Detecting Image Spam Using Principal Component Analysis & SVM Classifier

Minal Kamble
Research Scholar, Computer Science & Engineering
G.H. Raisoni College of Engineering,
Nagpur, India
minalkamble327@gmail.com

Dr. L. G. Malik
Professor, Computer Science & Engineering Department
G.H. Raisoni College of Engineering,
Nagpur, India
latesh.malik@raisoni.net

Abstract— **Today, the internet is the most powerful tools throughout the world. But the explosive growth of unsolicited emails has prompted the development of numerous spam filtering techniques Image spam is one of the most prevalent forms of spam ever since its inception. Spammers have developed new spamming techniques to use smaller, more colorful and photo quality images as spam. In spite of numerous efforts to build efficient spam filters against e-mail spam by researchers and free-mailing services like yahoo mail, Gmail etc spam filters still fail to arrest image spam. A set of ten features were identified based on observations and existing research in this area that can help in classification of image spam from photographs. In this paper PCA has been proposed for feature selection and SVM for the classification of spam images.**

Keywords- Spam; Image Spam; Feature Extraction; Feature Selection; classification

I. INTRODUCTION

With the increase in use of email for the communication , the number of unwanted 'spam' is also increasing[1]. For example, there's the occasional joke sent in mass from friend to friends and back again, or that all-important virus alert, or the occasional inspiration, etc[2].

Large amount of time is spent while detecting such messages[3].There is also cost related to server which manages the large amount of emails related to the system. When large number of messages are sent in bulk by the spammers, it badly affects the performance of the system. With the increase in unwanted messages users have to pay long distance connection charges. Most of the spammers send the emails in fraudulent way, by using the software which hides the identity[2]. Spammers uses various ways to get the email addresses of the users. They collect it from various companies and pay for that, acquire email addresses and sometimes they also hack the account.

Image Spam is an e-mail solicitation that uses graphical images of text to avoid filters[4]. Before one year , fewer than five out of 100 e-mails were image spam. Today, up to 40 percent are the image spam . It is expected to keep rising. The use of images in spam is well known, and has been going on for as long as it has been possible to send images in email messages.

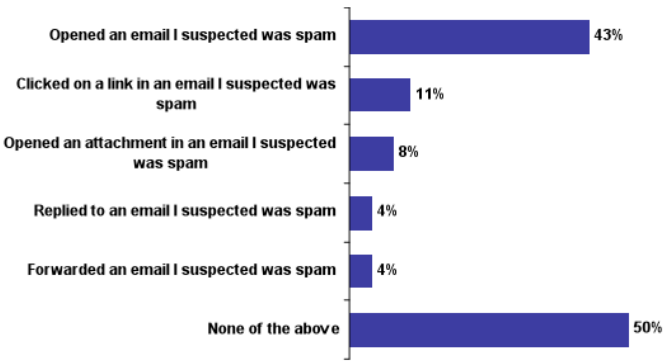

Fig. 1 Spam Survey

There are various aims of using images in email, from simply making the email more attractive, or adding a look of professionalism, to attempting to evade text based spam filters and signatures. The use of remote images in particular has been steadily increasing over the last 16 months.

Currently, the surest known countermeasure for image spam is to discard all messages containing images which do not appear to come from an already white listed E-mail address. However, this has the disadvantage that valid messages containing images from new correspondents must either be silently discarded, or that bogus "backscatter" bounce messages must necessarily be generated to the reply-to addresses in junk mail messages, enabling denial-of-service

attacks by spammers, as well as a directory harvesting attack. Another common technique for image spam detection is to analyze what percentage of the email is actually an image, as image spam often contains very little text content.

The remainder of this paper is organized as follows. Section II describe about image spam detection, section III explains about the related work, section IV deals with the overall system architecture , section V describes the expected result and section VI concludes the paper.

II. IMAGE SPAM DETECTION

Image spam is junk email that replaces text with images as means of fooling spam filters. If the recipient's email program downloads the image automatically, the image appears when the message is opened. The image itself may be a picture or drawing of alphanumeric characters that appears as text to the viewer, although it is processed as an image by the user's computer. The increase in more complex email spam attacks has caused spam capture rates across the email security industry to decline, resulting in wasted productivity and end-user frustration as more spam gets delivered to their inboxes. The root cause behind this sharp increase in spam volume is money. The more messages that are delivered to inboxes, the better the chances recipients take action on the messages, resulting in more income for spammers.

Fig. 2 Natural images

Fig. 3 Spam images

Various anti spam technologies are proposed in filtering text based spam emails which usually compare the contents of emails against specific keywords. The latest image spam is not possible to detect by the most anti spam software. Consequently, variety of methodology has been implemented in current anti spam system to filter the image spam.

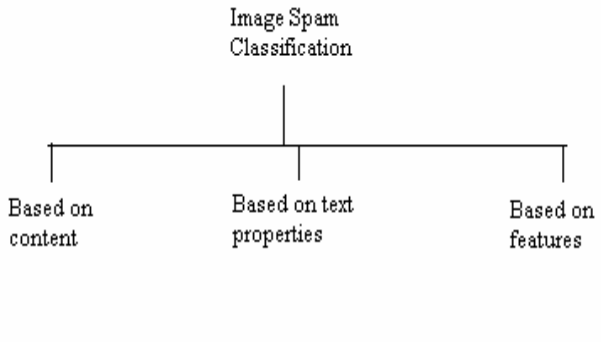

Fig. 4 Classification of image spam detection

III. RELATED WORK

Congfu Xu et. al [5] proposed approach based on Base64 encoding of image files and *n*-gram technique for feature extraction. It transformed normal images into Base64 presentation, and then it used *n*-gram technique to extract the feature. Using SVM, spam images were detected from legitimate images. This approach shows time efficient performance.

Tzong-Jye Liu et. al [6] came up with a three-layer image-spam filtering system by analyzing both the mail header and image. The first layer of the system deals with the mail header and the second and third layers analyze the high level feature and low–level feature of images.

Pattarapom Klangpraphant et. al [7] verity image with content-bases image retrieval. It also considers the partial similarity of e-mail spam from the normal e-mail.

Jen-Hao Hsia et. al [8] proposes method extracts topics in image to train classifier for detecting spam images, and achieves more accuracy than traditional filters. A detection cascade is provided to further reduce the overhead of the spam filter.

Cheng et al. [9] gives a framework called Binary Filtering with Multi-Label Classification (BFMLC) and considers both spam image filtering and user preferences into account. A file based on the BFMLC framework cannot only discriminate spam image from non spam images but also classifies spam image as several predefined topics.

IV. SYSTEM ARCHITECTURE

The objective of this paper is to develop a classifier that can differentiate legitimate from spam. The system consist of maximum likelihood classifier. Image spam can be classified based on their text properties, based on the content and based on the color histogram [3]. Here, we consider a global image feature for classification.

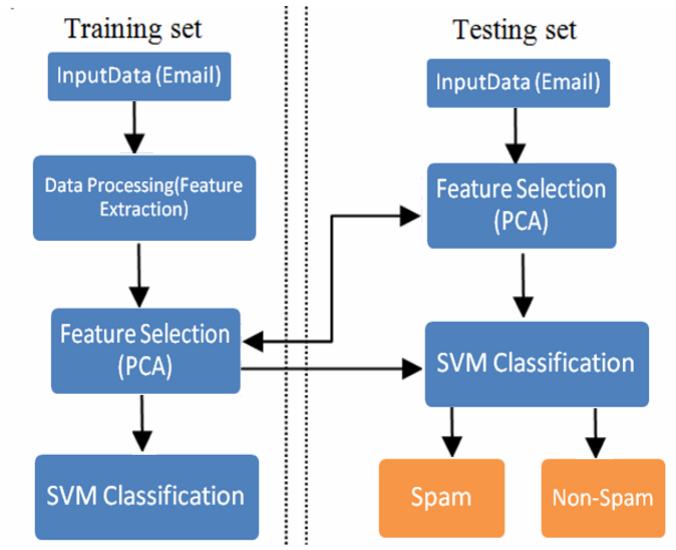

Fig. 5 System Architecture

A. *Image Data Set*

The spam archive images are taken from the website provided by Giorgio Fumera's Group. The natural images are taken from flickr's website. In total, the images consider to this proposed work is 1942 images with 1200 spam images and 742 ham images.

B. *Feature Extraction*

Ten features are taken into consideration for classification.[10] In image processing and photography, a color histogram is a representation of the distribution of colors in an image.

TABLE I. IMAGE FEATURES

Sr. No.	Feature	Values
1	Size	1
2	Height	1
3	Width	1
4	Area	1
5	Compression rate	1
6	Color number	1
7	Color variance	1
8	Consecutive color number	1
9	Main color coverage	1
10	Histogram of Oriented Gradient	80

The essential thought behind the Histogram of Oriented Gradient descriptors is that local object appearance and shape within an image can be described by the distribution of intensity gradients or edge directions.

In this paper 80 orientations of gradient in different directions have been considered. The reason for taking these values is that spammers use various techniques as shown in Fig. 6. So orientations are considered to defeat their efforts.

Fig. 6 Techniques used by spammers

The color range of the spam images are limited than that of the normal photos. Most of them are made on computer by spammers. Since transmission time and bandwidth are involved , all the listed features have relatively low values as compared to that of the natural scenes.

C. *Feature Selection Using PCA*

PCA is a technique of identifying patterns in data, and expressing the data in such a way as to highlight their similarities and differences. Since patterns in data can be hard to find in data of high dimension, where the luxury of graphical representation is not available, PCA is a powerful tool for analyzing data. The other main advantage of PCA is that once you have found these patterns in the data, and you compress the data, i.e. by reducing the number of dimensions, without much loss of information.[11]

The feature vectors formed then are used to compute the eigenvectors using Singular Value Decomposition. The input vectors to the reduced vector space is thus composed of the top ten eigenvectors. It is generally desirable to find or reduce the feature set to one that is minimal but sufficient.

PCA can be used to reduce the feature vector dimension while retaining most of the information by constructing a linear transformation matrix. The transformation matrix is made up of the most significant eigenvectors of the covariance matrix. The reduction in the dimensionality causes a reduction in the number of inputs to the support vector machine which makes it more efficient. In order to achieve these goals, PCA

computes new variables called principal components which are obtained as linear combinations of the original variables.

The use of PCA as a preprocessing step not only improves efficiency, but has also shown a consistent decrease in the error rate of the classifier. Out of 89 feature values, top 10 eigen values are selected.

D. SVM Classification

SVMs (Support Vector Machines) are a useful technique for data classification. An SVM classifies data by finding the best hyperplane that separates all data points of one class from those of the other class. The best hyperplane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyperplane that has no interior data points.

As with any supervised learning model, support vector machine is trained. The optimal feature values selected by PCA are gives as input the classifier. In total 100, 50 normal and 50 spam images are used for training.

E. Performance Measurement

The system performance is measured in terms of accuracy. The Accuracy tells the ratio of the number of spam which are identified accurately to the total number of images in the database.

The objective is to reduce the false positive rate of the classifier and to classify the images correctly into the actual class. The aim is to develop a classifier that can distinguish legitimate from spam. The idea is to develop a method to filter spam based on image content, rather than text content. Finally the focus is to reduce the false positive rate of the classifier i.e. if an image is spam, it should be detected as spam.

TABLE II. PERFORMANCE PARAMETER

Classifier	Natural Images	Spam Images
Natural Images	Number of images correctly classified as natural images	Number of spam images misclassified as natural image
Spam Images	Number of natural images misclassified as spam image	Number of images correctly classified as spam images

V. RESULT

We have measured the processing system using MATLAB. We have used all types of images which are collected randomly from spam archive data set. The classification can be done using support vector machine. The proposed method extracts the feature and then reduces them to train the classifier which distinguishes the spam and ham images. The classification accuracy of 98% was obtained.

VI. CONCLUSION

In this paper we have presented an image spam classification model by exploiting image feature. The spam images are growing continuously. They waste the storage on the network, also consumes the bandwidth. There is need for employing efficient method for differentiating spam and natural images.. In this paper we have reduced the number of features using Principal Component Anaysis. Finally classification of image spam is done using SVM classifier.

REFERENCES

[1]] Hrishikesh B. Ardhye, Gregory K. Myers, James A. Herson , (2005), "Image Analysis for Efficient Categorization of Image-based Spam E-mail"

[2] How do I stop spam?

[3] M. Soranamageswari, Dr. C. Meena, (2010), "An Efficient Feature Extraction Method for Classification of Image Spam Using Artificial Neural Networks" ,pp. 169-172

[4] Nobuo Kumagai Masayoshi Aritsugi, (2005) , "On Applying an Image Processing Technique to Detecting Spams"

[5] Congfu Xu, Yafang Chen, Kevin Chiew, (2010) "An Approach to Image Spam Filtering Based on Base64 Encoding and N-Gram Feature Extraction", pp. 171-177

[6] Tzong-Jye Liu, Wen-Liang Tsao, Chia-Lin Lee, (2010), "A High Performance Image-Spam Filtering System", pp. 445-459

[7] Pattarapom Klangpraphant, Pattarasinee Bhattarakosol, (2010), "Detect Image Spam with Content Base Information Retrieval", pp. 505-509

[8] Jen-Hao Hsia1 and Ming-Syan Chen, (2009), "LANGUAGE-MODEL-BASED DETECTION CASCADE FOR EFFICIENT CLASSIFICATION OF IMAGE-BASED SPAM E-MAIL"

[9] Hongrong Cheng, Zhiguang Qin, Chong Fu, and Y ong Wang, (2010), "A Novel Spam Image Filtering Framework with Multi -Label Classification",pp.282-285

[10] Feng Huamin, Yang Xinghua, Liu Biao, Jiang Chao,(2011), "A Spam Filtering Method Based on Multi-modal Features Fusion", pp.421-426

[11] Lindsay I Smith, "A tutorial on Principal Components Analysis", February 26, 2002

AUTHORS PROFILE

Minal Kamble received her Engineering degree in Computer Technology from Yeshvantrao Chavan College of Engineering, Nagpur, India in 2010, and currently pursuing for the MTech. degree in Computer Science and Engineering from G.H.Raisoni College of Engineering, Nagpur, India. She is a lecturer with Department of Computer Science and Engineering, GHRIETW, Nagpur University. Her research interests include digital signal processing, Data mining applications, artificial intelligence.