the neural network a word and their probabilities to be in Spam mail and for legitimate mail.

Network weight data: after optimum training with all the words in the bay's database the neural network weights are preserved for future use. When new data set is added to the system the bay's classifiers weights are updated and the neural network weights are cleared and recomputed.

Testing - in this phase the system accepts the emails to classify, therefore the trained neural network load their trained weights and performed the classification.

Classified emails - in this phase the neural networks classification results are listed.

VII. RESULT ANALYSIS

Our novel approach uses a combination of Neural Networks with Bayesian classifier to identify bad and good words in the textual content of an email. Words in the message are per-processed before using the Neural Network classifier. The word goes through stop words and noise removal steps then stemming process step to extract the word root or stem. The experiment shows positive results when compared with base method (Single Bayesian classifier). Results are shown in table below:-

(1) *Accuracy:* - This parameter shows the degree of correctness up to which the system is able to identify the junk emails from bulk amount of unsolicited emails. This is shown in the table below:-

Table I - Accuracy (%)

Sample Data Set	Proposed Method	Base Method
Dataset A	96.728971	93.488372
Dataset B	95.556235	86.956521
Dataset C	93.336589	86.656986
Dataset D	97.0297029	86.956521

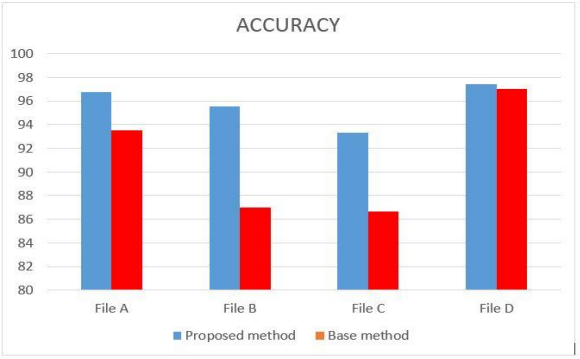


Figure V– Accuracy comparison (%)

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

(2) *Error rate:* - This parameter shows the error rate that occurred during classification of spam from available datasets. This comparison is shown in the table below:-

Table 2 – Error Rate Comparison (%)

Sample Data Set	Proposed Method	Base Method
Dataset A	3.271028	6.511627
Dataset B	4.442563	13.0434782
Dataset C	6.666536	13.333333
Dataset D	2.986206	13.970297

Figure VI – Error rate comparison (%)

3) *Recall:* - This parameter shows the proportion of relevant items that are retrieved, which in thiscase is the proportion of spam messages that are actually recognized. These comparisons are shown in the table below:-

Table 3 Recall

Sample Data Set	Proposed Method	Base Method
Dataset A	0.988655	0.850746
Dataset B	0.975666	0.7252631
Dataset C	0.8963333	0.692307
Dataset D	0.886666	0.758654

Figure VII –Recall value comparison

4) *Time taken:* - This parameter shows the time taken by two methods in executing the same files. It deals with the time complexity of proposed method in comparison with base method. This comparison is shown in the figure below:-

Table 4 – Time Taken Comparision (ms)

Sample Data Set	Proposed Method	Base Method
Dataset A	321	269
Dataset B	241	213
Dataset C	134	178
Dataset D	223	218

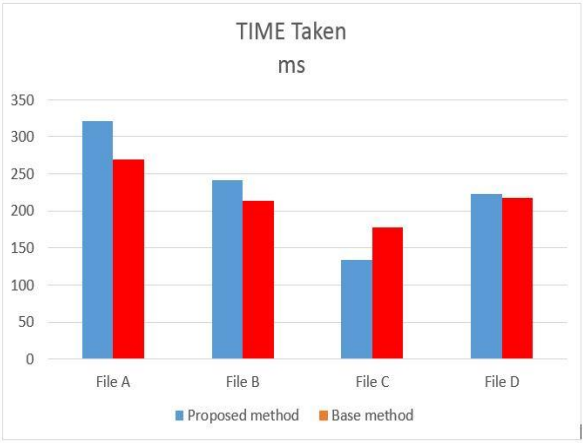


Figure VIII –Time taken comparison

VIII. CONCLUSION

Spam is becoming a very serious problem to the Internet community, threatening both the integrity of the networks and the productivity of the users. In this paper, we proposed an enhanced spam filter method. From the results shown above, it is clear that the proposed method is better as compared to standard method, hence it can be concluded that the proposed system is accurate and filters the spam more precisely. It has been justified using various measurement parameters such as recall, precision. In conclusion we can say that the proposed method i.e. the combination of Neural Network along with Bayesian Classifier shows better result as compared to Single Bayesian classifier.

ACKNOWLEDGMENT

We are thankful to Dr. Vivek Kapoor, Assistant Professor of IET, Davv Indore for guiding and manuscript formatting. We also like to thank to almighty for giving us strength to do fruitful work in the area of research to make our nation building.

REFERENCES

1. Aladdin Knowledge Systems, Anti-spam white paper,www.csisoft.com/security/aladdin/esafe_antispam_whitepaper.pdf Retrieved December 28, 2011.
2. F. Smadja, H. Tumblin, "Automatic spam detection as a text classification task", in: Proc. of Workshop on Operational Text Classification Systems, 2002.
3. Hassanien, H. Al-Qaheri, "Machine Learning in Spam Management", IEEE TRANS., VOL. X, NO. X, FEB.2009
4. P. Cunningham, N. Nowlan, "A Case-Based Approach to Spam Filtering that Can Track Concept Drift", [Online]Available:https://www.cs.tcd.ie/publications/techreports/reports.03/TCD-CS-2003-16.pdf RetrievedDecember 28, 2011
5. F. Roli, G. Fumera, "The emerging role of visual pattern recognition in spam filtering: challenge and opportunity forIAPRresearchers" http://www.iapr.org/members/newsletter/Newsletter07-02/index_files/Page465.htm Retrieved December 28, 2011
6. H. West, "Getting it Wrong: Corporate America Spams the Afterlife". Clueless Mailers. (January 19, 2008).
7. Parizo, "Image spam paints a troubling picture". Search Security. (2006-07-26)
8. Symantec (2011) VBS.Davinia.B, [Online] Available: http://www.symantec.com/security_response/writeup.jsp?docid=2001-020713-3220-99 Retrieved December 28, 2011
9. Androutsopoulos, J. Koutsias, "An evaluation of naïve bayesian anti-spam filtering". In Proceedings of the Workshop on Machine Learning in the New Information Age, 11th European Conference on Machine Learning (ECML 2000), pages 9–17, Barcelona, Spain, 2000.
10. Androutsopoulos, G. Paliouras, "Learning to filter spamE-mail: A comparison of a naïve bayesian and a memorybasedapproach". In Proceedings of the Workshop on Machine Learning and Textual Information Access, 4thEuropean Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2000), pages 1–13, Lyon, France, 2000.
11. J. Hidalgo, "Evaluating cost-sensitive unsolicited bulk email categorization". In Proceedings of SAC-02, 17th ACM Symposium on Applied Computing, pages 615–620, Madrid, ES, 2002.
12. K. Schneider, "A comparison of event models for naïve bayes anti-spam e-mail filtering". In Proceedings of the 10thConference of the European Chapter of the Association for Computational Linguistics, 2003.
13. Witten, E. Frank, "Data Mining: Practical MachineLearning Tools and Techniques with Java Implementations".Morgan Kaufmann, 2000.
14. N. Cristianini, B. Schoelkopf, "Support vector machines and kernel methods, the new generation of learning machines". Artificial Intelligence Magazine,23(3):31–41, 2002
15. V. Vapnik, "The Nature of Statistical Learning Theory, Springer; 2 edition (December 14, 1998)
16. S. Amari, S. Wu, "Improving support vector machine classifiers by modifying kernel functions". Neural Networks, 12, 783– 789. (1999).
17. Miller, "Neural Network-based Antispam Heuristics" ,Symantec Enterprise Security (2011), www.symantec.com Retrieved December 28, 2011
18. Wu, "Behavior-based spam detection using a hybrid method of rule-based techniques and neural networks" ,Expert Syst., 2009
19. A review of machine learning approaches to spam filtering. Thiago S Guzella, Walmir M. Caminhas.

Author Profile

Rahul Maheshwari is an M.Tech Scholar of Department of Computer Science and Engineering at Institute of Engineering and Technology,DAVV, Indore, India. His areas of research are Cryptography and Network security, mobile computing, parallel and distributed computing and Neural Networks.

Dr. Vivek Kapoor is an Assistant Professor of Department of Information Technology at Institute of Engineering and Technology,DAVV, Indore, India. His areas of research are Cryptography, Network security, mobile computing Computational Complexity, and Network Optimization and Computer Vision and Data Sciences

Sandeep Verma is an M.Tech Scholar of Department of Information Technology at Institute of Engineering and Technology,DAVV, Indore, India. His areas of interest are Biometrics, Supply Chain Management, Data Mining, Cryptography.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)