Detecting Image Spam Using Image Texture Features

Basheer Al-Duwairi*, Ismail Khater† and Omar Al-Jarrah†
Department of Network Engineering & Security
†Department of Computer Engineering
Jordan University of Science & Technology, Irbid, Jordan

Abstract

Filtering image email spam is considered to be a challenging problem because spammers keep modifying the images being used in their campaigns by employing different obfuscation techniques. Therefore, preventing text recognition using Optical Character Recognition (OCR) tools and imposing additional challenges in filtering such type of spam. In this paper, we propose an image spam filtering technique, called Image Texture Analysis-Based Image Spam Filtering (ITA-ISF), that makes use of low-level image features for image characterization. We evaluate the performance of several machine learning-based classifiers and compare their performance in filtering image spam based on low-level image texture features. These classifiers are: C4.5 Decision Tree (DT), Support Vector Machine (SVM), Multilayer Perception (MP), Naïve Bays (NB), Bayesian Network (BN), and Random Forest (RF). Our experimental studies based on two publicly available datasets show that the RF classifier outperforms all other classifiers with an average precision, recall, accuracy, and F-measure of 98.6%.

1. Introduction

Email spam, defined as unsolicited bulk email, continues to be a major problem in the Internet. With the spread of malware combined with the power of botnets, spammers are now able to launch large scale spam campaigns covering wide range of topics (e.g., pharmaceutical products, adult content, etc.) causing major traffic increase and leading to enormous economical loss. Recent studies [1], [2] revealed that spam traffic constitute more than 89% of Internet traffic. According to Symantec [3], in March 2011, the global Spam rate was 79.3%. According to the same report, spam accounted for approximately 52 billion email per day at the beginning of March and decreased to 33 billion email per day at the end of March. The cost of managing spam is huge compared to the cost of sending spam which is negligible. This cost includes the waste of network resources and network storage, the traffic and the congestion over the network, in addition to the waste in employees' productivity. It was estimated that an employee spends 10 minutes a day on average sorting through unsolicited messages [4].

Content-based email spam filtering represents one of the main approaches to combat spam. This approach involves digging into the content of email messages searching for certain signatures or specific patterns. Spammers are continuously adopting new techniques to evade detection. Image spam is one of these techniques that have gained a lot of popularity among spammers and that is being increasingly used in recent years. This type of spam began to appear in late 2005 and reached a peak of over 50% of spam emails from 2006 to 2007 [5]. In April, 2009 the amount of image spam was about 15-22% of all spam [6]. In this technique, spammers launch their campaigns through images attached to their emails instead of text based spam.

The main problem in dealing with such type of email spam is that spammers keep changing/modifying the images being used in their campaigns by employing different obfuscation techniques and randomly modifying them. Therefore, preventing text recognition by Optical Character Recognition (OCR) tools and imposing additional challenges in filtering such type of spam. This paper is an extended version of [7] which discusses the topic of filtering image spam as an important instance of content-based email spam filtering. Mainly, we propose an image spam filtering technique, called Image Texture Analysis-Based Image-Spam Filtering (ITA-ISF) that makes use of low-level image features for image characterization. We evaluate the performance of several machine learning-based classifiers and compare their performance in filtering email spam based on the features extracted from the images attached to emails. These classifiers are: C4.5 Decision Tree (DT), Support Vector Machine (SVM), Multilayer Perception (MP), Naïve Bays (NB), Bayesian Network (BN), and Random Forest(RF). The proposed method is evaluated experimentally based on real publicly available data sets. The rest of this paper is organized as follows: Section 2 discusses related work. Section 3 presents the proposed Image-spam-filtering technique. Section 4 presents performance evaluation. Finally, Section 5 concludes the paper.

2. Related work

Email spam filtering represents a major approach to combat spam. The goal of email spam filtering is to classify email messages into ham or spam. Typically, email spam filtering involves inspecting message content, header or both. Generally, in machine learning-based email spam filtering approach, a machine learning-based classifier is applied to certain features extracted from the email message in order to classify it as ham or spam. Machine learning-based spam filters may be further classified into two types [8], namely "Non content-based (Header-based) spam filtering", and "Content-based spam filtering". Content-based techniques inspect the body of an email searching for specific keyword(s) or features that are typically used by spammers or associated by certain spam campaign. Other techniques use pattern recognition to detect spam that follows certain behavior or pattern. Email body itself may be text, image, or both. Therefore, content-based filtering techniques usually deal with all these content types. The following subsections discuss the main approaches for image based spam filtering.

2.1. OCR-based techniques

The philosophy of OCR-based techniques is based on extracting the text embedded into attached images, then the same approaches used in spam filters to analyze emails' body text is used, which are keyword detection and text categorization techniques. The power of OCR-based techniques is determined by the OCR system itself. OCR errors is considered as one of the drawbacks of this kind of filters, especially when spammers obscure the content of the image by adding noise, dots, changing the background colors and rotating images, which affects the efficiency of OCR text extraction. This fact has led to other techniques based on low-level image features [9] and a combination of OCR with low-level image features [10], [11], [12]).

2.2. Techniques based on low-level image features

In these techniques, image classification is based on a set of low level features extracted from images. The classification process depends on the chosen features. For example, Wu et. al., [13] proposed a classification technique based on the presence of text features such as number of text regions, fraction of images with detected text regions, and the text area. L. Qiao et. al., [14] used corner and edge detection to characterize text area, and the color variance, the number of colors contained in the image, and the prevalent color coverage to characterize graphic properties of spam images. Low-level features such as color, shape and texture are used by [15], based on the fact that spam images often contain clearer and sharper objects than ham images.

A different approach based on image metadata was proposed in [16]. Image metadata and information include image width, height, aspect ratio, image area, image compression, image file extension and file size. Another technique is the near-duplicate detection technique. Spam images are often generated from a common template, and randomized to evade signature-based filters. Besides, the spam images are sent in batches to many users. Thus, images generated from the same template are visually similar (near-duplicate), these images can be recognized by a comparison with a known spam images stored in a database [17]. Table I provides a summary of the main image low-level features considered by these image spam filtering techniques. It also shows the main features that we consider in our work.

3. Image spam filtering based on image texture analysis

In this section, we discuss the main elements of the proposed Image Texture Analysis-Based Image Spam Filtering (ITA-ISF) technique. First, we discuss image texture features employed by this technique (Subsection 3.1). Then, we explain the methodology followed in our work (Subsection 3.2).

3.1. Texture Features

Image texture is a rich source of visual information [18]. Generally, textures are complex visual patterns composed of entities that have characteristic brightness, color, slope, size, etc. The main reason for choosing image texture features for image spam filtering is the fact that non-computer generated images have a different quality of texture as compared to textures in computer generated

Table 1. Image low-level features considered by different image spam filters

Image Spam Filter	Wu et. al., 2005 [13]	Liu et. al., 2010 [14]	Mehta et. al., 2008 [15]	Dredze. et. al., [16]	Our Approach (ITA-ISF)
Main Approach	Features related on the presence of text	Characteristics of text areas & graphic properties	Low-level graphic properties	Image metadata	Image texture analysis
Features used	# of detected text region, Fraction of images with text regions, Text area, Aspect ratio, Height & width	Corner and edge detection, Color variance, Prevalent color coverage, # of colors contained in an image, Color saturation	Color features, Shape features, Texture features	Image width, Image height, Aspect ratio, Image area, Image compression, Image file extension and size	Image histogram, Run-length matrix, Co-occurrence matrix, Image gradient, Autoregressive model, Wavelet transform

images. In our work, we use the following features which are considered to be among the most important features for texture analysis as pointed out in ([18], [19]):

- *Image Histogram:* is a graphical representation of the tonal distribution in a digital images (i.e., for each tonal value, it plots the number of pixels). The image can be described as a function *f(x,y)* of two space variables *x* and *y*, *x=0,1,...,N-1* and *y=0,1,...,M-1*. The function *f(x,y)* can take discrete values *i = 0,1,...,G-1*, where, *x*, *y* are the spatial coordinates (rows and columns), and the amplitude of *f* at any pair of coordinate (*x,y*). M and N can be any positive integers, for matrix representation an N × M image. *G* is the intensity levels in the image. The histogram *h(i)* is a function showing the number of pixels for each intensity level in the whole image, the normalized histogram *p(i)* whose entries are divided by the total number of pixels in the image. The histogram is an important feature of the image, for example, a narrow distributed histogram indicates a low contrast image. Many useful features can be computed from the histogram of the image, most often called central moments: histogram features have Mean, Variance, Skewness, Kurtosis, Energy and Entropy. The mean takes the average level of intensity of the image being examined, while the variance depicts the variation of intensity around the mean. If the histogram is symmetrical around the mean then the skewness is zero, otherwise, is positive or negative depending on whether it is skewed above or below the mean. The flatness of the histogram is measured by the kurtosis, and the entropy is the measure of histogram uniformity.
- *Image Gradient:* is a directional change in the intensity or color of an image.
- *Run-Length Matrix (RLM):* the run-length matrix p(i,j) is the number of runs with pixels of gray level *i* and run length *j*. Various texture features can be derived from RLM.
- *Co-Occurrence Matrix (COM):* is a matrix that is defined over an image to be the distribution of co-occurring values at a given offset.
- *Autoregressive Model (AR):* assumes a local interaction between pixels of the image in that the intensity is a weighted sum of neighboring pixel intensities.
- *Wavelet Transform:* in digital image processing, a Discrete Wavelet Transform (DWT) is used. DWT is any wavelet transform for which the wavelets are discretely sampled. It captures both frequency and location information (location in time) and considered as a key advantage over Fourier transform. The discrete wavelet transform (DWT) can also be defined as a linear transformation that operates on data vector whose length is an integer power of two, transforming it into a numerically different vector of the same length. It separates data into different frequency components, and then studies each component with a resolution matched to its scale.

It is to be noted that the second-order statistical features of texture analysis based on the co-occurrence matrix are the most popular features [20]. They were demonstrated to feature the capabilities for effective texture discrimination in biomedical images [18]. These features can be used in many applications that require discrimination between image areas that differ in texture characteristics. For example, image texture analysis can be used to analyze medical images to discriminate image areas that represent healthy tissues from that which represent pathological tissues. It is also important to mention that a variety of statistical image features are derived from image histogram, absolute gradient, run-length matrix, co-occurrence matrix, autoregressive model and wavelet analysis [18]. Some of these parameters are computed more than one time. For example, we compute run-length parameters four times (for vertical, horizontal,45^o and 135^o directions). The co-occurrence matrix parameters are computed twenty times *(for (d,0), (0,d), (d,d), (d,-d))*, where the distance d can take values of 1, 2, 3, 4, and 5. Overall, more than 270 features were extracted based on changing the parameters associated with each of the main features.

3.2. Methodology

The process of extracting features from the image attached to an email is depicted in Figure 1. This process consists of the following stages:

1. An email with different types of content (e.g., text, image) represents the input for the system.
2. Images attached to the email are extracted.
3. Image texture analysis processes is applied to the extracted image.
4. The image texture features vector is computed from the image texture parameters.
5. A pre-processing is performed on the computed features, this pre-processing include selection of the most informative features.
6. The selected image features are used to classify the image into the appropriate class (i.e., Ham or Spam). The classification process is done using a machine learning classifier.

The image texture analysis sub-block can be implemented using image processing and computer vision tools such as Matlab image processing toolbox, OpenCV, MaZda, etc. In our work, we used MaZda application [21], because it is specialized in image texture analysis. Initially, the extracted email image is entered to the system, then the region of interest (ROI) is chosen to study certain parts of the image for texture analysis. For image spam detection, it is required to include all the image and analyze its texture. Therefore, the ROI is superimposed on the required image in order to extract the needed texture features.

As an illustrative example, Figure 2 shows the numerical values of different image texture features obtained using Mazda application for two spam images (A and B) and two ham images (C and D), all taken from Dredze dataset [16]. In this example, we show the values of the features obtained after applying a feature selection algorithm as explained in Section 4. These features include:

- ATeta2:θ_2 of the Autoregressive (AR) model.

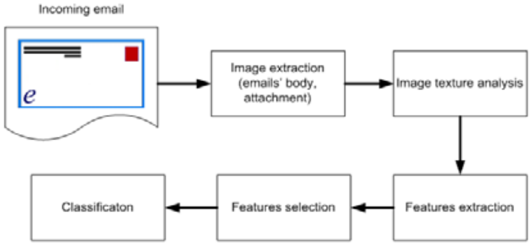

Figure 2. The process of extracting features of the image attached to an email

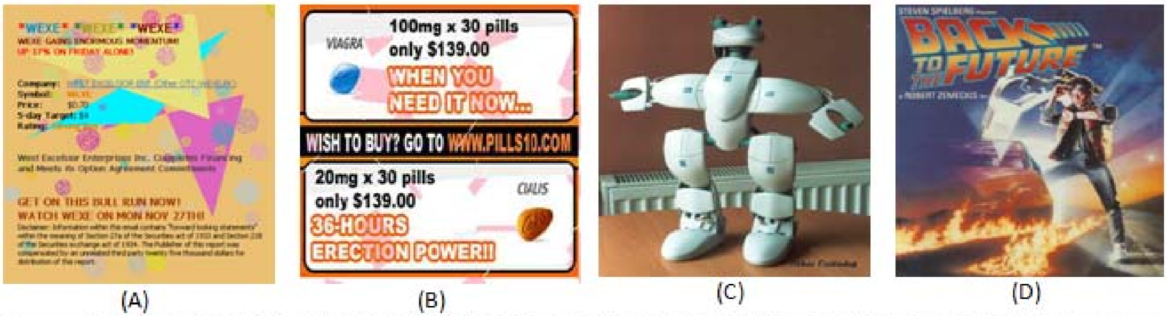

Image	ATeta2	ATeta1	RND6GLevNonU	RHD6RLNonUni	Z25D6AngScMom	CH5D6Correlat	CH4D6DifEntrp	C22D6DifVarnc	CH2D6SumEntrp	CH1D6Entropy	Perc99	Perc90	Perc01	Decision
A	-0.173894	0.407229	6075.6331	6015.4373	0.338929	0.348513	0.442207	146.60292	0.80601	0.858374	215	215	1	Spam
B	-0.717442	0.884905	2167.2362	26960.867	0.029418	0.670315	1.145363	310.04173	1.694521	2.233062	256	256	1	Spam
C	-0.428369	0.690574	3042.9522	49057.934	0.011479	0.862056	0.883437	24.659031	1.866816	2.264964	251	220	34	Ham
D	-0.431262	0.688793	2535.069	45222.699	0.003102	0.794949	1.03478	30.489383	1.895095	2.521677	220	195	42	Ham

Figure 1. The numerical values of different image texture features obtained using Mazda application [21] for two spam images (A and B) and two ham images (C and D) taken from Dredze dataset [16].

- This model is represented in Mazda by four model parameters vector $\theta = [\theta_1, \theta_2, \theta_3, \theta_4]$.
- ATeta1: θ_1 from the AR model.
- RND6GLevNonU: The gray level non-uniformity which is one of the Run length matrix-based parameters
- RHD6RLNonUni: The run length non-uniformity, which is one of the Run length matrix-based parameters.
- CZ5D6AngScMom: The angular second moment, which is one of the Co-occurrence matrix-derived parameters.
- CH5D6Correlat: The correlation, which is one of the Co-occurrence matrix-derived parameters.
- CH4D6DifEntrp: The difference entrop, which is one of the Co-occurrence matrix-derived parameters.
- CZ2D6DifVarnc: The difference variance, which is one of the Co-occurrence matrix-derived parameters.
- CH2D6SumEntrp: The sum entropy, which is one of the Co-occurrence matrix-derived parameters.
- CH1D6Entropy: The entropy, which is one of the Co-occurrence matrix-derived parameters.
- Perc99: it is a Histogram parameter, 1% – 99% option, which is the range between the brightness level at which image accumulated histogram is equal to 1% of its total to the level where the accumulated histogram is equal to 99% of its total (typically, different ROIs give different 1%- and 99%-levels).
- Perc90: the same as (Perc99) but with 90%
- Perc01: the same as (Perc99) but with 1%

The class label (Decision) is given for each image, because we are dealing with supervised learning, the class label is given from the dataset.

4. Performance evaluation

The performance of the proposed ITA-ISF technique has been evaluated experimentally based on real publicly available datasets as described in Subsection 4.2. Basically, our experiments involve evaluating the proposed scheme in terms of the performance metrics defined in Subsection 4.1 and for the following machine learning classifiers: C4.5 Decision Tree (DT), Support Vector Machine (SVM), Multilayer Perception (MP), Naïve Bays (NB), Bayesian Network (BN), and Random Forest (RF). Image datasets have been divided into a train and test sets according to the cross validation technique, where we used 10-fold cross validation. Weka tool [22] has been used for applying the machine learning techniques. Weka requires that the used features must conform to the input format of Weka. Therefore, the used features were ordered in a CSVfile in the following format:

feature 1, feature 2, , feature n, class label

By default the class labels are located at the end of each row. In our experiments, we have two class labels used to categorize the image in the email, a legitimate image is marked as *Ham*, while the spam image is marked as *Spam*.

4.1. Performance Metrics

We use the following standard performance metrics to evaluate the proposed technique: accuracy, precision, recall, F-measure, which are defined as follows:

- $\text{Accuracy} = \dfrac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$
- $\text{Precision} = \dfrac{\text{TP}}{\text{TP} + \text{FP}}$
- $\text{Recall} = \dfrac{\text{TP}}{\text{TP} + \text{FN}}$
- $\text{F} - \text{measure} = \dfrac{2.\text{Precision}.\text{Recall}}{\text{Precision} + \text{Recall}}$

Where *FP, FN, TP, TN* are defined as follows.

- *False Positive (FP):* The number of misclassified legitimate emails.
- *False Negative (FN):* The number of misclassified spam emails.
- *True Positive (TP):* The number of spam messages that are correctly classified.
- *True Negative (TN):* The number of legitimate emails that are correctly classified.

Precision is the percentage of correct prediction (for spam email), while spam Recall examines the probability of true positive examples being retrieved (completeness of the retrieval process), which means that there is no relation between precision and recall. On the other hand, F-measure combines these two metrics in one equation which can be interpreted as a weighted average of precision and recall. In addition, we use Receiver Operating Characteristics (ROC) curves which are commonly used to evaluate machine learning-based systems. These curves are basically a two-dimensional graphs where TP rate is plotted on y-axis and FP rate is plotted on x-axis.

Therefore, depicting the tradeoffs between benefits TP and costs FP. A common method to compare between classifiers is to calculate the Area Under ROC Curve (AUC). It is important to mention that our definition of the performance metrics is mainly based on the confusion matrix shown in Table 2.

Table 2. Confusion matrix representation

Prediction	Actual	
	Spam	Ham
Spam	TP	FN
Ham	FP	TN

4.2. Datasets

Our experiments are based on the following two publicly available recent datasets. It is to be mentioned that these datasets are usually used to evaluate Image spam filters (e.g., these datasets were used in the work conducted in, [16], [23]):

- Dredze data set [16]: this dataset contains only images that were extracted from both spam and legitimate emails. More specifically, this data contains 2021 ham images, and 3299 spam images. In our experiments, we preprocessed the dataset to exclude images tha does not provide enough information such as icon images, and images with sizes of tens of bytes which are either blank images, or images with no texture information. This resulted in a total of 1770 ham images and 3209 spam images.
- Image Spam Hunter ISH data set [23]: ISH dataset contains only images as well. In this dataset collection, there are 810 ham images that are randomly collected and downloaded from Flicker.com, and 926 spam images collected from real spam emails.

It is important to mention that these datasets were used for training and testing. In the following subsections, we present the results obtained without applying feature selection (Subsection 4.3) and with applying feature selection (Subsection 4.4) on both datasets.

4.3. Results without applying feature selection

In this subsection, we discuss the results obtained by the machine learning classifiers when applied on all extracted image features (about 270 features). Figures 3 and 4 depict the performance of the classifiers applied to the features extracted from Dredze and ISH datasets, respectively. It can be seen that the RF classifier outperforms all other classifiers with precision, recall, F-measure, accuracy, and ROC area of more than 0.98 for both datasets. This classifier achieved the lowest false positives of 0.014 in Dredze dataset and 0.013 in the ISH dataset. It is important to point out that the misclassification rate of ham emails was 0.006 which is considered to be very low. Typically, it is very challenging to obtain such results for image spam detection.

It is to be mentioned that the accuracy of an SVM model is largely dependent on the selection of its kernel parameters. In this paper, we evaluate the SVM model using the radial basis kernel with γ = 0.0. There are many difficulties in applying SVM, when we tried to apply it on the features extracted (about 270 features, and 4979 email images) from Dredze dataset, the required memory (Weka heap size) was not enough, so, we only applied it on the features extracted from ISH (about 270 features, and 1736 email images), and we obtained the results shown in Figure 3. Based on these results, it is clear that SVM does not perform well in this experiment. We also point out that we exclude the Multilayer Perceptron (MP) classifier from this study as it took a very long time without yielding any result. Which means the training phase was too long. However, we include it in the next study after applying feature selection algorithm.

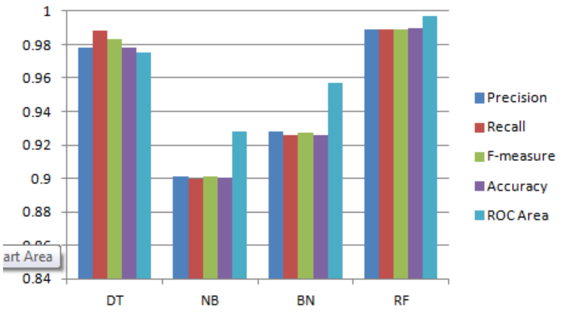

Figure 3. Plot of performance measures for different machine learning techniques applied on Dredze dataset - without feature selection

4.4. Results with feature selection

It is clear from the results of Subsection 4.3 that using large feature space does not necessarily provide good results because many features carry the same information and there is a correlation between features. In this subsection, we show the results obtained after applying the Principal Components Analysis (PCA) feature selection algorithm. The resulting feature space consists of only 8 features that represent the most important features from the original feature space.

Figure 5 depicts the performance of the classifiers applied to the features extracted from Dredze dataset. We also show the performance of SVM for different values the parameter (the radial basis kernel of the SVM classifier). It can be seen that both the RF classifier outperforms all other classifiers with precision, recall, F-measure, accuracy, and ROC Area of 0.986, 0.986, 0.986, 0.985, and 0.994 respectively. It is to be noted that the performance of SVM was very close to that of RF classifier, and it did not vary much for different values of γ. The ROC curves for these classifiers are shown in Figure 6. Based on the area under ROC for the two experiments (i.e., using ISH and Dredze datasets), it can be seen that the RF classifier has the best performance.

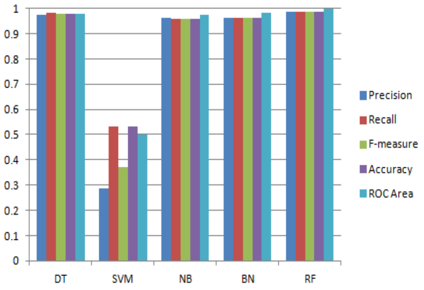

Figure 4. Plot of performance measures for different machine learning techniques applied on ISH dataset- without feature selection

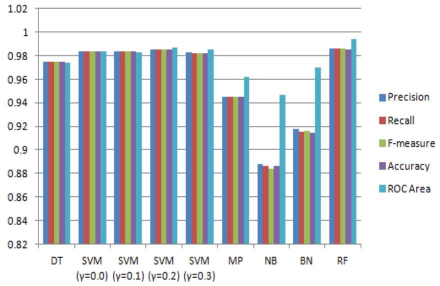

Figure 5. Performance of machine learning classifiers-Dredze dataset- with feature selection

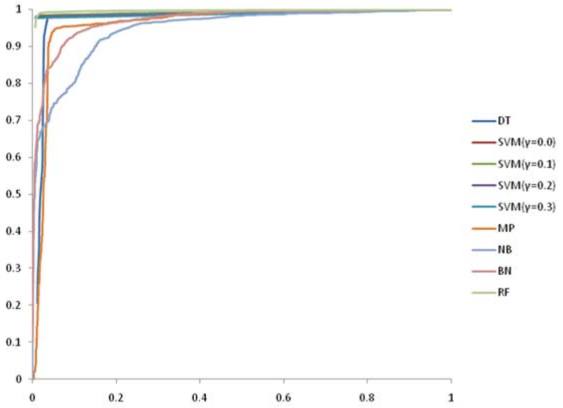

Figure 6. ROC curves for the classifiers applied on Dredze dataset after features selection with PCA (x-axis is FP rate, y-axis is TP rate)

Figure 7 depicts the performance of the classifiers applied to the features extracted from ISH dataset. We also show the performance of SVM for different values the parameter (the radial basis kernel of the SVM classifier). It can be seen that both the RF classifier and the SVM classifier (with γ = 0.1) performs very well. RF classifier achieved precision, recall, F-measure, accuracy, and ROC Area of 0:981, 0.981, 0.9810, and 0.995 respectively. While the same metrics for SVM classifier (with γ = 0.1) were as follows: 0.986, 0.986, 0.986, 0.9856, and 0.986. It is also obvious that as we increase the value of, the overall performance of SVM classifier decreases, but with a very low false positive. This means that the value of could be adjusted to obtain the increase or decrease *FP* while maintaining a good performance for this classifier. The ROC curves for these classifiers are shown in Figure 8.

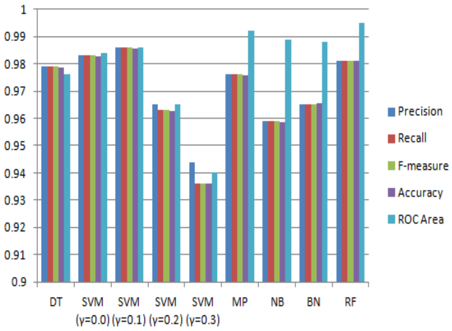

Figure 7. Performance of machine learning classifiers-ISH dataset- with feature selection

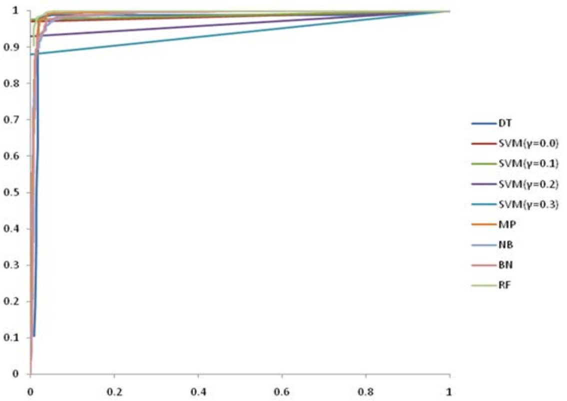

Figure 8. ROC curves for the classifiers applied on ISH dataset after features selection with PCA (x-axis is FP rate, y-axis is TP rate)

4.5. Comparison with Previous Work

In this subsection, we compare the performance of the proposed scheme with other image spam filtering techniques based on the results reported in the literature for these techniques. Table 3 shows the best performance obtained by four image spam filters and compare it to the results obtained using ITA-ISF. Also, a summary of the datasets used by each filter is provided.

It can be seen that our approach maintains a low false positive rate while achieving high classification accuracy and it performs very well compared to these image spam filters.

Table 3. Performance of ITA-ISF compared to other image spam filters. A: Accuracy, P: Precision, R: Recall, F: F-measure

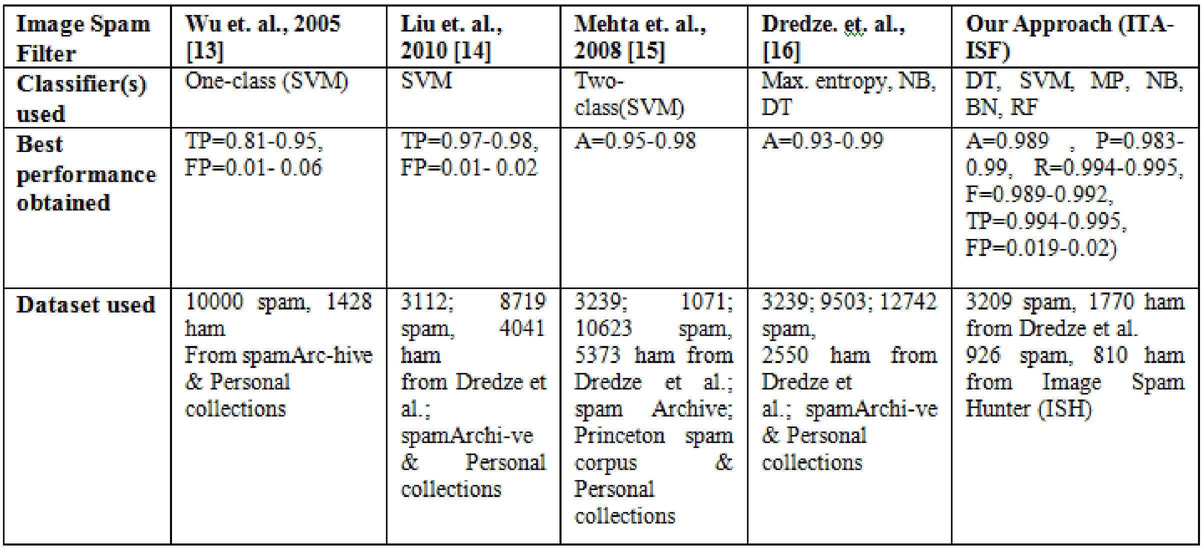

Image Spam Filter	Wu et. al., 2005 [13]	Liu et. al., 2010 [14]	Mehta et. al., 2008 [15]	Dredze. et. al., [16]	Our Approach (ITA-ISF)
Classifier(s) used	One-class (SVM)	SVM	Two-class(SVM)	Max. entropy, NB, DT	DT, SVM, MP, NB, BN, RF
Best performance obtained	TP=0.81-0.95, FP=0.01- 0.06	TP=0.97-0.98, FP=0.01- 0.02	A=0.95-0.98	A=0.93-0.99	A=0.989 , P=0.983-0.99, R=0.994-0.995, F=0.989-0.992, TP=0.994-0.995, FP=0.019-0.02)
Dataset used	10000 spam, 1428 ham From spamArc-hive & Personal collections	3112; 8719 spam, 4041 ham from Dredze et al.; spamArchi-ve & Personal collections	3239; 1071; 10623 spam, 5373 ham from Dredze et al.; spam Archive; Princeton spam corpus & Personal collections	3239; 9503; 12742 spam, 2550 ham from Dredze et al.; spamArchi-ve & Personal collections	3209 spam, 1770 ham from Dredze et al. 926 spam, 810 ham from Image Spam Hunter (ISH)

The results reported in [13] show that image spam detection rate ranges from 81.4% to 95% when applying one class SVM on the visual features extracted from image spam with outlier in the anti-spam filter of 20% and 5%, respectively. However, since the false positive rate increases by increasing the outlier of the anti-spam filter, the authors adopted the results that correspond to the 20% outlier (i.e., 81.4% spam detection rate). In addition, [13] reports the results obtained for the two class SVM classifier as well, showing that it is not suitable in practice because of its high false positive rate of 13.97%.

The image spam detection results obtained in [14] based on applying the L2-loss linear SVM and the non-linear SVM with Gaussian RBF kernel, respectively, for different types of datasets that represent a combination of spam archives and personal ham images, are comparable to the results obtained by the proposed spam filter where the spam detection accuracy ranges from 97% to 98% with false positive rates between 1% to 2%.

The visual features based and near duplicated detection approach proposed in [15] reports a prediction accuracy of over 95% for different datasets. Dredze et al. [16] showed that image spam can be characterized by a small feature set, which included metadata properties of the image file and some other low-level features directly drawn from the digitized image, such as file format, size, edge and ten randomly-generated color pixels. Although, the accuracy of the spam filter as reported in [16] reached high values of 96% and 99%, it remains unclear whether using features extracted from image metadata would be efficient in cases where spammers are able to evade detection by compromising image information (e.g., keep changing image background colors).

5. Conclusion

This paper presented an image spam filtering based on image texture analysis. The proposed technique called Image Texture Analysis-Based Spam Filtering (ITA-ISF), extracts features related to the histogram, gradient, run-length matrix, co-occurrence matrix, autoregressive model, and wavelet transform of an image and applies a feature selection algorithm to reduce the feature space while keeping the most informative features. The performance of C4.5 Decision Tree (DT), Support Vector Machine (SVM), Multilayer Perception (MP), Naïve Bays (NB), Bayesian Network (BN), and Random Forest (RF) machine learning classifiers were applied on the low-level image texture features of two publicly available datasets. Performance evaluation of the proposed image spam filter shows that the RF classifier outperforms all the other classifiers with an average precision, recall, F-measure and accuracy of 98.6%. It is also important to mention that the SVM classifier performed very well and that the number of false positives can be minimized by adjusting the kernel parameter of this classifier.

7. References

[1] C. Kreibich, C. Kanich, K. Levchenko, B. Enright, G. Voelker, V. Paxson, and S. Savage, "Spamcraft: An Inside Look At Spam Campaign Orchestration," Proceedings of the Second USENIX Workshop on Large Scale Exploits and Emergent Threats (LEET '09), Boston, Massachusetts, April 2009.

[2] M. Intelligence, "MessageLabs Intelligence: 2010 Annual Security Report," 2010.

[3] Symantec. March 2011 Intelligence Report. Available at: http://www.symantec.com/about/news/release/article.jspprid=20110329 01

[4] S. Hinde, "Spam, scams, chains, hoaxes and other junk mail" Computers & Security, vol. 21, pp. 592 - 606, 2002.

[5] M. Security. March, 2008, Whitepaper - The Rise and Fall of Image Spam Available at: http://www.m86security.com/newsimages/trace/RiseandFalofImageSpam March08.pdf

[6] I. X. F. Report. May, 2009, Image spam - reborn and trying to rejuvenate YOUR health! Available at: http://blogs.iss.net/archive/image-spam-rebirth.html.

[7] B. Al-Duwairi, I. Khater, O. Al-Jarrah, "Texture Analysis-Based Image Spam Filtering," in Proc. of International Conference of Internet Technology and Secured Transactions, ICITST 2011, Abu Dhabi, pp. 288-293.

[8] B. Agrawal, N. Kumar, and M. Molle, "Controlling spam Emails at the Routers," in Communications, 2005. ICC 2005. 2005 IEEE International Conference on, 2005, pp. 1588-1592 Vol. 3.

[9] B. Biggio, G. Fumera, I. Pillai, and F. Roli, "Image Spam Filtering Content Obscuring Detection," In Proceeding Fourth Conference on Email and Anti-Spam (CEAS 2007), Mountain View, California, 2007.

[10] G. Fumera, I. Pillai, F. Roli, and B. Biggio, "Image Spam Filtering Using Textual and Visual Information," In proceeding the MIT Spam Conference 2007, Cambridge, MA, USA 2007.

[11] P. Klangpraphant and P. Bhattarakosol, "PIMSI: A Partial Image Spam Inspector," in Future Information Technology (FutureTech), 2010 5th International Conference on, 2010, pp. 1-6.

[12] F. Gargiulo and C. Sansone, "Combining Visual and Textual Features Filtering Spam Emails," in Pattern Recognition, 2008. 19th International Conference on, 2008, pp. 1-4.

[13] C. Wu, K. T. Cheng, Q. Zhu, and Y. L. Wu, "Using Visual Features for Anti-Spam Filtering," in Image Processing, 2005. ICIP 2005. IEEE International Conference on, 2005, pp. 509-12.

[14] Q. Liu, Z. Qin, H. Cheng, and M. Wan, "Efficient Modeling of Spam Images," in Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on, 2010, pp. 663-666.

[15] B. Mehta, S. Nangia, M. Gupta, and W. Nejdl, "Detecting Image Spam Using Visual Features and Near Duplicate Detection," In Proceeding of the 17th international conference on World Wide Web, Beijing, China, 2008.

[16] R. G. Mark Dredze andA. Elias-Bachrach,"Learning Fast Classifiers for Image Spam," presented at the in Proc. CEAS 2007, Mountain View, California, August 2-3, 2007.

[17] W. J. Zhe Wang, Qin Lv, M. Charikar, and Kai Li, "Filtering Image Spam with Near-Duplicate Detection," In Proceedings of the Fourth Conference on Email and AntiSpam, CEAS'2007, 2007.

[18] M. S. Andrzej Materka "Texture Analysis Methods – A Review," Institute of Electronics, Technical University of Lodz, Brussels 1998.

[19] P. S. Andrzej Materka "MaZda User's Manual," Instytut Elektroniki Politechnika Lodzka, Lodz, Poland1998-2005.

[20] R. M. Haralick, "Statistical and structural approaches to texture," Proceedings of the IEEE, vol. 67, pp. 786-804, 1979.

[21] P.M. Szczypinski, A. Materka, and A. Klepaczko,

"MaZda A Software Package for Image Texture Analysis," Computer Methods and Programs in, vol. 94, pp. 66-76, 2009.

[22] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. "The WEKA Data Mining Software: An Update. SIGKDD Explorations", 2009.

[23] Y. Gao, M. Yang and X. Zhao,"Image Spam Hunter," in Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, 2008, pp. 1765, 1768.