基于主成分分析和 RBF 网络的泛盲掩密分析方案*

陈 丹, 张键红, 王育民
(西安电子科技大学 ISN 国家重点实验室, 陕西 西安 710071)

摘 要: 针对 Farid 泛盲掩密分析方法所选的图像分类特征数目多而且具有相关性的缺陷, 采用主成分分析技术对特征进行去相关性的预处理, 并基于 RBF 网络提出了新的掩密分析方案。该方案不但大大降低了用于分类的图像特征的维数, 从而提高了掩密分析速度, 而且提高了掩密分析的检测性能。分别利用该方案和 Farid 的方案对 JSteg 等软件掩密后的图像进行检测, 比较实验结果表明, 对于不同长度的嵌入消息, 该算法具有更好的检测性能。

关键词: 掩密分析; 泛盲掩密分析; 主成分分析; RBF 网络
中图分类号: TN918 TP391 文献标识码: A 文章编号: 0529-6579 (2004) S2-0066-05

近几年, 信息隐藏学科 (information hiding) 得到了广泛的关注, 其重要分支——掩密术 (又称隐写术, steganography) 的发展也在不断地深入。掩密术[1]是一种隐蔽通信技术, 其目的在于隐藏秘密信息的存在, 而不像密码术那样仅隐藏秘密信息的内容。勿庸置疑, 掩密术开辟了信息安全的新思路, 弥补了传统加解密系统的一些缺陷。它的发展, 使得我们可以利用它来进行敏感信息的秘密传递而不引起别人的注意, 但是与此同时, 敌人也可能利用同样的方法, 进行着非法的通信。因此, 掩密术的对抗技术——掩密分析技术 (steganalysis) 也越来越显示出举足轻重的地位。

掩密分析[2]是利用掩密术固有的弱点, 即信息的嵌入会引起载体数据分布特性或统计特性的改变, 分析各种可能的载体信息, 从而检测、提取或破坏隐藏的秘密消息。目前多数的掩密分析方法都是针对某种特定的掩密算法的。例如, 对于掩密术中最常见的 LSB (least significant bits) 算法, 许多学者都提出了相应的分析对策[3-5]。这些基于某一种掩密算法的检测方法虽然能够达到较准确的分析结果, 但是灵活性和可扩展性都很差。由于隐藏学科自身的特点, 掩密算法可能并不公开, 而且掩密算法将层出不穷, 所以无法得知要攻击算法的具体细节。Farid[6,7]首先将掩密分析看成是两类模式的识别问题, 即通过合理选择特征和设计分类器来区分没有隐藏消息的载体图像和隐藏了消息的掩密图像, 并由此提出泛盲掩密分析 (universal blind steganalysis) 的概念。这一概念的提出, 使得可以针对某一类而不是某一个掩密算法来检测隐藏消息的存在, 从而大大提高了掩密分析算法的适用范围和检测效率。然而, Farid 的算法所选的分类特征不但数目多而且具有相关性。该相关性使得所观测的数据在一定程度上有信息的重叠, 从而不利于分类器的设计和性能。为此, 提出采用主成分分析技术[8]对特征进行预处理, 以去除它们之间的相关性, 并基于 RBF 网络[9]构造掩密分析分类器, RBF 网络在解决非线性分类问题中具有收敛速度快、不易陷入局部极小点、鲁棒性好等优点。分别利用本文算法和 Farid 的算法对 JSteg、EZStego 和 S-Tools[10-12]掩密后的图像进行检测, 实验结果表明, 对于不同长度的嵌入消息, 该算法较 Farid 的算法都具有更好的检测性能。

1 Farid 的泛盲分析算法及其缺陷

Farid[6,7]的算法首先对图像进行可分离的正交镜像滤波器 (QMFs) 分解, 然后选取两组统计量作为区分载体图像和掩密图像的分类特征矢量。第 1 组统计量取自不同分解级和不同分解方向 (水平、垂直和对角) 上每一子带系数的均值、方差、偏度和峰度, 这些统计量描述了基本的系数分布。第 2 组统计量是基于对系数幅值的最优线性预测误差。由这一误差得到了另一组统计量, 即误差的均值、方差、偏度和峰度。

一般而言, 分类特征选取的好坏将直接影响到

* 收稿日期: 2004-09-02
基金项目: "十五" 军事通信预研项目 (41001040303)
作者简介: 陈 丹 (1976 年生), 女, 博士生; E-mail: chenndann@hotmail.com.

分类或检测性能[9]，因为所选择的特征是模式分类或检测的依据，它们会被作为样本参与分类器的设计和实现。Farid 的算法共选取了 12($n-1$) 个小波系数统计量和 12($n-1$) 个小波系数误差统计量作为分类特征矢量，n 是小波分解级数。为了获取足够多的分类信息，n 不能太小，如 n 取 4，则分类特征矢量的维数将达到 72，这么多的特征将会导致分类器十分复杂。更重要的是，它的检测效率也不高，因为它忽略了各特征之间存在的相关性。这些特征都是分别从不同的分解级和分解方向的子带中得出的，而小波分解的子带之间具有很强的相关性，所以这些特征之间也必然具有相关性，该相关性使得所观测的数据在一定程度上有信息的重叠，从而不利于分类器的设计和性能。这里，为证明各特征之间存在有相关性，取 100 幅由数码相机拍摄的自然图像（剪裁成 640×480），分别计算它们的小波系数统计量和小波系数误差统计量构成特征矢量，然后求取矢量相关系数，图 1 给出了部分数据，而且经计算得，相关系数中绝对值大于等于 0.5 的个数约占总体系数的 10%。由此可见，我们有必要对原始数据进行去相关性的处理，使得处理后的样本数据之间没有信息的混叠，从而简化分类器的设计并最终提高检测性能。

−0.5935	−0.1314	0.2986	0.2737	−0.0660	−0.2065
−0.6557	−0.2136	0.3158	0.3551	−0.0068	−0.2232
−0.0414	−0.5568	0.6316	0.0092	0.4396	−0.5020
0.8011	0.1427	−0.2526	−0.5406	0.0750	0.2078
0.7029	0.6681	−0.4871	−0.5676	−0.3810	0.4440
0.1688	0.2789	−0.1079	−0.1538	−0.2532	0.5275
−0.8864	−0.2730	0.2524	0.7245	0.0578	−0.2422
−0.5861	−0.7823	0.5033	0.5847	0.5376	−0.4836
0.0610	0.2242	−0.4677	−0.0007	−0.1160	0.0398
0.2841	0.0932	−0.2737	−0.1719	−0.0374	0.3642
0.1602	0.0114	−0.2368	−0.0310	−0.0014	0.3343
0.1412	0.5252	−0.6557	−0.0633	−0.3918	0.5226
−0.5706	−0.1009	0.2761	0.2460	−0.0788	−0.2104
−0.6433	−0.2293	0.3428	0.3288	0.0118	−0.2708
−0.1501	−0.5567	0.6622	0.0724	0.4157	−0.5368
1.0000	0.3301	−0.2132	−0.9108	−0.1039	0.2088
0.3301	1.0000	−0.4466	−0.3471	−0.9194	0.4690
−0.2132	−0.4466	1.0000	0.1409	0.3100	−0.8536
−0.9108	−0.3471	0.1409	1.0000	0.1697	−0.1633
−0.1039	−0.9194	0.3100	0.1697	1.0000	−0.3685
0.2088	0.4690	−0.8536	−0.1633	−0.3685	1.0000

图 1　部分矢量相关系数

Fig. 1　Part of the correlation coefficients of vectors

为此，我们提出对原始样本数据进行主成分分析预处理，该技术在对数据降维处理的同时还简化了原始数据的统计数字特征，不但简化了分类器设计而且能够提高总体的检测性能。下面将详细介绍该掩密分析方案。

2　掩密分析方案

2.1　系统模型

掩密分析方案分为学习和判决两个部分，如图 2 所示。学习过程是从大量的训练图像（包括载体图像和掩密图像两类）中提取并选择有利于分类的特征矢量，以此构造掩密分析分类器并对分类器进行训练，直到满足一定的精度要求。判决过程是利用学习过程中建立的分类器对被测图像进行分类。这样，即使在不知道掩密算法的情况下，也可以通过比较和分析学习样本，观察出由于消息嵌入所带来的数据特性变化，从而提取出可用于分类的特征并设计分类器，最终获得图像掩密与否的信息。

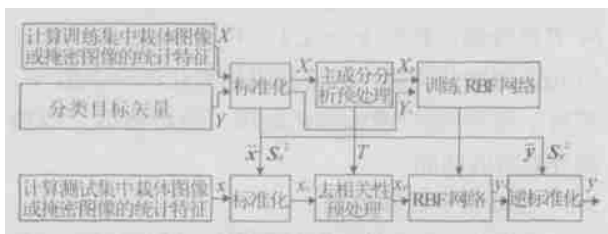

图 2　掩密分析系统模型

Fig. 2　System model of steganalysis

学习过程主要有以下几个步骤：

(1) 根据 Farid 的算法分别计算训练集中载体图像和掩密图像的特征矢量，将两组特征矢量合并构成原始特征矩阵

$$\boldsymbol{X}=\begin{bmatrix} x_{1,1} & \cdots & x_{1,n_1} & x_{1,n_1+1} & \cdots & x_{1,n_1+n_2} \\ x_{2,1} & \cdots & x_{2,n_1} & x_{2,n_1+1} & \cdots & x_{2,n_1+n_2} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{p,1} & \cdots & x_{p,n_1} & x_{p,n_1+1} & \cdots & x_{p,n_1+n_2} \end{bmatrix}$$

n_1 为载体图像数，n_2 为掩密图像数，$n_1+n_2=n$ 为总体样本数，p 是特征数。

(2) 建立分类目标矢量

$$Y=[\underbrace{0,\cdots,0}_{n_1\text{个}0},\underbrace{1,\cdots,1}_{n_2\text{个}1}]$$

前面 n_1 个“0”代表载体图像，后面 n_2 个“1”代表掩密图像。

(3) 对 $\boldsymbol{X}$ 进行标准化处理，形成标准化特征矩阵 $\boldsymbol{X}_S$，即

$$x_{Sk}=\frac{x_k-\bar{x}_k}{\sqrt{S_{Xk}^2}},\ k=1,\cdots,p$$

其中，$\bar{x}_k=\frac{1}{n}\sum_{i=1}^{n}x_{ik}$，$S_{Xk}^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_{ik}-\bar{x}_k)^2$。因为主成分分析的每个主成分依赖于测量初始变量所用的尺度，当尺度改变时，会得到不同的特征值。克服这个困难的方法是对初始变量进行标准化处理，使其均值为 0，方差为 1。同理，对 $\boldsymbol{Y}$ 也进行标准化处理生成 $\boldsymbol{Y}_S$。

(4) 对 $\boldsymbol{X}_S$ 进行主成分分析，形成主成分特征矩阵 $\boldsymbol{X}_P$，并保存转换矩阵 $\boldsymbol{T}$。

(5) 将处理后的数据 $\boldsymbol{X}_P$ 和分类信息 $\boldsymbol{Y}_S$ 输入初始的 RBF 网络，训练网络的权值参数，直到达到目标误差或训练次数最大值为止。

判决部分主要是计算测试集中图像的特征矢量 $\boldsymbol{x}$，然后利用预先计算的样本均值 $\bar{\boldsymbol{x}}$ 和样本方差 $\boldsymbol{S}_x^2$ 进行标准化变换，形成 $x_s=\frac{x-\bar{x}}{\sqrt{S_X^2}}$，然后经转换矩阵 $\boldsymbol{T}$ 预处理，获得 $\boldsymbol{x}_p=\boldsymbol{x}_s\boldsymbol{T}$，再将 $\boldsymbol{x}_p$ 输入设计好的 RBF 网络分类器，网络输出标准化的分类值 y_s，需用 $\bar{y}$ 和 S_Y^2 进行逆标准化处理，最后输出 y 以检测分类器的性能。

2.2 主成分分析预处理

主成分分析[8]是在保证数据信息损失最小（最小二乘准则）的前提下，经线性变换和舍弃一小部分信息，以少数新的综合变量取代原始的多维变量的技术。预处理的具体过程为：

(1) 计算标准化样本矩阵 $\boldsymbol{X}_S$ 的相关系数矩阵。

$$R=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$$

(2) 求特征方程 $|\boldsymbol{R}-\lambda\boldsymbol{I}|=0$ 的 p 个非负特征值 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p\geqslant 0$，对应于特征值的特征向量为

$$C^{(i)}=(C_1^{(i)},C_2^{(i)},\cdots,C_p^{(i)})\mathrm{T},\ \mathrm{i}=1,\cdots,\mathrm{p}$$

并且有

$$C^{(i)\mathrm{T}}C^{(j)}=\sum_{k=1}^{p}C_k^{(i)}C_k^{(j)}=\begin{cases}1\ (i=j)\\ 0\ (i\neq j)\end{cases}$$

(3) 计算 $m(m<p)$ 个主成分。求取当方差比例 $\alpha=\left(\sum_{i=1}^{m}\lambda_i\right)\Big\backslash\left(\sum_{i=1}^{p}\lambda_i\right)$ 接近于 1 时（一般取 $\alpha\geqslant 0.85$）的 m，将前 m 个特征值 $\lambda_1,\lambda_2,\cdots,\lambda_m$ 对应的特征向量 $C^{(1)},C^{(2)},\cdots,C^{(m)}$ 组成变换矩阵 T，计算主成分特征矩阵 $X_{pm\times m}=X_{sn\times p}T_{p\times m}$。

$$X_P=(x_{p1},x_{p2},\cdots,x_{pm})\ (m<p)$$

包含了原特征数据的 m 个主成分，其方差和占全部总方差的 85%以上，基本上保留了原来特征的信息，而特征数目由 p 个减少至 m 个，从而达到筛选特征的目的。

2.3 RBF 网络设计

分类器采用 RBF 网络[9]，即径向基函数网络，是一种较常用的前馈型神经网络。它可以实现对任意非线性函数的逼近，而且用于模式分类具有收敛速度快、不易陷入局部极小点、鲁棒性好和易于实现等优点。由于掩密分析属于两类模式识别问题，所以相应的 RBF 网络结构比较简单，如图 3 所示。图中 $\boldsymbol{x}$ 为输入矢量，$\boldsymbol{n}$ 表示隐层输出矢量，$\boldsymbol{w}'_i$ 表示第 i 个隐层神经元的权值矢量，即隐层神经元权值矩阵 $\boldsymbol{W}^1$ 的第 i 行，$\boldsymbol{b}^1$ 是隐层的阈值矢量，$\boldsymbol{W}^2$ 是输出权值矢量，$\boldsymbol{b}^2$ 是输出阈值。网络的输入维数为 m、隐层神经元个数为 s 个、输出个数为 1，隐层神经元采用高斯函数作为传递函数，输出层的传递函数为线性函数。根据训练样本采用迭代方法确定隐层神经元的个数，即每迭代一次就增加一个神经元，直到平方和误差下降到目标误差以下或者隐层神经元个数达到最大值时迭代停止。输出单元的权值直接由最小二乘法计算。

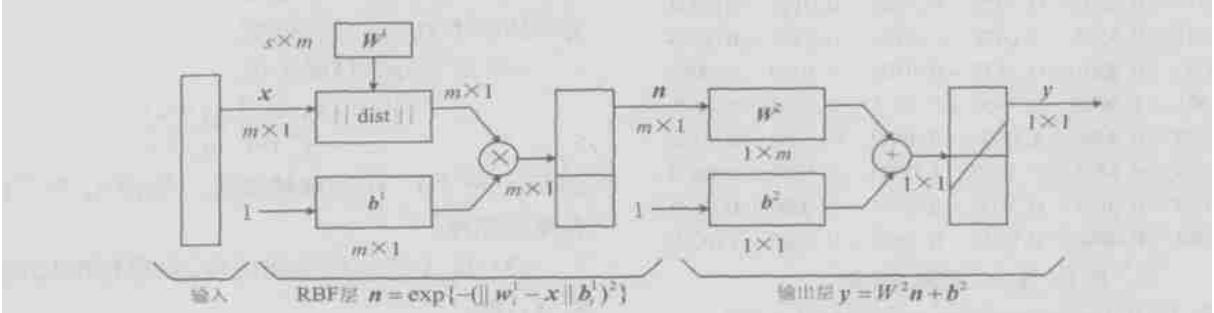

图 3 基于 RBF 网络的掩密分析分类器结构

Fig. 3 Structure of steganalysis classifier based on RBF network

3　仿真实验

将分别测试上述算法对 Jsteg、EzStego 和 S-Tools 软件[10—12]的检测能力。采用由数码相机拍摄的自然图像（剪裁成 640×480）进行仿真实验。取 500 幅这样的图像作为未嵌消息的载体图像，存为 JPEG（或 GIF 或 BMP）格式，对每幅图像进行 4 级小波分解，计算 72 维特征矢量。然后用 Jsteg（或 EzStego 或 S-Tools）软件在每幅图像中嵌入消息，形成掩密图像，并对每幅掩密图像计算 72 维特征矢量。这样，1000 幅图像的特征矢量构成了训练样本集。类似地，另取 1000 幅图像构成测试样本集。

经主成分分析预处理后，样本集特征矢量的维数从 72 分别降到了 15（Jsteg）、18（EzStego）和 12（S-Tools），特征数目的减少大大简化了分类器的设计。表 1 给出了该算法和 Farid 算法的检测性能比较。从实验数据上看，本文算法的检测性能对 Farid 的算法有了较大的改善。

表 1　本文算法和 Farid 算法的检测性能比较

Tab. 1　Comparison of detection performance between Farid's algorithm and our scheme

软件	消息大小	本文算法			Farid 算法		
		JPEG	GIF	BMP	JPEG	GIF	BMP
Jsteg	256×256	96.4			92.0		
	128×128	96.0			92.7		
	64×64	90.7			85.3		
EzStego	194×194		64.3			44.2	
	128×128		37.1			10.8	
	64×64		11.6			1.9	
S-Tools	256×256			96.9			97.0
	128×128			89.6			70.7
	64×64			61.8			46.1

4　结　论

本文在 Farid 的泛盲掩密分析思想的基础的上，提出了基于主成分分析预处理和 RBF 网络掩密分析方法。主成分分析不仅可以对原始特征数据进行降维处理，还简化了原始数据的统计数字特征，同时提供了很多对分类极其重要的信息。而分类器采用了 RBF 网络，是因为 RBF 网络具有收敛速度快、不易陷入局部极小点、鲁棒性好等优点。我们分别测试了该方案对 Jsteg、EzStego 和 S－Tools 软件的检测能力，实验结果表明，其性能较 Farid 算法有了大幅度的提高。

从实验结果可进一步得出，一旦确定了所使用的分类器，泛盲掩密分析的性能将主要取决于特征的选择。所以，在进行掩密分析时，对于不同类型的掩密算法应选择相应的图像特征，以获得尽可能高的检测效率。例如，针对空间域的掩密算法，所选的特征应该能够反映载体图像和掩密图像在空间特性上的差异，而对于频率域的掩密算法，则应选择反映两者之间频率变化的特征。

参考文献：

[1] PROVOS N, HONEYMAN P. Hide and seek: an introduction to steganography[J]. IEEE Security & Privacy, 2003: 32—44.

[2] FRIDRICH J, GOLJAN M. Practical steganalysis —— state of the art[C]. Proc SPIE Photonics West, Electronic Imaging 2002. Security and Watermarking of Multimedia Contents, 2002: 1—13.

[3] WESTFELD A, PFITZMANN A. Attacks on steganographic systems[C]. 3rd Int'l Workshop Information Hiding, Ney York: Springer-Verlag, 1999: 61—76.

[4] FRIDRICH J, GOLJAN M, DU R. Detecting LSB steganography in color and gray-scale images[C]. Magazine of IEEE Multimedia, Special Issue on Security, 2001: 22—28.

[5] ZHANG T, PING X J. Reliable detection of LSB steganography based on the difference image histogram[C]. ICASSP'2003, 2003: 545—548.

[6] FARID H. Detecting hidden messages using higher-order statistical models[C]. Int'l Conf Image Processing, 2002.

[7] LYU S, FARID H. Detecting hidden messages using higher-order statistics and support vector machines[C]. 5th Int'l Workshop on Information Hiding, New York: Springer-Verlag, 2002: 340—354.

[8] 任若恩, 王惠文. 多元统计数据分析——理论、方法、实例[M]. 北京: 国防工业出版社, 1997.

[9] 边肇祺, 张学工, 等. 模式识别[M]. 第 2 版. 北京: 清华大学出版社, 2000.

[10] DEREK U. JPEG-JSteg-V4. http://www.funet.fi/pub/crypt/steganography/jpeg—JSteg—v4.diff.gz.

[11] ROMANA MACHADO. EZStego. http://www.stego.com

[12] ANDREW BROWN. S-Tools 4. http://www.snapfiles.com/get/stools.html

A Universal Blind Steganalysis Based on Principal Component Analysis and RBF Network

CHEN Dan, ZHANG Jian hong, WANG Yu min
(State Key Laboratory on ISN, Xi'an University of Electronical Technology, Xi'an 710071, China)

Abstract: On the basis of the limitation of Farid's universal blind steganalysis that image features for the classification are too many and have correlations, a new universal blind steganalysis scheme is presented. Decorrelation preprocessing using principal component analysis on the image statistics features and steganalysis classifier using RBF network are proposed. The scheme not only reduces the dimension of the feature vector enormously so that the speed of steganalysis detection is increased, but also the performance of the detection is improved as well. Applying this scheme and Farid's scheme to detecting the stego images produced by JSteg, EZStego, and S-Tools respectively, the comparison of these simulation results shows that this scheme is quite efficient for different size of messages embedded.

Key words: steganalysis; universal blind steganalysis; principal component analysis; RBF network

---

(上接第 65 页)

Steganalysis Based on Contamination Data Analysis

GUAN Wei[1,2], *ZHANG Wei ming*[1,2], *LIU Wen fen*[1]
(1. Institute of Information Engineering, Information Engineering University, Zhengzhou 450002, China;
2. State Key Laboratory of Information Security,
Graduate School of the Chinese Academy Sciences, Beijing 100039, China)

Abstract: Steganalysis is the important branch of information hiding, a large number of steganographic programs use the least significant bit (LSB) embedding as the method of choice for message hiding in color images, and grayscale images. In this paper, based on contaminated data analysis, a new steganalytic technique capable of a reliable detection of spatial LSB steganography is proposed. And this method can not only fast and reliably detect the existence of hiden message embedded in images, but also accurately estimate the amount of hiden message embedded by sequentially or randomly scattered algorithm.

Key words: information hiding; steganography; steganalysis; LSB; contamination data analysis