10-2015

On Robust Image Spam Filtering via Comprehensive Visual Modeling

Jialie SHEN
Singapore Management University, jlshen@smu.edu.sg

DENG, Robert H.
Singapore Management University, robertdeng@smu.edu.sg

Zhiyong CHENG
Singapore Management University, zy.cheng.2011@phdis.smu.edu.sg

Liqiang NIE
National University of Singapore

Shuicheng YAN
National University of Singapore

DOI: https://doi.org/10.1016/j.patcog.2015.02.027

Citation

SHEN, Jialie; DENG, Robert H.; CHENG, Zhiyong; NIE, Liqiang; and YAN, Shuicheng. On Robust Image Spam Filtering via Comprehensive Visual Modeling. (2015). *Pattern Recognition*. 48, (10), 3227-3238. Research Collection School Of Information Systems.
Available at: https://ink.library.smu.edu.sg/sis_research/3196

Published in Pattern Recognition, 2015 Oct, 48 (10), pp. 3227-3238.
http://doi.org/10.1016/j.patcog.2015.02.027

On robust image spam filtering via comprehensive visual modeling

Jialie Shen [a], Robert H. Deng [a], Zhiyong Cheng [a], Liqiang Nie [b], Shuicheng Yan [c]

[a] *School of Information Systems, Singapore Management University, Singapore*
[b] *School of Computing, National University of Singapore, Singapore*
[c] *Department of ECE, National University of Singapore, Singapore*

Keywords:
Algorithm
Security
Experimentation
Spam

ABSTRACT

The Internet has brought about fundamental changes in the way peoples generate and exchange media information. Over the last decade, unsolicited message images (image spams) have become one of the most serious problems for Internet service providers (ISPs), business firms and general end users. In this paper, we report a novel system called RoBoTs (*Ro*bust *Boos*T*rap* based *s*pam detector) to support accurate and robust image spam filtering. The system is developed based on multiple visual properties extracted from different levels of granularity, aiming to capture more discriminative contents for effective spam image identification. In addition, a resampling based learning framework is developed to effectively integrate random forest and linear discriminative analysis (LDA) to generate comprehensive signature of spam images. It can facilitate more accurate and robust spam classification process with very limited amount of initial training examples. Using three public available test collections, the proposed system is empirically compared with the state-of-the-art techniques. Our results demonstrate its significantly higher performance from different perspectives.

1. Introduction

The Internet has brought about a big change in the way people generate and exchange media information. Over the last decade, unsolicited bulk electronic mails (spam e-mail) in different formats have become one of the most serious problems for Internet service providers (ISP), business firms and general end users [1,2]. Image spam refers to the e-mail spam presenting major content of the message as a picture. Since the most modern e-mail client softwares present the image messages to user directly by default, image spam becomes a highly effective way to break protection based on existing filtering scheme [1]. In recent years, rapidly increasing volume of image spams has been witnessed. Recent statistics show that image spams account for 27% of total number of email spams [3] and image spams rose to 55% spam e-mails in 2010 [4]. Consequently, there is an urgent need to design and evaluate advanced techniques for anti-spam filtering.

Developing effective spam detection system and corresponding algorithms has been recently the focus of much attention. The previous efforts can be generally classified into two widely accepted and popular streams: (1) IP-based blocking and (2) content-based detection. The basic idea for IP-based blocking approaches is to create/keep blacklists of IP addresses of possible spam clients and those information can be helpful in determining whether to block account of an email sender. The related schemes could be policy based or "reactive" based. Unfortunately, a piece of comprehensive blacklist is extremely hard to be automatically maintained in many real circumstances. Main reason is that existing blacklists could be generated using non-permanent Internet identifiers. Especially, the hosts in mobile environment could introduce more dynamism and make the case even worse. Comparing to natural images, image spams enjoy a lot of unique characteristics. As shown in Fig. 1, image spams embed a lot of text contents. Also, colour contrast of nature images is much smoother and their contents can be approximated more accurately using artificial distribution functions. In fact, this nice property enables ham message detection which could use nature pictures as attachment. Additionally, spam images can be generated using a certain templates, and thus contain similar colour and texture patterns. One of the most popular methods is to combine a set of basic patterns and automatically generate huge amount of near-duplicate messages. All the observations suggest the feasibility of image spam identification based on the proper combination of multiple visual features. In this research, our main focus is to design and test robust content-based scheme for filtering image spam (Fig. 1).

Similar to many classic visual classification problems (e.g., scene classification [5], image annotation [6] and visual object detection [7]), spam image filtering can be modelled as the binary classification. The ultimate goal is to categorize input images into two classes: ham and spam. The process can be divided into two main steps: feature extraction and classification. In the first step,

E-mail addresses: jlshen@smu.edu.sg (J. Shen),
robertdeng@smu.edu.sg (R.H. Deng), zy.cheng.2011@phdis.smu.edu.s (Z. Cheng),
nieliqiang@gmail.com (L. Nie), eleyans@nus.edu.sg (S. Yan).

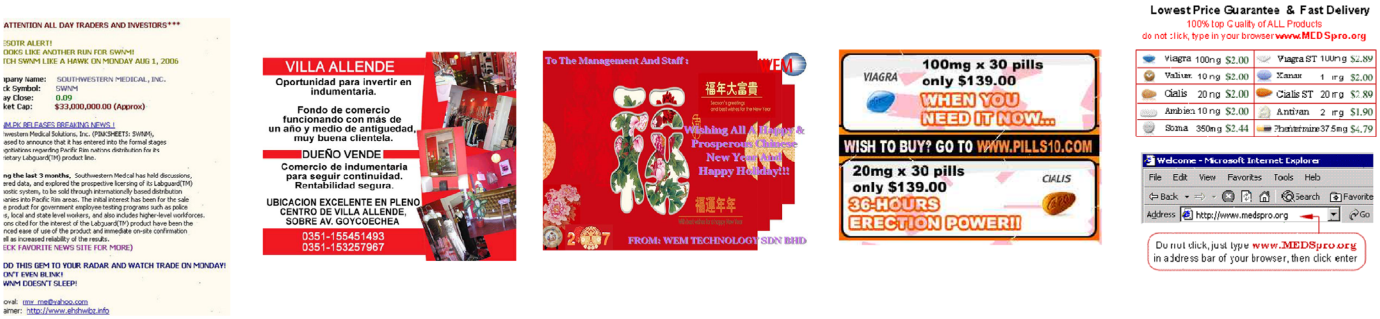
ATTENTION ALL DAY TRADERS AND INVESTORS***
VILLA ALLENDE
Oportunidad para invertir en indumentaria.
DUEÑO VENDE
0351-155451493
0351-153257967
To The Management And Staff:
Wishing All A Happy & Prosperous Chinese New Year And Happy Holiday!!!
FROM: WEM TECHNOLOGY SDN BHD
VIAGRA
100mg x 30 pills only $139.00
WHEN YOU NEED IT NOW...
WISH TO BUY? GO TO WWW.PILLS10.COM
20mg x 30 pills only $139.00
CIALIS
36-HOURS ERECTION POWER!!
Lowest Price Guarantee & Fast Delivery
100% top Quality of ALL Products
Welcome - Microsoft Internet Explorer
http://www.medspro.org
Do not click, just type www.MEDSpro.org in address bar of your browser, then click enter

Fig. 1. A few examples of image spams shows different visual patterns comparing to nature images and a lot of text contents embedded.

advanced visual representation needs to be extracted to model spam effectively. The second step is to apply a computational scheme (statistical model or machine learning classifier) to estimate labels of input messages based on their features. To achieve accurate classification, two outstanding concerns in modelling spam images and designing intelligent detection schemes need to be addressed properly:

1. How to model characteristics of spam image(s) effectively? To achieve robust and effective filtering, corresponding visual modelling scheme is essential to capture salient features of spams in a concise way. The features used for classifying spam and non-spam images are selected based on the observation of real spam images. The principle is that the selected features should be able to accurately discriminate the spam and ham images. Inspired by recent success of multimodal and cross-modal based visual analysis [8–10], different kinds of features have been used in existing image spam filters, including *email header features*, *image metadata*, *text-based features*, and *visual-based features*. Few image spam filters use only one type of other features besides the visual feature, because image spammers can easily develop tricks to fool these filters. Since the contents of image spams could be very complex and diverse, using single type of visual feature could be very hard to achieve satisfactory modelling performance. Moreover, discriminative characteristics of image spams could be found at different resolution levels. Thus, it would be desirable to develop a composite scheme to fuse multiple features from various levels of granularity.
2. How to develop robust detection system with small amounts of training examples? In many real applications, labelled training examples are very scarce and expensive to be acquired. This is because learning example annotation could be very time consuming task and requires huge amount of knowledge from domain experts. There are two ways to address the issue: (1) advanced visual classification scheme which requires small amount of training examples and (2) effective automatic scheme to generate high quality training set based on a small amount of seed.

Motivated by above, this paper reports a novel detection system called the RoBoTs (Robust BoosTrap based spam detector) to facilitate accurate and efficient image spam classification. It not only can effectively integrate multiple kinds of visual information at different resolution levels but also apply resampling scheme to address problem of small number of training examples. In brief, the contributions of the paper are as follows:

- Distinguished from previous approaches, we develop a comprehensive scheme to model visual contents of spam image. The approach is based on multiple kinds of features extracted at multiple resolution levels. It can provide a significant improvement on the quality of image modelling. To the best of our knowledge, very few work focused on developing multi-resolution based visual feature combination scheme for image spam filtering. Based on the comprehensive literature review presented in Section 2, no similar algorithm has been reported previously.
- We develop a novel detection framework naturally integrating bootstrapping resampling, Linear Discriminative Analysis (LDA), and Random forests [11]. Its architecture consist of two major components: the first one is a bootstrap based scheme for actively selecting high quality training examples. The cost sensitive scheme can greatly reduce size of initial training examples. Moreover, since performance of classifier plays a very critical role in determining final categorization accuracy, an effective ensemble method called Linear Discriminant Forests (LDFs) is designed to seamlessly combine LDA [12,13] and Random forests [11]. The key novelty of LDF is to apply feature selection over subsets of raw features and try to reconstruct a more comprehensive feature combination for superior classification performance via projection.
- To demonstrate the superiority of RoBoTs system, we have fully implemented the scheme and compared it with three other state-of-the-art systems. The empirical study has been carried out using three large spam image test collections. The results show that our approach achieves substantial performance improvement on spam detection in terms of effectiveness. In addition, RoBoTs demonstrates strong robustness under various training environments.

The rest of the paper is organized as follows: Section 2 gives a comprehensive review of existing work about spam image detection and analyses major characteristics (advantages and disadvantages) of the methods. Next, a short review about bootstrap resampling is given in Section 3. In Section 4, we introduce the details about proposed system architecture and core algorithms. In Section 5, the experimental configuration is present and after that, Section 6 presents the experimental results and the performance analysis to demonstrate the advantages of the proposed method. Finally, Section 7 concludes the research study and presents the major research findings, followed by a detail discussion on the possible extensions of the current study.

2. Related work

Along with the fast growth of image spams, researchers started to develop image spam techniques to fight with image spammers over decades ago, and since then there has been a steady development of the approach. Early work like SpamAssassin [14] and Fumera et al. [15] applies Optical Character Recognition (OCR) techniques to extract the texts embedded in the images, and then uses text-based spam filtering techniques. However, OCR requires high computational cost. Besides, high accuracy OCR by itself is a difficult problem especially when spammers are obfuscating the text in images, such as using different sizes and irregular fonts, rotating texts, and adding random colors [16,17]. Due to these reasons, recent works focus on directly classify email images into spam or non-spam using *near-duplicate detection methods* and *classification methods* by exploiting low-level features [18].

2.1. Near-duplicate detection based approach

Near-duplicate detection method is based on the assumption that spam images are derived from a common template by randomized alterations for circumventing signature-based detection, and are sent to many users in batches. Accordingly, spam images from the same template are visually similar. When processing a candidate image, the similarity between the images with each template in the database is computed separately, and then compared to a pre-defined threshold to determine whether it is spam or non-spam.

Several near-duplicate detection based image spam filters have been developed. Wang et al. [19] developed a system which first detect some image-based spam messages via traditional anti-spam methods, and then detect variations of those known spam images with fast near-duplicate detection filters based on low-level visual features (color histogram, Haar wavelet transform and edge orientation histograms). Mehta et al. [20] used Gaussian Mixture Models to describe spam image content and applied Agglomerative Information Bottleneck (AIB) principle to quantify difference between those image GMMs for the purpose of clustering. He et al. [21] first judged the similarity between the input and template images with file properties (e.g., file size, image dimensions, bit depth and aspect ratio), and then made another comparison based on gray-level or color image histogram using measures like histogram intersection and Euclidean distance, if necessary. Qu et al. [22] computed image similarity based on color moments, texture, and shape features, as well as an additional feature produced by a two-class SVM classifier trained on spam and ham images. The method in [23] first categorizes the images into illustrated images and text mainly images based on the foreground illustration objects of the images, and then clusters images into different categories based on different features (e.g., color and/or foreground layout for illustrated images, and text layout and/or background features for text mainly images), using an unsupervised ranked clustering algorithm.

The main disadvantage of near-duplicate detection methods is poor scalability. In general, they cannot detect new kinds of image spam, since they detect spam images by evaluating the similarity between input images with known spam images. In contrast, image classification methods are capable to generalize to new image spams because of the use of machine learning algorithms. Thus, the majority of existing image spam filters is based on image classification techniques.

2.2. Image classification based approach

The image classification for spam detection is to train a classifier on the feature vector representations of a set of spam and legitimate images. Many classification based image spam filters have been reported in literatures. Due to the space limitation, instead of describing each work, we summarize the used features and classification algorithms, which are the main characteristics of classification methods. For detail review, refer to [18,24,25].

The features used for classifying spam and non-spam images are selected based on the observation of real spam images. The principle is that the selected features should be able to best discriminate the spam and ham images. Different kinds of features have been used in existing image spam filters, including *email header features*, *image metadata*, *text-based features*, and *visual-based features*. Few image spam filters use only one type of other features besides the visual feature, because image spammers can easily develop tricks to fool these filters. Krasser et al. [26] and Uemura et al. [27] used only few image metadata (e.g., file name, file size, image dimensions, image file type, file size, etc.) for fast image spam detection. Other filters use either visual features or the combination of several types of features.

A large proportion of existing works uses only low-level visual features. The commonly used low-lever features are color features (e.g., color moments, color heterogeneity, grey histogram, number of colors, color saturation) [28,17,20,29], edge-based/shape features [20,30], texture features [31,20,32], as well as local variant features [33] and visual-of-bag-words [34]. Other works use the combination of image metadata and visual features [35,36,32], or text-based features and visual features [31,37,38,30]. The text-based features include text area features (e.g., the number of detected text regions, the fraction of images with detected text regions [31], corner and edge detection to characterize the text areas [30], etc.), and the text features based on the output of OCR (e.g., the number of words, length of text [38], etc.). The head features have been used in [36] with the combination of image metadata and visual features. A set of header features (such as precedence, list-help, sender, etc.) is used for the first step filtering in this work.

Among the existing classification based image spam approaches, the most popular classification methods are SVM [31,37,26,20,34,33] and decision trees [26,35,38,29]. Other types of classifiers are also used, like Probability boosting tree [17], Naïve Bayes [35], Bayesian filter [27], maximal figure-of-merit learning algorithm [28], ANN [16], and maximum entropy [35]. Among the investigated works, only [28,39] use a multi-class approach, others use either two-class classifiers or one-class classifiers (trained on spam images only).

The training of an effective classifier needs a large set of high quality labeled data, which requires lots of human labors and time to obtain. Few works study the problem of how to leverage small training data to learn an effective classifier. Gao et al. [40] proposed a semi-supervised approach, called regularized discriminant EM algorithm, (RDEM) to detect image spam e-mails. This method uses small amount of labeled data and large amount of unlabeled data to learn the model and identify spams simultaneously. In more recently, they use active learning to guide the users to label as few images as possible to maximize the classification accuracy in the client-side of a comprehensive image spam detection system, which consider both server-side filtering and client-side detection [41].

In this paper, we investigate the capability of the combination of multiple low-level visual features from multiple resolution levels on classifying spam and non-spam images. Besides, a bootstrap based scheme is developed to actively select high quality training examples, so as to reduce the requirement of large amount of initial training examples in the training of an effective classifier.

Table 1 summarizes and compares basic characteristics of different state-of-the-art approaches. FISND, VF-SVM and GMM-AIB denote the methods published in [19,20]. As the discussion given above, none of the above approaches use multiresolution and multifeature based method to characterize contents of image spam. Additionally, the classification supported by the schemes is not cost sensitive. It implies that if the number of training samples is limited, good performance might not be always guaranteed.

3. Bootstrap resampling

Resampling technique is a powerful scheme in many estimation or evaluation problems. The basic idea behind the statistical method lies in systematically recalculating the population parameters by using subsets of available data or drawing randomly with replacement from a set of data points. And the parameter could be a mean, median, proportion, odds ratio, and correlation coefficient. Distinguished from classical parametric tests comparing the observed statistics to theoretical sampling distributions, resampling is a novel methodology whose inference is based upon large member of repeated samplings. It means that the more we increase the number of times in sampling for estimation, the more accurate the average of various estimated population statics can be obtained.

Algorithm 1. Algorithm to select bootstrap evaluation samples.

```
Input: $X^1$: Initial sample set $X^1 = \{x_{11}, \ldots, x_{1N}\}$
    $B$: Number of Bootstrap sample
    $P$: Population
    $N$: Sample set size
Output: $BS$: Bootstrap sample: $\{X^1, \ldots, X^B\}$

1. Select a set of random samples containing $N$ values with
   replacement from $X$ to form new sample $X^b$;
2. Repeat Step 1 $B-1$ times and $B$ is set to be a large number;
3. return $BS$;
```

The Bootstrap is one of the most popular methods to obtain estimators for parameters in statistics and was invented by B. Efron in 1979. The key idea of the estimation technique is to resample data with replacement many and many times in order to produce a robust and reliable inference of a statistical parameter [42]. It has been successfully and widely applied to many fields of science, engineering and experimental medicine. Its basic procedure is quite simple and can be illustrated in Algorithm 1. The method involves resampling sample data with replacement many and many times in order to produce an empirical estimate of the entire sampling distribution of a parameter of interest. The inputs of the algorithm include target data set and number of bootstrap sample B. The output of the procedure is sets of bootstrap samples $BS = \{X^1, \ldots, X^B\}$. Using the samples, the bias between the estimated value and real value – $|E(\hat{\mu}) - \mu|$ – can be reduced with the bootstrap method significantly. Indeed, similar to other resampling methods, the nature of the bootstrap sampling scheme suggests that the more samples taken (more resampling done), the smaller difference between $E(\hat{\mu})$ and μ can be expected. Alternatively, when the more samples are taken, the better the convergence to zero of $|E(\hat{\mu}) - \mu|$ would be. Assuming that (1) resampling has been done B times, thus taking B samples from the population P and that (2) each of the B samples is equally picked from the population, the expected value $E(\hat{\mu}) = \sum_{\hat{\mu}} \hat{\mu} p(\hat{\mu})$ would be reduced to

$$E(\hat{\mu}) = \frac{\sum_{\hat{\mu}} \hat{\mu}}{B}. \quad (1)$$

This expected value is, in fact, the average value used by the bootstrap method as the estimator of μ. In fact, different extensions for bootstrapping are available.

Table 1
Summary and comparison of the state-of-the-art image spam filtering methods.

Filtering methods	Multiple feature based	Multiple resolution based	Training sample selection	Basic learning algorithm used
FISND [19]	Yes	No	Manual	Bayesian classifier
VF-SVM [20]	Yes	No	Manual	Support Vector Machines (SVMs)
GMM-AIB [20]	Yes	No	Manual	Gaussian Mixture Models (GMM)
RoBoTs	Yes	Yes	Resampling	LDA+Random Forest

4. RoBoTs system

This section introduce how RoBoTs can detect image spam using less training cost accurately and robustly. The architecture, as illustrated in Fig. 2, consists of three components: spam image content modelling and feature extraction, linear discriminative forest and resampling based training sample selection. In following, we firstly present basic methodology for spam image modelling and the feature extraction scheme in Section 4.1. Then, Sections 4.2 and 4.3 introduce a novel classification scheme called LDF and associated training algorithms (e.g. the procedure to select training examples) respectively. Section 4.4 gives a summary about spam identification process using the RoBoTs system. The notation used in this paper is defined in Table 2.

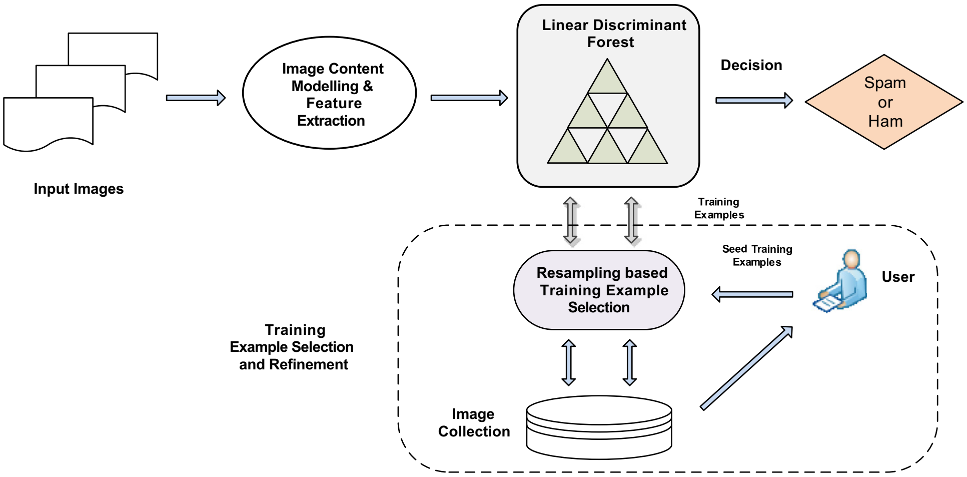
Input Images
Image Content Modelling & Feature Extraction
Linear Discriminant Forest
Decision
Spam or Ham
Training Examples
Seed Training Examples
Resampling based Training Example Selection
User
Training Example Selection and Refinement
Image Collection

Fig. 2. Architecture of RoBoTs spam image detection system.

Table 2
Summary of symbols and definitions

Symbols	Definitions
c	Message image class (where $c=1$, message is spam and $c=0$, message is ham)
η^c	Relevance score for class c
Ω_j	Transformation matrix for CART j
$g(x,y)$	Gabor function for texture feature extraction
fv_r	Composite feature vector containing various features extracted from $r \times r$ partition
co, tx, sh	Color feature, texture feature, shape feature
H	Linear discriminative forest classifier
τ_j	jth CART in the random forest
J	Total number of CARTs in the random forest
B	Total number of bootstrap samples
BS	Bootstrap sample set with B bootstrap samples
K	Number of disjoint feature subsets for LDF training

4.1. *Spam image modelling*

Effective content modeling is critical in various image spam analysis applications. Due to high complexity, it is essential to integrate multiple kinds of visual features to gain a high quality visual representation. On the other hand, subregions in image could include salient characteristics to facilitate accurate classification. However, very surprisingly, less existing approaches consider those for the purpose of comprehensive spam image modeling [43–45]. Motivated by the key observations, the proposed RoBoTs system applies multiresolution and multifeature based image modeling scheme. The raw visual features applied are similar to the ones used in [46]. The details can be found in Fig. 3.

For each partition at level r, RoBoTs extract three different kinds of low level features from every subblock as multimodal based representation for image. They include color, texture, and shape. The composite feature vector fv_r from partition level r can be expressed as

$$fv_r = (fv_r^{co} - fv_r^{tx} - fv_r^{sh}) \tag{2}$$

where fv_r^{co}, fv_r^{tx} and fv_r^{sh} denote color, texture and shape feature vectors, respectively, extracted from $r \times r$ partition. In our system, color feature vector consists of two main components – color histogram and color layout. To calculate color histogram, we firstly partition the images into multiple 8×8 sub-blocks and the average color over the blocks are calculated. The 8×8 DCT (Discrete Cosine Transform) is conducted to calculate a series of histogram coefficients. The color space considered is YCrCb and the dimensionality of color layout is 30. This includes top 10 values from Y, Cr and Cb coefficients. To extract texture feature, we apply Gabor filter to calculate effective signature to effectively optimize the joint uncertainty cross the space and frequency. Also, its frequency and orientation have been proven to be very similar to those of the human visual system. In our system, filter bank, which contains a set of the Gabor filters (4 scales and 6 orientations) are used to generate a 48 dimensional feature vector. Expect color and texture feature, we also use Canny edge operator to calculate edge histogram, which is a 30 dimensional shape feature.

Algorithm 2. Linear Discriminant Forest construction.

```
Input: $X^t$: Training examples ($N \times n$ matrix)
  $N$: Number of training examples
  $Y^t$: Labels of training example ($N \times 1$ matrix)
  $K$: Number of disjoint feature subsets
  $J$: Number of CARTs to build
  $\mu$: Percentage of bootstrap sample
  $\{1,\ldots,c\}$: class labels, where $c$ is 2
Output: Linear Discriminant Forest: $H$

for $j = 1, 2, \ldots, J$ do
    Partition feature vector $x$ randomly into $K$ disjoint subsets : $x_{i,k}$;
    for $k = 1, \ldots, K$ do
        Extract subset of training examples $X^t_{i,k}$ from $X^t$ for $x_{i,k}$;
        Draw a bootstrap sample $X^{ts}_{i,k}$ containing $\mu$%of number of objects in $X^t_{i,k}$;
        Run LDA over $X^{ts}_{i,k}$ to obtain coefficients of matrix $\Omega_j$;
    Construct matrix $\Omega_j$ with coefficients generated in Step 6;
    Train a randomized CART $\tau_j$ with $(X^t\Omega_j, Y^t)$;
    $H \Leftarrow H \bigcup \tau_j$;
Return $H$;
```

4.2. *Linear Discriminant Forest*

Spam filtering can be essentially modelled as binary image classification. Classifier design plays a very critical role in determining final performance of whole system. Motivated by the concern, we develop a novel ensemble method called the Linear Discriminant Forest (LDF) to seamlessly integrate Linear Discriminative Analysis (LDA) [12,13] and Random forests [11]. It adapts LDA, which is linear feature selection scheme, over subsets of raw features. The main advantage of LDA over other linear subspace methods is to generate a discriminative feature space to maximize the ratio of between-class scatter against within-class scatter (Fisher's criterion). Meanwhile, a more comprehensive feature combination is reconstructed for better performance. Random forest is a ensemble method for fusing a collection of J randoon and regression trees (CARTs). It can be denoted as

$$H = \{\tau_j(fv_r, \theta_j), \quad j = 1, \ldots, J\} \tag{3}$$

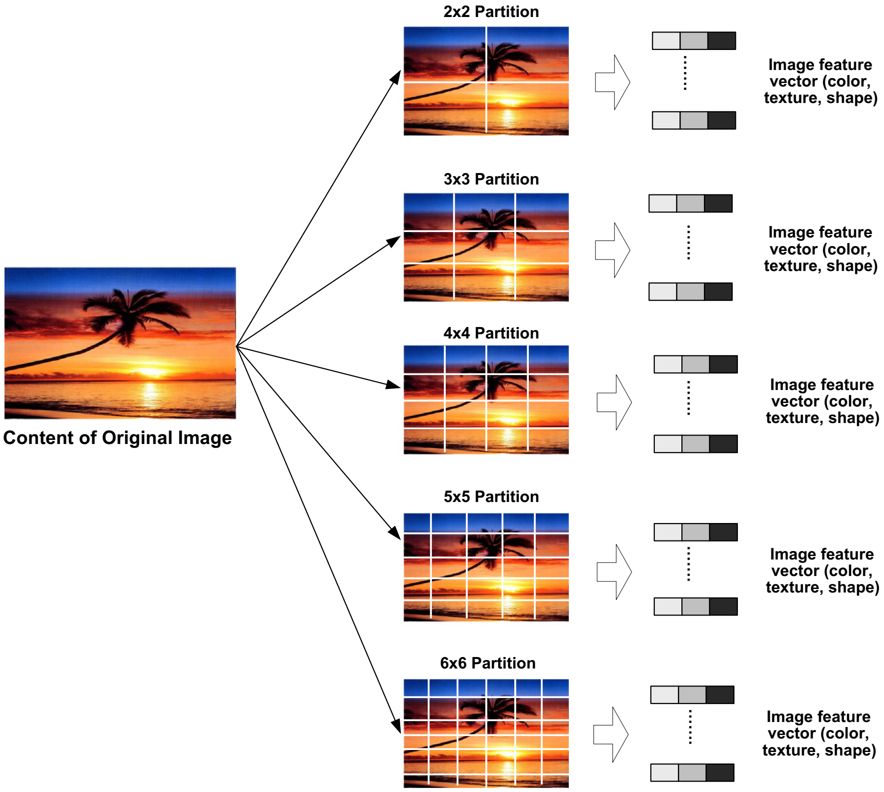
2x2 Partition
3x3 Partition
4x4 Partition
5x5 Partition
6x6 Partition
Image feature vector (color, texture, shape)
Image feature vector (color, texture, shape)
Content of Original Image
Image feature vector (color, texture, shape)
Image feature vector (color, texture, shape)
Image feature vector (color, texture, shape)

Fig. 3. Multiresolution and multifeature based spam image content modeling scheme. Image is partitioned using five different configurations (2 × 2, 3 × 3, 4 × 4, 5 × 5 and 6 × 6) and after partition, three different visual features are extracted from each subregion.

where τ_j is the jth CART, J is the total number of CARTs in the random forest and fv_r is an input composite feature vector. Each CART predictor τ_j is grown with learning examples and a random vector. To estimate model's parameters, learning examples are obtained by randomly resampling the original training sets with replacement. Random vector is generated via independently sampling and follows same distribution as the past random vectors $\{\theta_1, \ldots, \theta_{j-1}\}$. To improve the effectiveness and efficiency of the original random forest, our LDF employs the two phrase process to reduce the size of feature vectors. Details about its training procedure are illustrated in Algorithm 2.

The first step of training set construction for CART τ_j is to randomly split input visual feature vector x into K disjoint subsets. Thus the size of each feature subset is $M = n/K$, where n is the dimensionality of input features. In the second step, the ith subset of features $x_{i,j}$ is selected to train CART τ_j in a random fashion. In this study, for each of those subsets, the value of μ is set to be 70% and thus we draw a bootstrap sample with 70% size of original training set. After applying LDA over M features and the selected subset of training set X_t, the coefficients of LDA transformation $\mathbf{w_{ij}} = \{w_{ij}^1, \ldots, w_{ij}^{M_j}\}$ can be obtained, where $M_j < M$. If same procedure is used for all K disjoint feature subsets, a transformation matrix Ω_j can be generated to preprocess input features for CART τ_j

$$\Omega_j = \begin{pmatrix} \mathbf{w_{11}} & 0 & \cdots & 0 \\ 0 & \mathbf{w_{22}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{w}_{Kj} \end{pmatrix}. \tag{4}$$

Algorithm 3. Automatic training example generation based on AdaBoost based bootstrap resampling.

```
Input: $EL_t$: Seed labeled training examples
  $EUL_t$: Initial unlabeled training examples
  $H$:LDF classifier
  $M$: Number of examples for sampling subset
  $T$:Number of iteration
Output: $EL_T$: Labeled training examples

for $t = 1, 2, \ldots, T$ do
  Select $B$ bootstrap subset of samples $X_b$ from $EL_t$;
  Train $H$ with each sample $X_b$ using Algorithm2 and get a set of LDF classifiers $(H_1, \ldots, H_B)$;
  Apply AdaBoost to $(H_1, \ldots, H_B)$ and generate a final classifier $G$;
  for each sample $e$ in $EUL_t$ do
    Use classifier $G$ to classify $e$;
    if $e$ is spam then
      Add $e$ into $EL_{t+1}$;
      Remove $e$ from $EUL_t$;

return $EL_T$;
```

4.3. *Resampling based training example selection*

A large number of high quality learning examples cannot be expected for training under many real applications because manual labeling can be very costly in term of time and labor. To address this issue, we design a algorithm using Adaboost based Bootstrap resampling to satisfy two properties:

- *Effectiveness*: Generate a LDF classifier with accurate probability prediction for binary classification.
- *Efficiency*: Reduce the size of initial learning examples for training process. This can be very helpful to minimize the cost related to manual selection at the beginning of classifier training. This is a desirable feature for spam filtering.

The detail procedure can be found in Algorithm 3 and its basic idea is to employ resampling strategy to derive a discriminant classifier for selecting high quality training examples in each iteration. In the first step, a set of B bootstrap subsets (X_b, where $b = 1, \ldots, B$), are generated from labeled samples EL_t (line 2). Then, LDF learning models are trained using each of those subset X^b and it creates B LDF classifiers ($H_1,\ldots,H_B$) (line 3). They can be treated as a set of weak classifiers. For the purpose of effective train example selection, *RoBoTS* constructs a final classifier with discriminant function G using AdaBoost

$$G = \sum_{b=1}^{B} \alpha_b H_b \tag{5}$$

where α_b is a set of weights for combining weak classifiers. AdaBoost has been successfully applied to a variety of classification problem and has empirically proven to be highly competitive in various domain applications. It takes the majority vote principle by classifiers and aims to minimize the classification risk by linear combination of classifiers. Once AdaBoost classifier G is ready, we apply it to identify whether each sample e in EUL_t is spam or not. If so, the sample e is included the seed learning sample set for next round training – EL_{t+1}.

4.4. *Image spam filtering with RoBoTs*

The proposed RoBoTs system aims to detect whether incoming image based e-mail message is spam or not. With the system architecture introduced in the previous section, we introduce details about the filtering process now. To training the system, a small amount of seed samples EL_1 need to be labelled in the first step and then Algorithm 4 is used to generate training examples EL_T. We apply feature extraction scheme to each resolution level introduced in Section 4.1. After that, a group of *LDF* classifiers $\{H_2, \ldots\ldots, H_{R+1}\}$ are constructed using training example EL_T, where R is the value of predefined resolution level. In this study, five different resolution levels are considered for spam image modeling and training procedure is repeated five times, one for each grid configuration. Thus, R is set to be 5.

Algorithm 4. Algorithm for automatic spam identification using RoBoTs.

```
Input: $I$: Input image message
   $R$: Resolution level
Output: $c$: Label of class

Process the image message $I$ to obtain different grid based
    image representations;
Extract various kinds of features from image with grid based
    representations at different resolutions;
Using a set of LDF classifiers – ($H_2,\ldots, H_R$) to derive relevance
    scores for two classes (ham and spam);
Assign $I$ to the class with the largest relevance score with Eq.
    (6);
```

Once training is completed, the task of spam image identification can be carried. The basic procedure is shown in Algorithm 4 and consists of four steps. For a given image based message, at the initial stage of the process, the system partitions the image using different grid configuration (line 1). After that, the feature extraction procedure generates three different kinds of visual features using the techniques described in Section 4.1. Next, the features serve as input to *LDF* classifiers ($H_2,\ldots\ldots, H_{R+1}$). The likelihood score η_c based on the *LDF*s can be calculated for image class c using

$$\eta^c = \frac{\sum_{r=2}^{R+1} \eta_r^c}{R} \tag{6}$$

where η_r^c is the *LDF* for resolution *r*. η^c serves as distance between the incoming image message and the label. In the final step, the label for the class with the largest relevance score is assigned to input image message *I*.

5. Experimental configuration

This section presents the experimental settings for the empirical evaluation with goal of performance comparison. The details include test collections and evaluation metrics for performance assessment and technical summary of different competitive systems considered.

Unlike general image classification task, very few standard image spam corpora are public available due to privacy concerns. In order to gain better and insightful results for performance comparison, we select three datasets for the our empirical study. They are:

- *TSI*: This test collection includes 13,261 files and was constructed using image spam e-mail from SpamArchive.[1] Only 10,623 files are in image format and can be handled by image processing algorithm. TSI has been used for a empirical study in [15].
- *TSII*: Alternative name for this collection is the *PersonalSpam* dataset and it was created by Dredze et al. [35] using their personal e-mails. For TSII, a total number of spam images is 3300.
- *TSIII* : It consists of 1071 images belonging to 178 different categories [19]. All those spam images are collected from several personal e-mail accounts including accounts from two popular online web-based e-mail service providers, one IT company account and three education accounts. For each category, there are permutations of image spam templates. Examples can be found in Fig. 4. The minimum, maximum, average and standard deviation for image sizes in those batches are 2, 50, 6.02 and 6.39. TSIII is also called the Princeton Image Spam Benchmark and has been used for empirical study in [19,20].

All test collections above are publicly available. Table 3 summarizes the basic information about each testbed used in our empirical study. In addition, since no information about how to generate home grown ham for TSI and TSIII can be found in related

[1] http://www.spamarchive.org/.

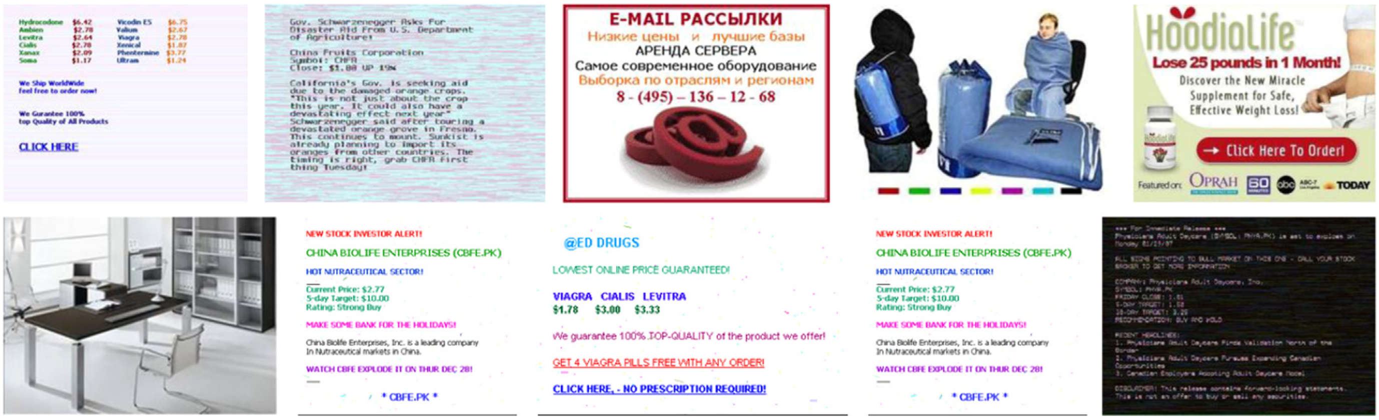
E-MAIL РАССЫЛКИ
Низкие цены и лучшие базы
АРЕНДА СЕРВЕРА
Самое современное оборудование
Выборка по отраслям и регионам
8 - (495) – 136 – 12 - 68
HoodiaLife
Lose 25 pounds in 1 Month!
Click Here To Order!
NEW STOCK INVESTOR ALERT!
CHINA BIOLIFE ENTERPRISES (CBFE.PK)
HOT NUTRACEUTICAL SECTOR!
MAKE SOME BANK FOR THE HOLIDAYS!
WATCH CBFE EXPLODE IT ON THUR DEC 28!
@ED DRUGS
LOWEST ONLINE PRICE GUARANTEED!
VIAGRA CIALIS LEVITRA
$1.78 $3.00 $3.33
GET 4 VIAGRA PILLS FREE WITH ANY ORDER!
CLICK HERE, - NO PRESCRIPTION REQUIRED!
CLICK HERE

Fig. 4. Examples of spam images used from the test collection TSIII.

literature [19,20], we use our own one, which consists of images, logos, and other items appeared in real world applications.

The main aim of the system is to identify which inputs are image spams and which are ham. Essentially, the task is a binary classification process. Thus, our evaluation method focuses on how accurate the filtering process using different approaches on a particular dataset can be. The spam classification accuracy ratio – *sm* is the key performance measures used here

$$sm = \frac{d}{c+d} \tag{7a}$$

Details about *a*, *b*, *c*, *d*, and the possible outcomes by a spam detection scheme can be found in Table 4 (contingency table). The performance of the proposed system and other three the state-of-the-art schemes are studied and analyzed from two different aspects – identification effectiveness and robustness. The three methods include VF-SVM [20], GMM-AIB [20] and FISND [19]. Detail information about the methods can be found in Section 2 (Literature review). It is worth noting that since GMM-AIB is based on GMM, its performance is sensitive to parameter settings. In this experiment, we use the same methodology present in [20] is used to select the GMM parameters.

In addition, Table 5 summarizes the system configurations (e.g., parameters and features considered) which may potentially have an impact on our performance study. In our experiments, all parameters use default values unless otherwise specified.

6. An empirical study

In this section, we report a set of experimental studies to evaluate and compare the proposed RoBoTs system with other existing approaches from different aspects. The spam detection schemes considered include FISND, VF-SVM, and GMM-AIB.[2] First, Section 6.1 presents the experimental results and associated analysis about identification accuracy comparison. Then, Section 6.2 presents empirical results about the robustness improvement of RoBoTs over several existing schemes. Finally, we discuss a few technical issues associated with the effects of feature integration and visual resolution on the RoBoTs in Section 6.3.

Table 3
Summary of three image spam test collections used in this empirical study.

Name	Size	Download website	Notes
TSI	13,261	http://www.seas.upenn.edu/~mdredze/datasets/imagespam	Only 10623 can be handled by image processor
TSII	3300	http://www.seas.upenn.edu/~mdredze/datasets/imagespam	Alternative name: *PersonalSpam* dataset
TSII	1071	http://www.cs.princeton.edu/cass/spam/spambench	Alternative name: Princeton Image Spam Benchmark

6.1. On filtering accuracy

This section describes a comparative study on the accuracies of the various spam filtering schemes. Basic methodology is to select a series of training sets and size of the training sets ranges from 5% to 95% of corresponding testbed. Fig. 5 summarises the results gained. Overall, the experimental results show that RoBoTs always outperforms other approaches. The empirical outcomes also demonstrate that as long as trained using sufficient amount of learning examples, all four different methods achieve good spam detection accuracy. For example, when the size of training examples is more than 10%, at least 90% accuracy can be observed for all the methods tested.

The second experiment tests the performance of all the methods with different training example sizes. We find that all filtering accuracies decrease to some extent when the size of training examples becomes smaller. For example, for TSI, when using 10% testbed as learning examples, detection accuracy of VF-SVM is 89%. With 25% training examples, accuracy increases to 96.5%, which is about 8.5% gain.

As shown previously, hybrid system architecture of RoBoTs, which combines LDA and random forest, demonstrates promising performance for spam detection. It is interesting to test performance of LDA and random forest under same environment and do comparison with RoBoTs. Table 6 shows a detail summary of related experimental results. We observe that LDA achieves the lowest accuracy. Further, random forest provides 15.7% performance gain over LDA. In comparison with LDA and random forest, detecting spams with RoBoTs results in a significant improvement in accuracy for all the different testbeds. For example, the improvement with RoBoTs against random forest is from 8.98% to 16.26%, depending on the testbed used.

6.2. On detection robustness

Learning-based classification schemes are often required to work under the environment with certain kinds of resource constraints.

[2] We use notation introduced in Section 2 to represent the methods considered.

Table 4
Contingency table of spam filtering.

Category	Ham	Spam
Ham	a	b
Spam	c	d

Table 5
System parameters and configurations.

Notation	Definition (default values)
TE	Size of training example (10%)
R	Resolution level (5)
μ	Size of bootstrap sample (70%)
co, *tx*, *sh*	Color, texture and shape feature

Table 6
Performance comparison of three classification scheme: RoBoTs, random forest and LDA.

Classification scheme	TSI (%)	TSII (%)	TSII (%)
RoBoTs	91.5	96.8	91
random forest	78.7	85.5	83.5
LDA	65.3	73.1	71.5

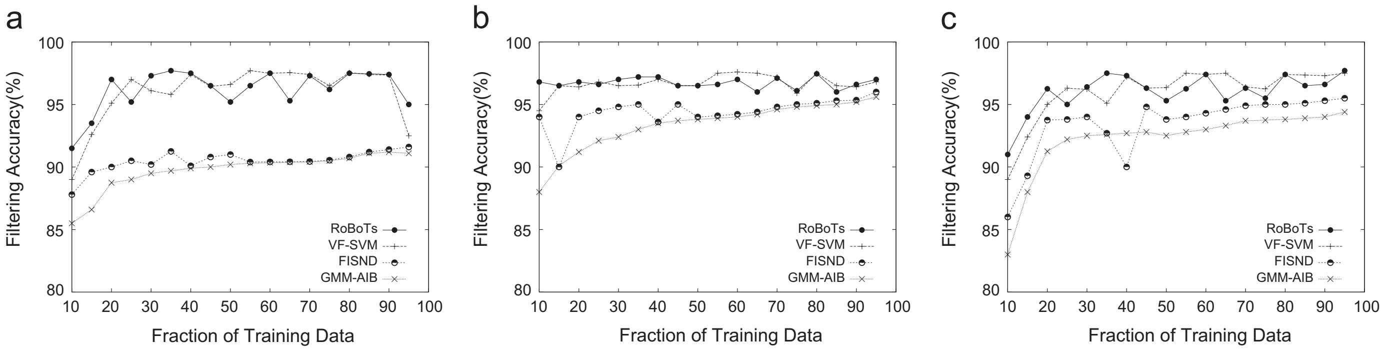

Fig. 5. Spam detection accuracy comparison over three testbeds: (a) TSI; (b) TSII; (c) TSIII (Princeton Dataset).

For example, in many real applications, training resources can be reduced by using limited amount of or lower quality training examples. Such scarification may lead to a less accurate or robust classifier. In the following section, we compare RoBoTs with several systems on their robustness under various training environments. The cases include

- *Small training examples*: Training example is one of the most crucial learning resources. Labelling training examples is expensive since it relies on a manual selection. If the system can perform well with a small number of training examples, this will significantly reduce the training cost.
- *Mislabelled training examples*: Human based labelling might not be always reliable and consistent. In many cases, the training datasets labelled by human could include mislabelled training examples. It is desirable for a spam detection system to enjoy superior robustness against this type of error.

In many practical applications, a large amount of good quality training examples might not be always available. Thus, it would be great if learning based spam detection system can achieve high performance only using small training examples. The first experiment aims to investigate the effects of small training set size on the accuracy of RoBoTs and other methods. In order to make the experimental results more reliable, we select data belonging to different categories uniformly from our test collections. After selecting different portions of learning examples, we study how RoBoTs and other competitors behave with size changes of training examples (from 2% to 10%). As illustrated in Fig. 6, the accuracy of all methods degrades after the size of the training examples is decreased to a certain level. However, when compared to the other competitors, RoBoTs demonstrates more robust performance when using relatively small size of training examples. We believe the better robustness of RoBoTs is due to the fact that multiple resolution and multiple feature combination can be very helpful to gain more informative signatures for spam images. Further, AdaBoost based resampling scheme for training example selection can capture more high quality with much less initial learning examples.

The second experiment investigates the effects of mislabelled training examples. To carry out experiments under different sizes of incorrect training data, 1%, 5%, 10%, 15%, 20%, 25% and 30% training data from both datasets were randomly selected and their original labels were reversed. Fig. 7 summarizes the precision rate versus the different settings on the various methods. The results show that RoBoTs is superior to the other approaches when proportion of mislabelled data is gradually increased. As shown in the Figure, the performance of the three other approaches degrade dramatically after the size of the mislabelled data is greater than 10% of the whole training set. In contrast, RoBoTs maintains reasonable accuracy even with 20% incorrect training data in some cases. Based on the results, we can easily conclude that by taking advantage of the resampling based training example selection and comprehensive visual content modeling scheme, our scheme is able to achieve much better robustness against mislabelled training data.

6.3. *On effects on system configuration*

It is important to experimentally evaluate the spam detection accuracy under different system settings. Particularly, this study aims to find out how different strategies for visual feature combinations and image partition combination can influence the performance of RoBoTs.

Firstly, we investigate the effects of various visual feature combinations on accuracy improvement of spam detection using RoBoTs. Basic methodology for the test is that we progressively incorporate additional visual feature into RoBoTs and compare the results. The RoBoTs system was tested based on four different visual feature combinations: (colour, texture), (colour, shape), (texture, shape) and (colour, texture, shape). Table 7 summarises the results of this study (of visual feature configurations) via the precision measurement. The main observation gained is that additional visual feature integration introduces significant

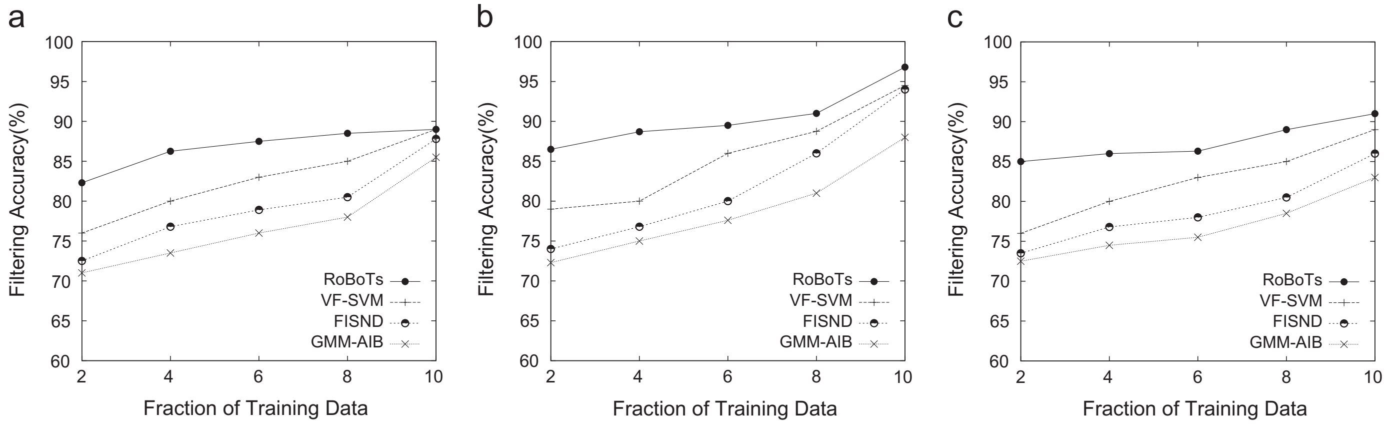

Fig. 6. Spam detection accuracy comparison with small size of training examples: (a) TSI; (b) TSII; (c) TSIII (Princeton Dataset).

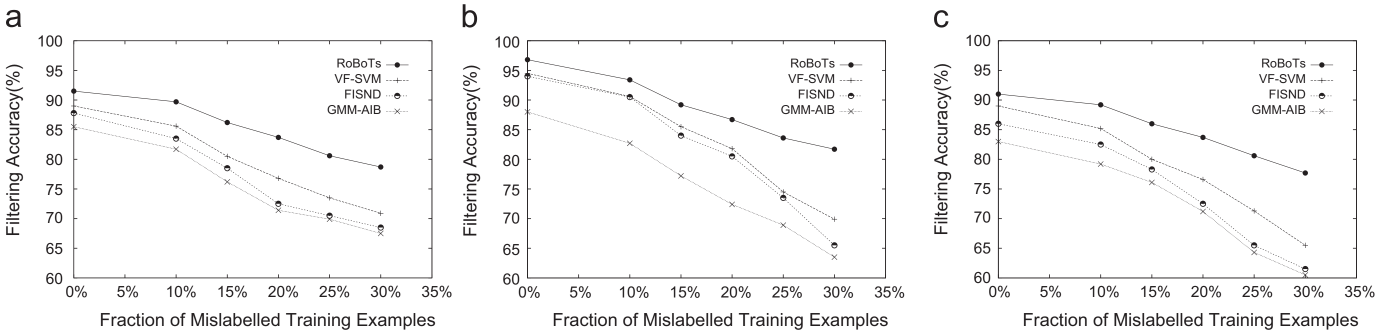

Fig. 7. Performance comparison with different size of mislabelled training examples: (a) TSI; (b) TSII; (c) TSIII (Princeton Dataset).

Table 7
Performance summary of RoBoTs with different feature configurations. For feature configuration, C, T and S denote colour, texture and shape. The size of training example is 10% of whole test collection.

Visual feature combination	TSI (%)	TSII (%)	TSII (%)
C,T,S	91.5	96.8	91
C,S	81.7	87.5	82.5
C,T	82.3	87.1	83.5
T,S	80.5	85.6	81.4

Table 8
Performance summary of RoBoTs with combinations of partition levels. The size of training example is 10% of whole test collection.

Partition level	TSI (%)	TSII (%)	TSII (%)
5,4,3,2	91.5	96.8	91
4,3,2	85.1	90.5	86.4
3,2	82.5	85.6	82.6
2	77.6	81.4	78.5

improvement on the effectiveness of the system. For example, using colour, texture and shape gives an additional 12.5% gain in detection precision over using only colour and texture on TSI.

The objective of the second experimental study addresses the question about how the combinations of various image partitions influence detection accuracy of RoBoTs. Four different settings tested include (5×5, 4×4, 3×3 and 2×2), (4×4, 3×3 and 2×2), (3×3 and 2×2) and (2×2). The experimental results are reported in Table 8. It shows that more partition level considered by RoBoTs, the significant performance improvement can be observed. For example, when considering four different partition levels (5×5, 4×4, 3×3 and 2×2), RoBoTs achieves 7.5% accuracy gain over RoBoTs with three partition levels (4×4, 3×3 and 2×2). The main reason is that more comprehensive modeling for an image can be gained when more partition levels are considered in RoBoTs. Consequently, we can expect better detection accuracy.

7. Conclusion

In this paper, we present a novel system, called RoBoTs, to facilitate effective and robust image spam detection based on (1) an efficient learning sample selection scheme and (2) an effective ensemble method – Linear Discriminant Forests (LDFs). We have fully implemented the system and assess its performance using large scale test collections. As demonstrated in the empirical study, RoBoTs system not only enjoys significantly better accuracy over the existing approaches but also achieves strong robustness.

Effective image spam detection is of importance in many different domain applications. The system framework, associated learning algorithms and empirical results present in this study elaborate on several importance issues for further scholarly investigation. It would be interesting to study how to develop intelligent feature extraction scheme for spam identification accuracy improvement based on deep learning technique [47,48]. Meanwhile, in order to gain fair and reliable performance comparison, test collection plays a key role. As such, developing large scale test collection is another promising direction for future exploration. We hope that this work can provide an impetus for further investigation on this important research area.

Conflict of interest

None declared.

References

[1] L. Weinstein, Spam wars, Commun. ACM 46 (8) (2003).
[2] N. Holmes, In defense of spam, IEEE Comput. 38 (2005).
[3] Secure Computing. Image Spam: The Latest Attack on the Enterprise Inbox, Available online, November 2006.
[4] A. Anti-Spyware, Available at ⟨http://www.symantec.com⟩.
[5] L. Zhang, X. Zhen, L. Shao, Learning object-to-class kernels for scene classification, IEEE Trans. Image Process. 23 (8) (2014) 3241–3253.
[6] D. Blei, M.I. Jordan, Modeling annotated data, in: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Ser. SIGIR '03, 2003, pp. 127–134.
[7] N. Pinto, D.D. Cox, J.J. DiCarlo, Why is real-world visual object recognition hard? PLoS Comput. Biol. 4 (1) (2008).
[8] N. Rasiwasia, J.C. Pereira, E. Coviello, G. Doyle, G.R. Lanckriet, R. Levy, N. Vasconcelos, A new approach to cross-modal multimedia retrieval, in: Proceedings of the International Conference on Multimedia, Ser. MM '10, 2010, pp. 251–260.
[9] L. Shao, L. Liu, X. Li, Feature learning for image classification via multiobjective genetic programming, IEEE Trans. Neural Netw. Learn. Syst. 25 (7) (2014) 1359–1371.
[10] F. Zhu, L. Shao, Weakly-supervised cross-domain dictionary learning for visual recognition, Int. J. Comput. Vis. 109 (1–2) (2014) 42–59.
[11] L. Breiman, Random forests, Mach. Learn. 45 (1) (2001) 5–32.
[12] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley & Sons, New York, 2000.
[13] V. Vapnik, Statistical Learning Theory, John Wiley & Sons, New York, 1998.
[14] A. Schwartz, SpamAssassin, O'Reilly Media, Inc., Germany, 2004.
[15] G. Fumera, I. Pillai, F. Roli, Spam filtering based on the analysis of text information embedded into images, J. Mach. Learn. Res. 6 (2006).
[16] M. Soranamageswari, C. Meena, Statistical feature extraction for classification of image spam using artificial neural networks, in: ICMLC, 2010, pp. 101–105.
[17] M. Gao, Y. Yang, X. Zhao, B. Pardo, Y. Wu, T.N. Pappas, A. Choudhary, Image spam hunter, in: ICASSP 2008, 2008, pp. 1765–1768.
[18] B. Biggio, G. Fumera, I. Pillai, F. Roli, A survey and experimental evaluation of image spam filtering techniques, Pattern Recognit. Lett. 32 (10) (2011) 1436–1446.
[19] Z. Wang, W.K. Josephson, Q. Lv, M. Charikar, K. Li, Filtering image spam with near-duplicate detection, in: CEAS, 2007.
[20] B. Mehta, S. Nangia, M. Gupta, W. Nejdl, Detecting image spam using visual features and near duplicate detection, in: Proceedings of the 17th International Conference on World Wide Web (WWW'08), 2008, pp. 497–506.
[21] P. He, X. Wen, W. Zheng, A simple method for filtering image spam, in: ICIS, 2009, pp. 910–913.
[22] Z. Qu, Y. Zhang, Filtering image spam using image semantics and near-duplicate detection, in: ICICTA, 2009, pp. 600–603.
[23] Z. Chen, et al., Image spam clustering: an unsupervised approach, in: Proceedings of the First ACM workshop on Multimedia in Forensics, 2009, pp. 25–30.
[24] M.A. Khanum, L.M. Ketari, Trends in Combating Image Spam E-Mails, arXiv preprint arXiv: arXiv:1212.1763, 2012.
[25] A. Attar, R.M. Rad, R.E. Atani, A survey of image spamming and filtering techniques, Artif. Intell. Rev. 40 (1) (2013) 71–105.
[26] S. Krasser, Y. Tang, J. Gould, D. Alperovitch, P. Judge, Identifying image spam based on header and file properties using c4. 5 decision trees and support vector machine learning, in: IAW, 2007, pp. 255–261.
[27] M. Uemura, T. Tabata, Design and evaluation of a Bayesian-filter-based image spam filtering method, in: ISA, 2008, pp. 46–51.
[28] B. Byun, C.-H. Lee, S. Webb, C. Pu, A discriminative classifier learning approach to image modeling and spam image identification, in: CEAS, 2007.
[29] K.J. Liszka, C.-C. Chan, Application of learning algorithms to image spam evolution, in: Emerging Paradigms in Machine Learning, Springer, Berlin Heidelberg, 2013, pp. 471–495.
[30] Q. Liu, Z. Qin, H. Cheng, M. Wan, Efficient modeling of spam images, in: IITSI, 2010, pp. 663–666.
[31] C. Wu, Q. Cheng, K. Zhu, Y. Wu, Using visual features for anti-spam filtering, in: ICIP, vol. 3, IEEE, Genova, Italy, 2005, pp. III–509.
[32] C. Wang, F. Zhang, F. Li, Q. Liu, Image spam classification based on low-level image features, in: ICCCAS, 2010, pp. 290–293.
[33] H. Zuo, W. Hu, O. Wu, Y. Chen, G. Luo, Detecting image spam using local invariant features and pyramid match kernel, in: WWW, pp. 1187–1188.
[34] J. Hsia, M. Chen, Language-model-based detection cascade for efficient classification of image-based spam e-mail, in: ICME, 2009, pp. 1182–1185.
[35] M. Dredze, R. Gevaryahu, A. Elias-Bachrach, Learning fast classifiers for image spam, in: Proceedings of the Fourth Conference on E-mail and Anti-Spam (CEAS'07), 2007.
[36] T. Liu, W. Tsao, C. Lee, A high performance image-spam filtering system, in: DCABES, 2010, pp. 445–449.
[37] H.B. Aradhye, G.K. Myers, J.A. Herson, VisualSEEk: a fully automated content-based image query system, in: Proceedings of the 2005 Eighth International Conference on Document Analysis and Recognition, 2005, pp. 87–98.
[38] F. Gargiulo, C. Sansone, Combining visual and textual features for filtering spam emails, in: ICPR, 2008.
[39] H. Cheng, Z. Qin, C. Fu, Y. Wang, A novel spam image filtering framework with multi-label classification, in: ICCCASn, 2010, pp. 282–285.
[40] Y. Gao, M. Yang, A. Choudhary, Semi supervised image spam hunter: a regularized discriminant em approach, in: ADMA, 2009, pp. 152–164.
[41] Y. Gao, A. Choudhary, G. Hua, A comprehensive approach to image spam detection: from server to client solution, IEEE Trans. Inf. Forens. Secur. 5 (4) (2010) 826–836.
[42] B. Efron, R. Tibshirani, An Introduction to the Bootstrap, Chapman & Hall, New York, 1993.
[43] C. Carson, S. Belongie, H. Greenspan, J. Malik, Blobworld: image segmentation using expectation-maximization and its application to image querying, IEEE Trans. Pattern Anal. Mach. Intell. 24 (8) (2002) 1026–1038.
[44] R. Unnikrishnan, C. Pantofaru, M. Hebert, Toward objective evaluation of image segmentation algorithms, IEEE Trans. Pattern Anal. Mach. Intell. 29 (June (1)) (2007) 929–944.
[45] C. Pantofaru, M. Hebert, A Comparison of Image Segmentation Algorithms, Robotics Institute, Pittsburgh, PA, Technical Report CMU-RI-TR-05-40, September 2005.
[46] J. Shen, J. Shepherd, A.H.H. Ngu, An empirical study of query effectiveness improvement via multiple visual feature integration, Int. J. Image Graph. 7 (3) (2007) 551–581.
[47] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, A.Y. Ng, Multimodal deep learning, in: International Conference on Machine Learning (ICML), 2011.
[48] L. Shao, D. Wu, X. Li, Learning deep and wide: a spectral method for learning deep networks, IEEE Trans. Neural Netw Learn. Syst. 25 (12) (2014) 2303–2308.

Jialie Shen is an Assistant Professor in Information Systems, School of Information Systems, Singapore Management University, Singapore. Jialie's main research interests include information retrieval, economic-aware media analysis, and statistical machine learning. His recent work has been published or is forthcoming in leading journals and international conferences including ACM SIGIR, ACM Multimedia, CVPR, ICDE, WWW, IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT), IEEE Transactions on Multimedia (IEEE TMM), ACM Multimedia Systems Journal, ACM Transactions on Internet Technology (ACM TOIT) and ACM Transactions on Information Systems (ACM TOIS). He received the Lee Foundation Fellow for Research Excellence from the Singapore Management University in 2008 and is also winner of Microsoft Mobile plus Cloud Computing Theme Award.

Robert H. Deng has been a Professor with the School of Information Systems, Singapore Management University, since 2004. He was a Principal Scientist and Manager of the Infocomm Security Department, Institute for Infocomm Research, Singapore. His research interests include data security and privacy, multimedia security, network, and system security. He was an Associate Editor of the IEEE Transactions on Information Forensics and Security from 2009 to 2012 and Security and Communication Networks from 2007 to 2013. He is currently an Associate Editor of the IEEE Transactions on Dependable and Secure Computing, and a member of the editorial board of the Journal of Computer Science and Technology (the Chinese Academy of Sciences) and the International Journal of Information Security. He is the Chair of the Steering Committee of the ACM Symposium on Information, Computer and Communications Security. He received the University Outstanding Researcher Award from the National University of Singapore in 1999 and the Lee Kuan Yew Fellow for Research Excellence from the Singapore Management University in 2006. He was named Community Service Star and Showcased Senior Information Security Professional by under its Asia-Pacific Information Security Leadership Achievements Program in 2010.

Zhiyong Cheng received the B.S. degree in Thermal Energy and Power Engineering from Huazhong University of Science and Technology, China, in 2007 and the M.S. degree in Power Machinery and Engineering from Xi'an Jiaotong University, China, in 2010. He was a Research Engineer in School of Information Systems, Singapore Management University, Singapore, from 2009 to 2011. Since August 2011, he has been pursuing the Ph.D. degree in the School of Information Systems, Singapore Management University, Singapore. His research focuses on multimedia retrieval and recommendation.

Liqiang Nie received the B.Sc. degree and Ph.D. from Xi'an Jiaotong University of China, Xi'an and National University of Singapore, Singapore, in 2009 and 2014 respectively. Now He is a research fellow at the School of Computing, National University of Singapore. His current research interests include multimedia content analysis, search, large-scale computing as well as multimedia applications such as multimedia question answering, image reranking and expert mining. Various parts of his work have been published in top forums including ACM SIGIR, ACM MM, TOIS and TMM, etc. Dr. Nie has been a Reviewer for various journals and conferences.

Shuicheng Yan is currently an Associate Professor with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore, and the Founding Lead of the Learning and Vision Research Group. His research areas include computer vision, multimedia, and machine learning, and he has authored or co-authored over 300 technical papers over a wide range of research topics, with Google Scholar citation over 9400 times and H-index-42. He is an Associate Editor of the IEEE Transactions on Circuits and Systems for Video Technology and the ACM Transactions on Intelligent Systems and Technology, and has been serving as the Guest Editor of the special issues for TMM and CVIU. He received the Best Paper Awards from ACM MM in 2012 (demo), PCM in 2011, ACM MM in 2010, ICME in 2010, and ICIMCS in 2009, the winner prizes of the classification task in PASCAL VOC from 2010 to 2012, the winner prize of the segmentation task in PASCAL VOC in 2012, the Honourable Mention Prize of the detection task in PASCAL VOC in 2010, the 2010 TCSVT Best Associate Editor Award, the 2010 Young Faculty Research Award, the 2011 Singapore Young Scientist Award, and the 2012 NUS Young Researcher Award.