

# Bayesian Neural Network Impact on Email Spam Filter

S. Prince Sahaya Brighty<sup>1</sup>; Dharsana.R<sup>2</sup>; HariniParthiban<sup>3</sup>

Assistant Professor, Department of CSE, Sri Ramakrishna Engineering College<sup>1</sup>

II year BE CSE Student, Sri Ramakrishna Engineering College<sup>2</sup>

II year BE CSE Student, Sri Ramakrishna Engineering College<sup>3</sup>

[brighty.s@srec.ac.in](mailto:brighty.s@srec.ac.in) [dharsuravi797@gmail.com](mailto:dharsuravi797@gmail.com) [hariniparthiban@gmail.com](mailto:hariniparthiban@gmail.com)

## ABSTRACT

*The volume of mass unsolicited electronic mail, often known as Spam, bulk e-mail or junk mail, has recently increased enormously and has become a serious threat to not only the Internet but also to society and Spam mail, that is sent to a group of recipients who have not requested it. These unsolicited mails have already caused many problems such as filling mailboxes, engulfing important personal mail, wasting network bandwidth, consuming users' time and energy to sort through it, other problems associated with Spam mails are crashed mail-servers, pornography adverts sent to children, and so on. The task of Spam filtering is to rule out unsolicited mails automatically from a user's mail stream. The present Spam filters are static rule based and are only have the email header analysis, so they are unable to filter the Spam in full-fledged manner.*

*A new Spam detection method using Text Categorization, which uses Rule based heuristic approach and statistical analysis tests to identify "Spam". This method includes Header Analysis - Spot the Common Spammer tricks in headers, Content Text Analysis - Spot the Common Spammer phrases / format in the Body of email, Blacklists - Spot e-mails from suspected Spammer networks, Learning Classifier, Bayesian probability analysis. Bayesian methods have been applied to ANNs in order to regularize training to improve the robustness of the classifier. The goal of training a Bayesian ANN with finite sample sizes is, as with unlimited data, to approximate the ideal observer. This strategy can provide improved filtering of Spam than existing Static Spam filters.*

**KeyWords :** Email Spam, Neural Network, Bayesian Text Classification, Meta Classification, Neural Network.

## 1. INTRODUCTION

The volume of junk e-Mail (spam) transmitted by the Internet has arguably reached epidemic proportions. While the inconvenience of spam is not new – public comments about unwanted e-mail messages identified the problem as early as 1975 – the volume of unsolicited commercial e-Mail was relatively limited until the mid-1990s. Spam volume was estimated to be merely 8% of network e-mail traffic in 2001 but has ballooned to about 40% of e-Mail today. One research firm has predicted that the cost of fighting spam across the U.S. will approach \$25 billion in 2005.

Most e-Mail readers must spend a non-trivial amount of time regularly deleting spam messages,

even as an expanding volume of junk e-Mail occupies server storage space and consumes network bandwidth. An ongoing challenge, therefore, rests within the development and refinement of automatic classifiers that can distinguish legitimate e-mail from spam. Many commercial and open-source products exist to accommodate the growing need for spam classifiers, and a variety of techniques have been developed and applied toward the problem, both at the network and user levels. The simplest and most common approaches are to use filters that screen messages based upon the presence of common words or phrases common to junk e-mail. Other simplistic approaches include *blacklisting* (automatic rejection of messages received from the addresses of known spammers) and *whitelisting* (automatic acceptance of

message received from known and trusted correspondents).

In practice, effective spam filtering uses a combination of these three techniques. The primary flaw in the first two approaches is that it relies upon complacency by the spammers by assuming that they are not likely to change (or forge) their identities or to alter the style and vocabulary of their sales pitches. White listing risks the possibility that the recipient will miss legitimate e-mail from a known or expected correspondent with a heretofore unknown address, such as correspondence from a long-lost friend, or a purchase confirmation pertaining to a transaction with an online retailer. A variety of text classifiers have been investigated that categorize documents topically or thematically, including probabilistic, decision tree, rule-based, example-based ("lazy learner"), linear discriminate analysis, regression, support vector machine, and neural network approaches. A prototype system has also been designed to recognize hostile messages ("flames") within online communications. However, the body of published academic work specific to spam filtering and classification is limited. This may seem surprising given the obvious need for effective, automated classifiers, but it suggests two likely reasons for the low volume of published material. First, the effectiveness of any given anti-spam technique can be seriously compromised by the public revelation of the technique since spammers are aggressive and adaptable. Second, recent variations of Naïve Bayesian classifiers have demonstrated high degrees of success. In general, these classifiers identify attributes that are assigned probabilities by the classifier. The product of the probabilities of each attribute within a message is compared to a predefined threshold, and the messages with products exceeding the threshold are classified as spam.

## 2. EXISTING APPROACHES

Detecting and filtering spam creates a number of complex challenges due to the dynamic nature of junk e-mail. An effective spam filter must block the maximum unwanted e-mail, with the minimum number of false positives (messages, wrongly identified as spam). This is further complicated by the fact individual users have different views on what they consider to be spam. Many users are very happy to receive adverts from well-known Internet retailers while others consider

this junk. There are a number of techniques available today for spam detection, including:

- 1 Real Time Black Lists
- 2 Lexical Analysis
- 3 DCC
- 4 White & Black Lists

## 3. PROPOSED IDEA

### 3.1. Objective

In this Paper we propose the Solution for this problem, a Perl-based application supervised neural network based learning and Bayesian classification for filtering, which is usually used to filter all incoming mail for one or several users. It can be run as a standalone application or as a client that communicates with a spammed. The latter mode of operation has performance benefits, but under certain circumstances may introduce additional security risks. Typically either variant of the application is set up in a generic mail filter program, or it is called directly from a mail user agent that supports this, whenever new mail arrives. Mail filter programs such as procmail can be made to pipe all incoming mail through this project with an adjustment to Organization rules file.

This Spam filter works with a large set of rules which are applied to determine whether an email is spam or not. To decide, specific fields within the email header and the email body are typically searched for certain regular expressions, and if these expressions match, the email is assigned a certain score, depending on the test, and several (customizable) headers are added to the mail. The total score resulting from all tests or other criteria can then be used by the end user or by the ISP to set the conditions under which email is moved to a separate spam folder, deleted, flagged etc.

Each test has a label and a description. The label is usually an all upper case identifier separated with underscores, such as "LIMITED\_TIME\_ONLY", with the description for that label being "Offers a limited time offer". A mail that passes that test might be assigned a score of +0.3. With a spam threshold of 5, several other tests would usually have to pass for the mail to be classified as spam. On the other hand, some tests, such as those for

invalid message IDs or years, result in a very high score being assigned, where even a single test can almost put a mail "over the edge".

When a mail's total score is higher than the required-hits setting in Spam filter's configuration, the mail is treated as spam and rewritten according to several options. In the default configuration, the content of the mail is appended as a MIME attachment, with a brief excerpt in the message body, and a description of the tests which resulted in the mail being classified as spam. If the score is lower than the defined settings, by default the information about the passed tests and total score is still added to the email headers and can be used in post-processing for less severe actions, such as tagging the mail as suspicious.

## 1. PROPOSED METHODOLOGIES

### 4.1. Statistical Filter

This Spam Filter uses a wide range of heuristic tests on mail headers and body text to identify "spam", also known as unsolicited commercial email. Once identified, the mail can then be optionally tagged as spam for later filtering using the user's own mail user-agent application. Spam Filter typically differentiates successfully between spam and non-spam in between 95% and 100% of cases, depending on what kind of mail you get and your training of its Bayesian filter. Specifically, Spam Filter has been shown to produce around 0.9% false negatives (spam that was missed) and around 0.1% false positives (ham incorrectly marked as spam).

This Paper proposes, a new Spam detection method using Text Categorization, which uses Rule based heuristic approach and statistical analysis tests to identify "Spam". This method includes Header Analysis - Spot the Common Spammer tricks in headers, Content Text Analysis - Spot the Common Spammer phrases / format in the Body of email, Blacklists – Spot e-mails from suspected Spammer networks, Learning Classifier, Bayesian probability analysis. Bayesian methods have been applied to ANNs in order to regularize training to improve the robustness of the classifier. The goal of training a Bayesian ANN with finite sample sizes is, as with unlimited data, to approximate the ideal observer.

This strategy can provide improved filtering of Spam than existing Static Spam filters.

### 4.2. Bayesian Classifier

Using Bayesian inferential statistics is a relatively new innovation in spam detection. The concept of Bayesian filtering is to create two databases or 'corpus' of e-mail: A corpus of spam e-mail and a second of valid e-mail. Each corpus is then 'tokenized' and analyzed looking for tokens that frequently appear in each type of e-mail. Each token is then given a probability weighting, suggesting if it is likely to appear in spam or valid e-mail. Each new message is then processed and tokenized, and the tokens compared against the existing database to determine the probability that the message is valid or spam. This approach creates some unusual but effective results, for example the token: "ff0000" frequently appears in spam corpuses. This is the HTML tag for bright red, something spammers often use to highlight their special offers. A key benefit of this approach is that it is possible to tune the filters to different customer environments. For example the token 'Viagra' would typically appear in the corpus of Spam with a high Spam probability. However for pharmaceutical companies selling Viagra it is also likely to appear in their valid e-mail; in this case 'Viagra' would be a less significant indicator of spam

### 4.3. Meta Classifier

The Meta-classifier approach is one of the simplest approaches to this problem. Given a base classifier, the approach is to learn a Meta classifier that predicts the correctness of each instance classification of the base classifier. The source of the Meta training data are the training instances. The Meta label of an instance indicates reliable classification, if the instance is classified correctly by the base classifier; otherwise, the Meta label indicates unreliable classification. The Meta classifier plus the base classifier form one combined classifier. The classification rule of the combined classifier is to assign a class predicted by the base classifier to an instance if the Meta classifier decides that the classification is reliable.

#### 4.4. Neural Network Learning

One of the most interesting properties of a neural network is the ability to learn from its environment in order to improve its performance (measured through a predefined performance measure) over time. Learning in an artificial neural network stands for an iterative process of adjusting the synaptic weights and threshold values. In artificial neural networks the following learning procedures can be distinguished:

- Supervised learning (teacher or target is available)
- Unsupervised learning (without teacher)
- Reinforcement learning (without target, but with reward)

#### 5. CONCLUSION

We present a statistical algorithm for automatic personalized spam filtering that does not require users (Administrator) to provide feedback regarding the classifications of e-mails in their inboxes. The algorithm builds a statistical model of words from training corpus and then adapts it to the distribution of words and e-mails in each individual user's inbox. Overall, the algorithm requires one pass over e-mails in the training corpus and three-passes over e-mails in the individual user's inbox. Our experiments confirm the benefit of personalization with significant performance gains over a filter that assumes that training and evaluation (users' inboxes) datasets follow the same distribution. We present extensive results of our algorithm including a discussion on the estimation of its two parameters. The problem of automatic personalized spam filtering has generated much interest recently. It is a technically challenging problem that promises significant benefit to email users and e-mail service providers.

#### 6. REFERENCES

- [1] Eirinaios Michelakis, Ion Androutsopoulos, Georgios Paliouras, George Sakkis, and Panagiotis Stamatopoulos, "Filtron: A Learning-Based Anti-Spam Filter", First Conference on Email and Anti-Spam (CEAS), 2004 Proceedings, Mountain View, CA, July 30 and 31, 2004.
- [2] Georgios Sakkis, Ion Androutsopoulos, Georgios Paliouras, "A Memory-Based Approach to Anti-Spam Filtering for Mailing Lists" IEEE on Information Retrieval Volume 6, Issue 1 (January 2003) Pages: 49 - 73 Year of Publication: 2003, ISSN:1386-4564
- [3] Le Zhang, Jingbo Zhu, and Tianshun Yao, "An Evaluation of Statistical Spam Filtering Techniques", ACM Transactions on Asian Language Information Processing (TALIP), Volume 3, Issue 4 (December 2004), Pages: 243 - 269, Year of Publication: 2004, ISSN:1530-0226
- [4] Marwan A. Mattar, Allen R. Hanson, and Erik G. Learned-Miller, "Sign Classification using Local and Meta-Features", Computer Vision and Pattern Recognition, 2005, IEEE Computer Society Conference, Publication Date: 20-26 June 2005, Volume: 3, On page(s): 26 - 29.
- [5] Wei-Hao Lin, Rong Jin, Alexander Hauptmann, "Meta-Classification of Multimedia Classifiers", International Workshop on Knowledge discovery in multimedia and complex data, Taipei, Taiwan, May 6, 2002.
- [6] Xin Jin; Anbang Xu; Bie, R.; Xian Shen; Min Yin, Granular Computing, 2006, "Spam email filtering with Bayesian belief network: using relevant words", IEEE International Conference on Volume 3, Issue 1, 10-12 May 2006 Page(s): 238 - 243
- [7] <http://www.symantec.com/searchlanding/antispam/>
- [8] <http://www.ironport.com>
- [9] <http://www.spamassassin.apache.org>
- [10] <http://w2.syronex.com/jmr/safe>