Image Spam Filtering using Support Vector Machine and Particle Swarm Optimization

T. Kumaresan
Assistant Professor (Sr. Grade)
Department of CSE
Bannari Amman Institute of Technology
Sathyamangalam, India

S. Sanjushree
PG Scholar
Department of IT
Bannari Amman Institute of Technology
Sathyamangalam, India

K. Suhasini
PG Scholar
Department of IT
Bannari Amman Institute of Technology
Sathyamangalam, India

C. Palanisamy, Ph.D.
Professor and Head
Department of IT
Bannari Amman Institute of Technology
Sathyamangalam, India

ABSTRACT
Spam is most often considered to be electronic junk mail. Spam is defined as unsolicited bulk mail. Image spam is a kind of email spam where the spam text is embedded with an image. Spam email has become difficult in the survival of internet users, causing personal injury and economic losses. In this paper, we propose a feature extraction scheme which focuses on low-level features, like metadata and visual features of images. This technique makes classification better and it is an effective method because it does not depend on extracting text and examining the content of email. A SVM classifier with kernel function is used to identify an image spam and also the accuracy will be calculated.

General Terms
Spam image, Ham image, Support Vector Machine, Particle Swarm Optimization

Keywords
Email, ham image, spam image, svm classifier

1. INTRODUCTION
Email spam, also known as junk email or unsolicited bulk email (UBE), is a subset of electronic spam involving nearly identical messages sent to numerous recipients by email. Clicking on links in spam email may send users to phishing web sites or sites that are hosting malware. Spam email may also include malware as scripts or other executable file attachments. Spam has several interpretations varying by source.

- Unsolicited bulk email (UBE)—unsolicited email, sent in huge amounts.
- Unsolicited commercial email (UCE) — it is a legal term used to describe an electronic promotional message sent to a consumer without the consumer's prior request or consent. It is mainly used for commercial gain.

1.1 Image Spam
Image spam is a confusing method in which the text of the message is stored in the image format (i.e., gif, jpeg, png, etc) and displayed in the email. It prevents text based spam filters from detecting and blocking spam messages. Habitually, image spam contains senseless, computer-simulated text which simply irritates the reader. A new-fangled technique, however, is to use an image that does not contain clear text in its initial frame, or to twist the shapes of letters in the image (as in CAPTCHA) to avoid detection by OCR tools. Affording to the content of image spam, it can be divided into types as follow:

- Advertisement Image Spam: The content of image spam is sale promotion, product advertising, or shopping online, and so on.
- Pornography Image Spam: The content of image spam includes pornographic image or pornographic product.
- URL (Uniform Resource Locator) Image Spam: The content of image spam contains a short textual description and a URL web site. Since the URL web site and HTML tags are difficult to be recognized, URL image spam becomes the chief means of phishing.
- Reactionary Image Spam: The content of image spam is negative, and against government.

1.2 Characteristics
Image spam has many characteristics, such as rich in content, various forms, and large size. The harm of the image spam is larger than majority spam. The size of majority spam is less than 5K. While the size of image spam is usually more than 12k. The statistics of Commtouch show that image spam which is 35% of all spam causes 70% bandwidth bulge. Image spam can thwart traditional text filter by using image process techniques. To circumvent anti-spam system, Image spam has several advantages for spammers:

- It is easy to let the message across the anti-spam which uses scanning the message body for spam like text to detect spam.
- Pretty graphics let the email has a colorful and "professional" message.

Using several new techniques, spammers can randomize each message, again by-passing antispam techniques based on signatures.

2. RELATED WORK

Spam e-mail with moneymaking text embedded in images presents a great challenge to anti-spam filters. In this paper, the authors present a fast method to detect image spam from an incoming e-mail. A simple edge-based feature method is used to evaluate a vector of similarity scores between an image and a set of patterns. Then, the evaluated similarity vector is used with support vector machines to separate spam images from other common kinds of images. This method does not require expensive OCR or even text extraction from images. Empirical results show that the method is fast and has good classification accuracy. [1]

In this paper, a framework for filtering image spams by using the Fourier-Mellin invariant features. Fourier-Mellin features are robust for most kinds of image spam variations. A one-class classifier, the support vector data description (SVDD), is exploited to model the boundary of image spam class in the feature space without using information of legitimate emails. Experimental results demonstrate that their framework is effective for fighting image spam. [2]

Image spam has become the main form of spam, which is a problem which is looking for solutions to effectively filter such spam nowadays. In this paper, author proposed an image spam detection method that is based on image semantics and near-duplicate detection, for solving the difficulties of existing anti-image spam skills: decreased accuracy rate, difficultly in recognizing image spam by making use of complication techniques and so on. The trial results show that the system has better filtering effect than previous systems, with increasing accuracy rate and better anti-obfuscation effect, and effectively solves the above described problems. [3]

In recent eras to escape from the spam detection of text-based spam filtering system, spammers insert junk facts into the piece of mail with images, and attach it to the message body. The conventional text-based filter cannot handle such spam image. In order to deal with the spam which contains text and images, a filtering technique which fuses text, image and other multi-modal features is proposed in this paper. Initially, mining the text features and image features to build multiple classifiers, and then by using the fusion method to choose the output of multiple classifiers. Experimental results show that the fusion method can have a better result than that of a single classifier. [8]

Spam is no more junk but risk since it recently includes virus attachments and spyware agents which make the recipients' system abandoned; therefore, there is an evolving need for spam detection. Many spam detection techniques based on machine learning algorithms have been proposed. As the amount of spam has been increased tremendously using bulk mailing tools, spam detection techniques should deal with it. For spam detection, parameters optimization and feature selection have been proposed to reduce processing overheads with guaranteeing high detection rates. However, the previous approaches have not taken into account variable importance and optimal number of features and there are no approaches using both of them together so far. In this paper, the authors propose an optimal spam detection model based on Random Forests (RF) which enables parameters optimization and feature selection. [12]

3. EXISTING SYSTEM

Email is considered as the most suitable way for transferring message through Internet. But the flooded of spam becomes fast now a days and the developers also paid attention towards spam. To solve this problem, many outcomes have been recommended to reduce spam, which can classify spam from emails successfully. The battle between virus and antivirus, spammers explore new technologies to keep one step ahead of spam filters. Spammers' most recent method involves image spam, in which the main payload of the spam message is carried as an embedded image. Generally, the body of image spam contains no text or only fake text. The conventional text based spam filters are failed to detect and to block it. Most of spam that breaks through the personal anti-spam is image spam. For the time being, image spam is more interesting and definite than text alone.

4. PROPOSED SYSTEM

In this paper, we aim to classify the spam and ham images. A feature extraction scheme and a one-class SVM classifier with RBF kernel as the kernel function are proposed, which mainly focus on low-level features of image. The basic idea of image spam is based on two steps. The first step is to get image features, which contains file properties and texture properties. Normally first step is fast, because of not extracting the text from image and not considering the content of text. The second step is to run a one-class SVM classifier with RBF kernel as the kernel function to classify image spam.

4.1 Overview of Architecture

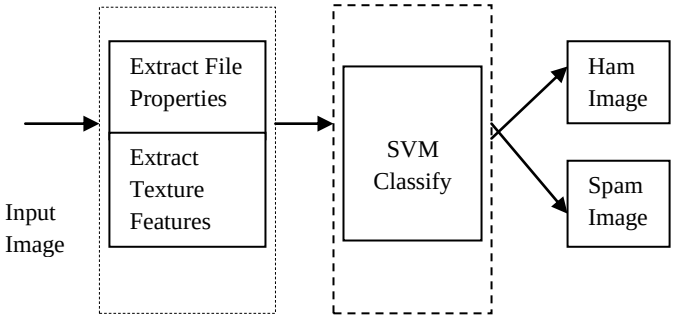

Fig 1: System Architecture

4.2 Modules

4.2.1 Feature Extraction

Feature extraction is an image processing technique is used to reduce the dimensionality of an image. It involves reducing the amount of resources required to describe a large set of data. Evaluation with a large number of variables normally requires a large amount of memory and computation power. It is a common term for methods of constructing groupings of the variables to get around these problems while still describing the data with sufficient accuracy.

4.2.2 Metadata Features

The five basic metadata features that can quickly derived from an image at extremely low computational cost. These features are: Image Size, Width, Height, Bit Depth, and Image File Type. Based on these raw features, we generate a 10 dimensional feature vector. The image file type features (f7, f8 and f9) are binary features that are set to 1 if the file is of specified type and to 0 otherwise.

Table 1. File Properties

Symbol	Description
f_1	Width
f_2	Height

f_3	Aspect ratio : f_1/f_2
f_4	File size
f_5	Image area : f_1*f_2
f_6	Compression : f_4/f_5
f_7	Binary : JPEG image
f_8	Binary :GIF image
f_9	Binary :PNG image
f_{10}	Bit depth

4.2.3 Texture Features

Histogram is a statistical presentation of image. For a gray image, the gray histogram reflects the statistics of the different gray scale appearance. The definition of global gray histogram is in Eq. (1)

$$H(k) = \frac{r_k}{N} \qquad (1)$$

k is the gray value of gray image, rk is the number of pixels which the gray value is k. N is the total number of image pixels. Texture is a reaction to an image in a region of the spatial distribution of pixel gray-level properties, the inherent properties of the structure of this space neighborhood pixels can be directly related to portray. The simplest way to describe the demographic characteristics of the texture is the moment of gray histogram.

$$\mu(r) = \sum_{i=1}^{L} (r_i - m) n\, H\,(r_i) \qquad (2)$$

L is the dimension of histogram, m the mean histogram, μn is directly related to the shape of H(r), but these moments with the texture of the absolute position in space has nothing to do. In the histogram of the n-order moments, variance μ2 is a measure of gray scale contrast, an expression of the curve relative to the mean of the distribution. It describes the extent of a relatively smooth histogram reflects the dispersion degree of gray scale images; μ3 describes skewness, which expresses the curve relative to the mean symmetry. It describes the degree of histogram skew, which is the case whether or not asymmetric histogram points; μ4 can be defined kurtosis, which indicating the relative flatness of the histogram. It is histogram distribution of gathered in the vicinity of or close to the mean at both ends of the situation, which describing the gray-scale image texture contrasts. The paper utilizes μ2, μ3, μ4 as texture features.

4.2.4 Classification

A large number of classification algorithms have been applied to spam detection area, where support vector machine classification for its decent generalization performance effect and very popular. SVM is a powerful technique used for data classification. Even though people consider that it is easier to use than Neural Networks. Each instance in the training set contains one class labels and several features. The main aim of SVM is to produce a model which predicts class labels of data instances in the testing set which are given only the features. At present, the support vector machines have been widely used in text-based anti-spam system. SVM is a splendid solution to the small sample size problem, by constructing a separating hyper plane to complete the classification. As the support vector machine in spam detection in the good performance, the paper uses this algorithm to identify spam images.

4.2.5 Support Vector Machine

A support vector machine (SVM) can be used when our data has absolutely two classes. An SVM classifies data by finding the perfect hyperplane that separates all data points of one class from those of the other class. The hyperplane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the portion parallel to the hyperplane that has no interior data points.

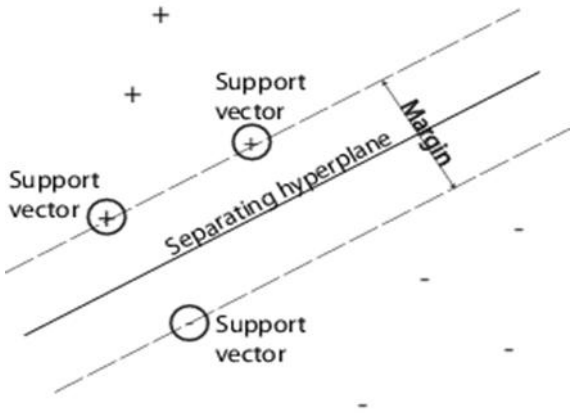

Fig 2: Support Vector Machine

4.2.5.1 Properties of SVM

Support Vector Machine belongs to a family of generalized linear classifiers and it can be interpreted as an extension of the perception. A special property is that they simultaneously minimize the empirical classification error and maximize the geometric margin; hence they are also known as maximum margin classifiers.

4.2.6 Particle Swarm Optimization

Particle swarm optimization (PSO) is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with respect to a given degree of quality. The algorithm optimizes a difficulty by taking inhabitants of candidate solutions, here labeled particles, and moving those particles around in the search-space according to simple calculated formulae over the particle's location and velocity. Each particle's movement is inclined by its local best known position but, it is guided towards the best known positions in the search-space, which are updated as better locations are bring into being by other particles. PSO is a metaheuristic method as it makes few or no assumptions about the problem being optimized and can search very huge spaces of candidate results. However, metaheuristic method such as PSO does not guarantee an optimum solution is ever bringing into being. More specifically, PSO does not use the gradient of the problem being enhanced, which means PSO does not require that the optimization problem be differentiable as is required by classic optimization methods such as gradient descent method and quasi-newton method. It can also be used on optimization problems that are moderately irregular, noisy, change over time, etc.

4.2.7 Dataset

An image-based spam has been widespread several years, but the publicly available image corpus is quite lacking. Personal Ham subset contains 1786 images and personal spam subset contains 3203 images. Addition to that the majority of the spam image is from the large spam corpus namely spamarchive.

Table 2. A Summary of the Dataset

Corpus	Number of images
Personal Spam	3203
Personal Ham	1897

5. EXPERIMENTAL RESULTS

The experimental results show that the method is based on metadata and texture features using Support Vector Machine learning and PSO is able to achieve 90% detection rate. By changing the training and testing values, we obtain the below accuracy shown in the table 2. But the time complexity still is a problem, because it takes 82 sec to evaluate all images in the dataset. In this paper, we have extracted metadata features and texture features of an image for easy classification. SVM classifier is used to detect an image spam. PSO optimizer is used to optimize the obtained results and it works only for small dataset. If we increase the dataset means the accuracy of the system get reduced.

Table 3. Experimental Results

Spam Images (1886 images)		Ham Images (930 images)		Accuracy	Time Complexity
Train	Test	Train	Test		
40	100	20	40	82.5	81.34sec
50	150	50	90	30.71	82.57sec
60	120	30	60	67.78	81.98sec
90	150	20	50	83.3	81.75sec
100	160	30	60	78.88	81.93sec
100	160	40	60	90	82.44sec
200	260	100	120	75	238.89sec
1000	1060	500	600	48.75	238.64sec

6. CONCLUSION

Image spam is a kind of email spam where the spam text is embedded with an image. Image spam erodes the limited network resources, and brings troubles to people. In this project, we propose a feature extraction scheme which focuses on low-level features, like metadata and visual features of images. This technique makes classification better and it is an effective method because it does not depend on extracting text and examining the content of email. A SVM classifier with the kernel function is used to identify an image spam. Experimental results shown that this extraction scheme is efficient. But still accuracy will be a problem in this method, because of corrupted and multiframe images. In future we can implement our project for more number of features and can use some optimization algorithms which optimize the detection rate according to the dataset. The time complexity problem will also be reduced, when considering the large dataset. The corrupted images will be removed and also we construct the system to work well with multiframe images.

7. REFERENCES

[1] N. Nhung and T. Phuong. "An Efficient Method for Filtering Image Based Spam mail". Proc. IEEE International Conference on Research, Innovation and Vision for the Future (RIVF 07), IEEE Press, Mar. 2007, pp. 96-102. doi: 1O.1109/RIVF.2007.369141.

[2] B. Mehta, S. Nangia, M. Gupta. “Detecting image spam using visual features and near duplicate detection”. WWW, 2008.

[3] HQ. Zuo, x. Li, O. Wu, W.M. Hu , G. Luo. “Image spam filtering using Fourier-Mellin invariant features”. ICASSP, 2009.

[4] Z. Qu and Y Zhang, "Filtering Image Spam using Image Semantics and Near-Duplicate Detection," Proc. the 2nd International Conference on Intelligent Computation Technology and Automation (ICICTA 2009), IEEE Press, Oct. 2009, in press.

[5] Z. Wang, W. Josephson, Q. Lv, M. Charikar, K. Li. "Filtering Image Spam with Near Duplicate Detection". In Fourth Conference on Email and Anti-Spam, August 2-3, 2007, Mountain View, California USA.

[6] C. Wu, K. Cheng, Q. Zhu, and Y Wu. "Using Visual Features for Anti-spam Filtering". Proc. IEEE International Conference. Image Processing (ICIP 05). IEEE Press, Sept. 2005 pp. 509-12.doi:10.1109/ICIP.2005.1530440.

[7] M. Dredze, R. Gevaryahu, A. E.Bachrach. "Learning Fast Classifiers for Image Spam". In Fourth Conference on Email and Anti-Spam, August 2-3, 2007, Mountain View, California USA.

[8] H Aradhye, G. Myers, and J. Herson. "Image Analysis for Efficient Categorization of Image-based spam E-mail". Proc. IEEE Conf. Document Analysis and Recognition (ICDAR 05), IEEE Press, Aug 2005 pp.914-918, doi: 1O.1109/lCDAR2005.135.

[9] G. Fumera, I. PilIai, F. Roli."Image Spam Filtering using textual and visual Information". MIT Spam Conference 2007, 30 March 2007, Cambridge, MA, USA.

[10] W. Ma, D. Tran, and D. Sharma, "Detecting Image based Spam Email," Proc. First International Conference on Hybrid Information Technology (ICHIT 06), Springer Press, Nov. 2006, pp. 168-177, doi: 10.1007/978-3-540-77368-9.

[11] J. Shih and L. Chen, "Color Image Retrieval based on Primitives of Color Moments," Lecture Notes in Computer Science, vol 2314, 2002, pp. 19-27, doi: 10.1007/3-540 45925-1_8.

[12] D. Gavilan, H. Takahashi, and M. Nakajima, "Image Categorization Using Color Blobs in a Mobile Environment," Computer Graphics Forum (EG 2003), 22(3), 2003, pp. 427-432.

[13] A.K. Jain and A. Vailaya, "Shape-based retrieval: a case study with trademark image database", Pattern Recognition 31 (9) (1998) 1369-1390

[14] K. Tsuda, "Support vector classification with asymmetric kernel function". Proc. of 7th European symposium on Artificial Neural Networks, 1999, pp. 183-188.

[15] A. Androutsopoulos, J. Koutsias, K. V. Cbandrinos, and C. D. Spyropoulos. "An experimental comparison of naïve bayesian and keyword-based anti-spam filtering with personal e-mail messages". In Proc. ACM Int. Conf. on Research and Developments in Information Retrieval, pages 160–167,2000.

[16] H. Drucker, D. Wu, and V. N. Vapnik. "Support vector machines for spam categorization". IEEE Transaction on Neural Networks, 10(5):1048–1054, 1999.