2009 Second International Conference on Intelligent Computation Technology and Automation

Filtering Image Spam using Image Semantics and Near-Duplicate Detection

Zhaoyang Qu, Yingjin Zhang
School of Information Engineering
Northeast Dianli University
Jilin City, China
e-mail: qzywww@mail.nedu.edu.cn, zhangyj84@163.com

***Abstract*—Image spam has become the main form of spam, it is a problem crying out for solutions to effectively filter such spam nowadays. This paper proposes an image spam detection system, which is based on image semantics and near-duplicate detection, for solving the problems of current anti-image-spam technologies: low accuracy rate, difficultly recognizing image spam making use of obfuscation techniques and so on. The experimental results show that the system has better filtering effect than previous systems, with increasing more than 10% in accuracy rate and better anti-obfuscation effect, and effectively solves the above-mentioned problems.**

Keywords-spam; image semantics; near-duplicate detection; image spam filtering

I. Introduction

Several years ago, image spam emerged and spread widely later. Comparing with text spam, it consumes more network resources and has greater negative influence. Image spam can easily "escape" traditional spam filter, and becomes more and more difficult to recognize because of the utilization of obfuscation techniques.

An effective image spam detection system should satisfy three requirements: high accurate rate, high efficiency and extensibility [1]. However, in previous filtering technologies of image spam, some of them can only recognize spam images that have certain characteristics and fail to be applied to filter other spam; some are in pursuit of high detection rate but have low accuracy rate; some achieve well filtering effect but don't take account of efficiency.

Previous works mainly focus on study of image low-level visual features (color features, texture features and shape features) [1, 2, 3]. Because different images have similar (even same) visual features, previous filtering technologies of image spam may be wrong in detection and have low accuracy rate. High false positive rate means so many image hams are judged as image spam that would cause losses of E-mail users.

This paper applies image semantics to image spam filtering, and design a new-style image spam detection system based on near-duplicate. The extracted image features of the system combines low-level visual features with high-level semantics features, and are more exact to describe images for increasing accuracy rate of detection. On that basis, this paper proposes relative similarity measurements methods and algorithms. Finally, by comparing the system with previous systems, this paper evaluates the filtering effect of the system.

II. System Architecture

Because spammers usually send image spam in batches that consist of similar features, similarity detection method can effectively filter those on the basis of collecting known image spam. The basic principle of near-duplicate detection system for image spam filtering in this paper is: firstly, extract the features of the detected image (including low-level visual features and high-level semantics features). Secondly, compare the features of it with the features in two feature databases (spam-features DB and ham-features DB) by calculating their similarity, and respectively count the numbers of images that are similar to it in two DB. Finally, judge it is spam or ham by the numbers. The architecture of the system is shown in Fig. 1.

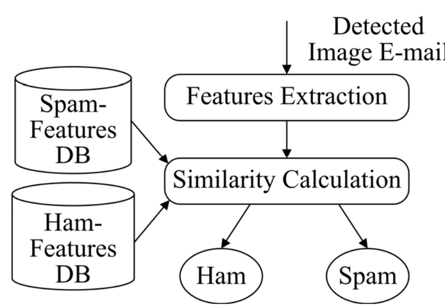

Figure 1. Image spam detection system architecture

III. Image Features Extraction

The process of image features extraction is shown in Fig. 2. Firstly, extract color features, texture features and shape features of the detected image, and combine them into visual feature vector of it. Secondly, use the vector to mapping low-level features to high-level semantics, and make semantics feature vector. Finally, combine the two vectors into the image feature vector of the detected image.

978-0-7695-3804-4/09 $26.00 © 2009 IEEE
DOI 10.1109/ICICTA.2009.151

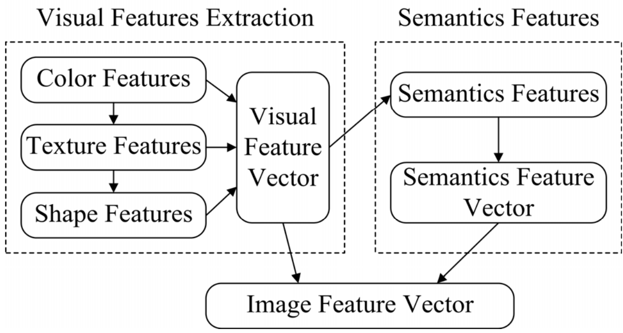

Figure 2. Process of features extraction

A. Image Low-level Visual Features Extraction

1) Color features extraction

Color is a general feature used for describing image in the manner of simplicity and intuition for basic attribute description of image. Color Moment is a simple and effective representation of color features [4]. According to distribution and nature of color, this paper adopts 3 moments of each color component of the RGB model: mean, standard deviation and skewness. Then use these 9 parameters for representing color features.

2) Texture features extraction

Texture is another important but difficult to represent feature of image. Above all, it needs to convert color image into gray level image before extracting texture features. Then get Co-occurrence Matrix and calculate 4 frequently-used texture parameters (Contrast, Angular Second Moment, Entropy and Correlation) [5]. In the end, use mean and second deviation of those for composing 8 dimensions texture feature vector.

3) Shape features extraction

Moment invariants are most effective for describing shape feature of image, they can keep invariable and accurate when image is shifted, rotated or resized [6]. Their extraction is also based on gray level image. Use Sobel algorithm for getting edge information of the image, and get 7 geometric moments by calculating relative parameters using the information.

B. Image High-level Semantics Features Extraction

If extracting each high level semantics features, the system would be very complex and inefficient. Therefore, this paper adopts a method of directly mapping low-level visual features to high-level semantics. The main idea is: extract low-level features, and then repeatedly train low-level features by semantics classification technology to get high-level semantics.

SVM is based on statistical learning theory of VC dimension and structural risk minimisation. According to model complexity of finite sample information (learning precision for given training sample) and learning capacity (faultless recognition capacity to any sample), seek for a best value among them. SVM has many unique advantages for solving problems of small sample, nonlinear and high dimensional mode recognition, and with satisfactory results in many practical applications.

1) SVM design

The form of training sample is:

$$\{(x_1,c_1),(x_2,c_2),\ldots,(x_n,c_n)\} \tag{1}$$

where

$$x_i = [F^{(i)}_{color}, F^{(i)}_{texture}, F^{(i)}_{shape}] \tag{2}$$

$$c_i = \begin{cases} -1, & x_i \in SpamFeatureDB \\ 1, & x_i \in HamFeatureDB \end{cases} \tag{3}$$

x_i is visual feature vector of training image i included color, texture and shape features. c_i is the category of image i.

Considering image content is generally non-deterministic and various, this paper adopts linear inseparable SVM. It converts input space into high dimensional space by nonlinear transforms, making question linear and separable, and solves optimal classification face in new space. Classification function is (where $k(x, y)$ is kernel function) [7]

$$f(x) = sign(\sum_i a_i y_i k(x, x_i) + b) \tag{4}$$

2) Kernel function choice

Because different kernel function will produce difference SVM algorithm, kernel function is directly critical factor for classification and determines final filtering effect. This paper repeatedly experiments and tests the filtering effect of two common-used kernel functions, the formulas and results are as follows

Polynomial kernel function [7]:

$$K(x,x_i) = [(x \bullet x_i) + l]^d, d = 1,2\ldots \tag{5}$$

Gaussian radial basis function [7]:

$$K(x,x_i) = \exp\{-|x - x_i|^2 / \sigma^2\} \tag{6}$$

TABLE I. EXPERIMENTAL RESULTS OF SVM KERNEL FUNCTIONS

	Polynomial (*d*)			**Gaussian (σ)**		
	1	***3***	***5***	***0.3***	***0.7***	***1.0***
Detection Rate (%)	71.16	86.78	91.84	79.42	90.32	82.36
Precision Rate (%)	72.70	88.97	88.67	76.48	91.77	80.78
Accuracy Rate (%)	74.34	82.54	79.10	79.87	90.89	88.48
False Positive Rate (%)	25.66	17.46	20.90	20.13	9.11	11.52

The results show that Gaussian radial basis function under $\sigma = 0.7$ is better.

3) Get high-level semantics features

This paper adopts Gaussian radial basis function as kernel function of SVM (formulas (6), where $\sigma = 0.7$), and use formulas (2) and (4) to get high-level semantics features.

C. Image Features Extraction Algorithm

1- $F^{(i)}_{color} = ExtractColorFeature(Image_i)$;
2- $F^{(i)}_{texture} = ExtractTextureFeature(Image_i)$;
3- $F^{(i)}_{shape} = ExtractShapeFeature(Image_i)$;
4- $F^{(i)}_{low} = F^{(i)}_{color} \cup F^{(i)}_{texture} \cup F^{(i)}_{shape}$;
5- $F^{(i)}_{high} = ExtractHighFeature(F^{(i)}_{low})$;
6- $F^{(i)} = F^{(i)}_{low} \cup F^{(i)}_{high}$;

IV. Similarity Measurements

A. Similarity Measurements Methods

In practical application, each feature vector plays different role. Therefore, this paper adopts a weighted Euclidean distance to calculate similarity measurements. The smaller distance value is, the more similar two images are. Computational formula is as follow [8].

$$D_2 = \sqrt{\sum_{i=1}^{N} \omega_i (A_i - B_i)^2} \quad (7)$$

Because the result of SVM classification function is 1 or -1, it cannot use Euclidean distance formula. So, for catering to the above-mentioned method of spam image filtering, this paper proposes a novel similarity measurement, the formula is

$$J = \omega_l [N(S_l) - N(H_l)] + \omega_h [N(S_h) - N(H_h)] \quad (8)$$

Where $N(x_i)$ is amount of images in feature DB x that are similar to the detected image using feature extraction method i. ω_i is weighted value of method i. $J > 0$ means more spam images similar to the detected image, so the image is judged as spam image. Contrarily, $J < 0$ means more ham images similar to it, so it is judged as ham image.

B. Similarity Measurements Algorithm

1- *for* $n=1$: *spamnum*{
2- *if EuclideanDistance*$(F^{(i)}_{low}, F^{(n)}_{low}) < S\_threshold$ *then* $N(S_l)$+ =1;
3- *if* $F^{(i)}_{high} == -1$ *then* $N(S_h)$ += 1; }
4- *for* $m=1$: *hamnum*{
5- *if EuclideanDistance*$(F^{(i)}_{high}, F^{(n)}_{low}) < S\_threshold$ *then* $N(H_l)$ +=1;
6- *if* $F^{(i)}_{high} == 1$ *then* $N(H_h)$ +=1; }
7- $J = \omega_l(N(S_l) - N(H_l)) + \omega_h(N(S_h) - N(H_h))$;
8- *if* $J > 0$ *then return*('*spam*');
9- *else return*('*ham*');

V. Experimental Results and Evaluation

A. Image Feature DB

The data of image feature DB are shown in Table 2. Where unique image is spam image using obfuscation techniques of shift, resizing, rotation, adding dots, fuzzy or color change.

TABLE II. IMAGE AMOUNT AND SOURCE

	Spam DB	Ham DB	Test DB
Image Amount	8444	2021	791
Image Source	SpamArchive	Individual DB	SpamArchive Unique Image

B. Experimental Results

For high detection rate and low false positive rate, the system mainly relies on high-level semantics features, so ω_l =0.4 and ω_h=0.6. This paper respectively makes an experiment on filtering performance of the new system in this paper comparing with the old system used by previous studies. The old system only extracts low-level visual features excluding high-level semantics features, its architecture is similar to the new system's but without semantics extraction in "Features Extraction". The methods of visual features extraction in the old system are same with the methods in the new system, and use formula (8) (where $N(H_l) = 0$ and $N(H_h) = 0$) to judge whether detected image is spam or ham. The following are the results.

TABLE III. EXPERIMENTAL RESULTS

		Image Total	Old System	New System
Normal Spam Image		371	349	356
Unique Spam Image	***Shift***	16	11	14
	Resize	25	15	21
	Rotate	18	15	18
	Dots	30	17	29
	Fuzzy	21	7	17
	Color Change	45	24	37
Ham Image		265	198	257

C. Figures and Tables

According to the experimental results, the detection rate of the new system is 93.54%, precision rate is 98.4%, accuracy rate is 94.69% and false positive rate is 5.31% by calculation. In comparison, the indexes of the old system respectively are 83.27%, 86.73%, 80.40% and 19.6%. Obviously, the new system proposed by this paper is better than the old system, particularly increasing 14.29% in accuracy rate. Additionally, because obfuscation techniques disturb image description only using low-level visual features, this paper introduces high-level image semantics that can solve the problem, for the reason that the new system is superior to the old one.

VI. CONCLUSIONS

By combining image semantics extraction with near-duplicate detection, this paper proposes a novel image spam filtering system. The experimental results show that firstly, the system can solve the problems of the old system and improve filtering effect that detection rate, precision rate and accuracy rate respectively increase by 10.27%, 11.67% and 14.29%. Secondly, anti-obfuscation technique is also improved, filtering performance of the system for unique image is obviously better than that of the old system.

In conclusion, the new system proposed by this paper gets well-improved as a whole comparing with the system using old methods.

REFERENCES

[1] Z Wang, W Josephson, Q Lv, M Charikar and K Li, "Filtering Image Spam with Near-Duplicate Detection," Proc. the Fourth Conference on Email and Anti-spam (CEAS 2007), Aug. 2007, pp. 346-357.

[2] B Mehta, S Nangia, M Gupta and W Nejdl, "Detecting Image Spam using Visual Features and Near Duplicate Detection," Proc. the 17th International Conference on World Wide Web (WWW 2008), the ACM Digital Library Press, Apr. 2008, pp. 497-506.

[3] M Dredze, R Gevaryahu and A Elias-Bachrach, "Learning Fast Classifiers for Image Spam," Proc. the fourth conference on email and anti-spam (CEAS 2007), Aug.2007, pp. 487-493.

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

[5] R. M. Haralick, K. Shanmugam and I. Dinstein, "Textural Features for Image Classification," IEEE Transactions on Systems, Man, and Cybernetics, vol. 3, No.6, Nov. 1973, pp.610-621, doi: 10.1109/TSMC.1973.4309314.

[6] Ming-Kuei Hu, "Visual Pattern Recognition by Moment Invariants," IRE Transactions on Information Theory, vol.8, No.2, Feb. 1962, pp.179-187, doi: 10.1109/TIT.1962.1057692.

[7] Steve R. Gunn, "Support Vector Machines for Classification and Regression," ISIS Technical Report, May 1998.

[8] Markus Stricker and Markus Orengo, "Similarity of Color Images," Proc. Storage and Retrieval for Images and Video Database (SPIE 95), 1995, pp. 381-392.