J. Vis. Commun. Image R. 34 (2016) 103–107

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Short Communication

Unbalanced JPEG image steganalysis via multiview data match ☆

Anxin Wu, Guorui Feng *, Xinpeng Zhang, Yanli Ren

School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China

ARTICLE INFO

Article history:
Received 7 June 2015
Accepted 19 October 2015
Available online 9 November 2015

Keywords:
Steganalysis
Unbalanced images
Dimension reduction
Multiview match resampling
Weighted fisher linear discriminant
K-means
High-dimensional feature
Semi-supervised learning

ABSTRACT

Image steganalysis must address the matter of learning from unbalanced training sets where the cover objects (normal images) always greatly outnumber the stego ones. But the research in unbalanced image steganalysis is seldom seen. This work just focuses on the problem of unbalance JPEG images steganalysis. In this paper, we propose a frame of feature dimension reduction based semi-supervised learning for high-dimensional unbalanced JPEG image steganalysis. Our method uses standard steganalysis features, and selects the confident stego images from the unlabeled examples by multiview match resampling method to rebalance the unbalanced training images. Furthermore, weighted Fisher linear discriminant (WFLD) is proposed to find the proper feature subspace where K-means provides the weight factor for WFLD in return. Finally, WFLD and K-means both work in an iterative fashion until convergence. Experimental results on the MBs and nsF5 steganographic methods show the usefulness of the developed scheme over current popular feature spaces.

1. Introduction

Steganography can be regarded as the study of the art and science of convert communication, which conceals the existence of communication. Image is very suitable for steganography due to its many attributes and JPEG images are used popularly, which make the research in the field of JPEG steganography has become active currently. Many steganographic techniques operating on JPEG images have been proposed, and MBs (Model-Based steganography) [1] and nsF5 (no-shrinkage F5) [2] are the typical embedding algorithms. People can easily access and use a variety of steganography tools, which brings a new threat to information security. Thereby, the study of effective anti-steganographic technique is an urgent task for the related field.

Steganalysis is the most important anti-steganographic technique, which detects whether the object contains secretly hidden data. Nowadays, with the complexity of steganographic algorithms increasing, image steganalysis needs to start designing the higher dimension feature space. Pevny and Fridrich [3] formed 274-dimensional features by merging Markov and DCT features, and then was extended to twice size by Cartesian calibration [4] to obtain the 548-dimensional PEV features, Kodovsky et al. [5] took advantage of co-occurrence matrix to model the distribution pair of DCT coefficients on a mode-by-mode basis and constructed 7850-dimensional $\mathcal{CF}^*$ features. Li et al. [6] calculated the co-occurrence matrices of DCT coefficients to designed 15,700-dimensional features. Furthermore, Kodovsk$\grave{y}$ and Fridrich [7] constructed multiple submodels based on the joint distribution of DCT coefficients and then constructed 11,255-dimensional DCT features by combining these submodels, and finally, used the Cartesian calibration [4] to construct 22,510-dimensional transform domain rich models JRM features (Cartesian-calibrated JPEG Rich Model).

Classification algorithms will suffer from performance degradation when the class distribution is unbalanced. Many works have been done in addressing the class imbalance problem in other many fields. Usually, the methods with the data perspective rebalance the training data prior to learning. Cost-sensitive learning tries to learn more characteristics of samples with the minority class by setting a high misclassification cost [8].

Class imbalance problem also exists in steganalysis generally, but the research in unbalanced image steganalysis is seldom seen. In real-world network, the vast majority of digital images are normal images with the exception of a few stego objects. Therefore, it is inevitable that the minority of stego images compared to cover images are mainly concerned. It is more practical significance to research unbalanced image steganalysis. Fig. 1 simply describes the application scenarios. Where, Alice and Bob represent the communication parties. It is assumed to intercept a few stego images, and then we select a large number of suitable cover images. We combine these stego images with the majority of cover images to construct imbalance training sets, and then detect the testing sets.

☆ This paper has been recommended for acceptance by M.T. Sun.

* Corresponding author.
E-mail address: fgr2082@aliyun.com (G. Feng).

http://dx.doi.org/10.1016/j.jvcir.2015.10.013
1047-3203/

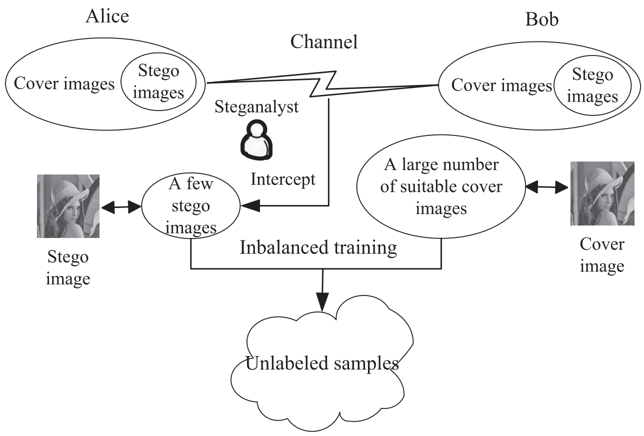

Fig. 1. The application scenarios of unbalanced image steganalysis.

This work focuses on the problem of unbalance JPEG images steganalysis. We construct two views of data set by subsampling the original feature space. The suspicious stego images which they agree in two views are selected. We add them to unbalanced training sets. WFLD (weighted Fisher linear discriminant) and K-means clustering are combined into a single framework that consider the clustering and the feature reduction simultaneously.

2. Steganalysis via weighted Fisher discriminative clustering

As we all know there exist many redundant features in high-dimensional spaces, so this paper proposes a method of feature dimensionality reduction for image steganalysis. The proposed approach first builds two views of the sets by sampling randomly features from original feature spaces, and each of two views of data is projected into the distinct feature subspace through WFLD, then cluster these sets by K-means in the feature subspace and get the weight factor of WFLD simultaneously. Later we select the part of the confident stego examples from the unlabeled data to rebalance the imbalance training sets. These modified data are used to train WFLD to obtain the suitable feature space where the K-means can implement well. Finally, WFLD and K-means work in an iterative fashion until convergence. The detailed process of the proposed algorithm is showed in Fig. 2. Process 1 is the schematic of the preprocessing of unbalance data by multi-view data match (see Section 2.2) and Process 2 is the diagram of iterative WFLDK algorithm (see Section 2.1).

2.1. Weighted Fisher linear discriminative K-means

Consider a data set $\{\mathbf{x}_i\}_{i=1}^{n} \in \mathbb{R}^m$ with n images, let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ denote the data matrix whose i-th column is given by $\mathbf{x}_i$. A typical steganalysis is a binary classification problem, thus, K-means clustering minimizes the following objective function [9]

$$\min\ J_K, J_K = \sum_{k=1}^{2}\sum_{i\in C_k}\|\mathbf{x}_i - \mathbf{u}_k\|^2 \tag{1}$$

where $\mathbf{u}_k$ is the mean of the k-th cluster C_k. C_1 can be denoted as the cluster of cover images, while C_2 can be denoted as the cluster of stego images. We also denote the vector $\mathbf{l} = \{0, 1\}^n$ as the cluster label. $l_i = 0$ if x_i belongs to the cluster C_1, and $l_i = 1$ is otherwise. In FLD, between-class scatter $\mathbf{S_b}$ and within-class scatter matrices $\mathbf{S_w}$ are defined as follows:

$$\mathbf{S_b} = (\mathbf{u}_1 - \mathbf{u}_2)(\mathbf{u}_1 - \mathbf{u}_2)^{\mathrm{T}}, \quad \mathbf{S_w} = \sum_{k=1}^{2}\sum_{i\in C_k}(\mathbf{x}_i - \mathbf{u}_k)(\mathbf{x}_i - \mathbf{u}_k)^{\mathrm{T}} \tag{2}$$

Thus, the K-means objective function can be transformed as

$$J_K = tr(\mathbf{S_w}) \tag{3}$$

where $tr(\mathbf{S_w})$ means the trace of the matrix $\mathbf{S_w}$. We can see that K-means clustering minimizes the within-class scatter. On the other hand, FLD objective function is

$$\max_{\mathbf{a}} \frac{tr(\mathbf{a}^{\mathrm{T}}\mathbf{S_b}\mathbf{a})}{tr(\mathbf{a}^{\mathrm{T}}\mathbf{S_w}\mathbf{a})} \tag{4}$$

where $\mathbf{a}$ is the projected vector. It is clear that K-means clustering has the similar properties as FLD to maximize the between-cluster scatter $\mathbf{S_b}$ and minimize the within-cluster scatter $\mathbf{S_w}$. Therefore, there should be ways to integrate them into a coherent framework [9,10]. Since the clustering labels are not entirely correct, we consider that the labels obtained by K-means have different credibility. It is not entirely reasonable that we treat equally in all examples. Therefore we develop a WFLD method to introduce weight factors which present the credibility metrics for different samples.

Let $\{g_i\}_{i=1}^{n} \in \mathbb{R}$ denote the weight factors and g_i represents the weight of $\mathbf{x}_i$. Since every instance has its own weight factor, the mean of k-th cluster must consider the weight and has to be redefined as shown in Eq. (5). Similarly, the between-cluster scatter and the within-cluster scatter also must consider the weight as well and can be redefined as Eqs. (6) and (7) using the weighted mean.

$$\mathbf{u}_k^{\mathbf{g}} = \frac{\sum_{i\in C_k} g_i\mathbf{x}_i}{\sum_{i\in C_k} g_i} \tag{5}$$

$$\mathbf{S_b^g} = (\mathbf{u}_1^{\mathbf{g}} - \mathbf{u}_2^{\mathbf{g}})(\mathbf{u}_1^{\mathbf{g}} - \mathbf{u}_2^{\mathbf{g}})^{\mathrm{T}} \tag{6}$$

$$\mathbf{S_w^g} = \sum_{k=1}^{2}\sum_{i\in C_k} g_i(\mathbf{x}_i - \mathbf{u}_k^{\mathbf{g}})(\mathbf{x}_i - \mathbf{u}_k^{\mathbf{g}})^{\mathrm{T}} \tag{7}$$

In order to avoid the problems with numerical instability in practice when $\mathbf{S_w^g}$ is singular or ill-conditioned. The regularization technique is commonly applied to solve the problem as follows:

$$\tilde{\mathbf{S}}_{\mathbf{w}}^{\mathbf{g}} = \mathbf{S_w^g} + \lambda\mathbf{I}, \quad \lambda > 0 \tag{8}$$

where λ is a regularization parameter which can be fixed to a small constant value and $\mathbf{I}$ is the identity matrix. Thus, WFLD is easy described using the generalized eigenvector

$$\tilde{\mathbf{a}} = (\mathbf{S_w^g} + \lambda\mathbf{I})(\mathbf{u}_1^{\mathbf{g}} - \mathbf{u}_2^{\mathbf{g}}) \tag{9}$$

We solve the optimal $\mathbf{g}$ when obtaining the eigenvector $\tilde{\mathbf{a}}$ and project the data matrix $\mathbf{X}$ into one-dimensional space to obtain the data vector $\mathbf{z} = \tilde{\mathbf{a}}^{\mathrm{T}}\mathbf{X}$, and then we use K-means clustering in the weighted Fisher discriminative subspace to achieve result. Once the cluster label vector $\mathbf{l}$ is solved, we can obtain the cluster centroid $\mathbf{u}$. Here, we consider that the smaller distance between the sample and the cluster centroid, the greater chance of sample is classified to the cluster. Therefore, we assure that the sample's weight is inversely proportional to the distance and the experimental results also prove that the hypothesis can achieve good effect. Then the weight is defined as follow:

$$g_i = \begin{cases} \dfrac{1}{1+\frac{\|z_i-u_1\|^2}{\|z_i-u_2\|^2}} & \text{if } z_i \in C_1 \\ \dfrac{1}{1+\frac{\|z_i-u_2\|^2}{\|z_i-u_1\|^2}} & \text{if } z_i \in C_2 \end{cases} \tag{10}$$

Therefore, when $\mathbf{g}$ is fixed, the eigenvector $\mathbf{a}$ using WFLD can be computed. Essentially, the optimization is alternatively carried out by fixing one of two components ($\mathbf{a}$ and $\mathbf{g}$) and then optimize the other. Our algorithm is called weighted Fisher linear discriminative K-means (WFLDK) and presented in Algorithm 1.