

# Universal Steganalysis for Image Based on Genetic Algorithm and Grey SVC

Yuehong Wu, Hongbo Wang

Troops 65067 PLA, Shenyang Liaoning  
Email: zhuxg11@chinaunicom.cn

Received: Sep. 2<sup>nd</sup>, 2017; accepted: Sep. 16<sup>th</sup>, 2017; published: Sep. 22<sup>nd</sup>, 2017

---

## Abstract

The isolated samples can produce some affect on distinguishing the best classifying plane, which becomes one of causes of less performance of universal steganalysis that uses Support Vector Classifier (SVC) as classifier. This paper proposes a new universal steganalysis algorithm. The algorithm firstly catches characteristic of noise signal in wavelet domain. Then utilizes Gentic Algorithm (GA) search the best characteristic of species. Finally makes grey relational degree between sample characteristic and the best characteristic of species participate in training of SVC, thus constructs a Grey Support Vector Machines (GSVM) to be a classifier.

## Keywords

Steganalysis, Genetic Algorithm, Grey Relational Analysis, Support Vector Machines

---

## 基于GA和灰色SVC的图像通用密写分析

武月红, 王洪波

中国人民解放军65067部队, 辽宁 沈阳  
Email: zhuxg11@chinaunicom.cn

收稿日期: 2017年9月2日; 录用日期: 2017年9月16日; 发布日期: 2017年9月22日

---

## 摘要

在以支持向量分类机(Support Vector Classifier, SVC)为分类器的图像通用密写分析中, 离群样本对最优分类面判别的影响成为导致分析性能不高的原因之一。文中提出一种新的密写分析算法。该算法首先在小波域上提取噪声信号的特征, 然后利用遗传算法(Genetic Algorithm, GA)搜索最优类特征, 最后将样本特征与最优类特征的灰色关联度参与到SVC的训练中, 构造一个灰色支持向量机作为分类器。与采用

相同特征的Holotyak算法相比, 文中所提的通用密写分析算法的分类结果更好。

## 关键词

密写分析, 遗传算法, 灰色关联分析, 支持向量机

Copyright © 2017 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

密写分析的目的在于揭示媒体中秘密信息的存在性。目前有很多针对特定隐藏方法的密写分析法, 然而随着密写的不断发展, 方法多样性使得对隐藏信息的检测变得非常困难, 其结果是密写分析的发展永远落后于密写。显然, 仅持有可能含密的图像、对可能使用的密写算法全然不知的完全盲分析是密写分析的重要研究内容。这种密写分析方法不局限于单一的密写算法, 通常对几种密写算法都适用, 所以又称通用密写分析。

著名的通用密写分析技术, 例如 I. Avcibas 等[1]提出的利用图像最低位平面和次低位平面的多项二值相似性测试作为特征空间, 采用 SVC 分类的算法。Siwei 等[2]进一步采用 QMF 分析图像小波系数及其预测误差的高阶统计量, 对一类支持向量机进行训练。T. Holotyak 等[3]提出一种新的在图像小波域上提取噪声成分特征, 用两类 SVC 进行分类的方法。这些技术的相同之处是都用支持向量机作为分类器, 但是在支持向量机方法中, 确定最优分类超平面的支持向量位于类边缘, 而离群样本往往也位于类边缘附近, 就这个问题, 线性不可支持向量分类机引入松弛变量来容许一定数量的错分类样本的存在, 这些错分样本会影响最优分类面的判别, 从而影响分类器的性能, 成为分析性能不高的原因之一。

文中将提出一种新的以 GSVM 为分类器的图像通用密写分析方法。首先选取 T. Holotyak 通用密写分析中的 33 个噪声高阶统计量作为特征向量, 利用遗传算法对训练样本进行搜索, 得到最优类特征向量, 然后运用灰色关联分析求出不同训练样本的特征向量与最优类特征向量的关联程度, 将其参与到支持向量机的训练中, 以减小离群样本的分类误差在支持向量机目标函数中的影响, 称经过这样训练的支持向量机为 GSVM, 最后运用训练好的 GSVM 进行分类检测。

## 2. 遗传算法和灰色关联分析

最早系统地提出 GA 的是 J. H. Holland [4]. GA 将问题的可能解集通过基因编码构成初始种群(父代)开始迭代, 计算父代个体的适应度, 如果满足优化准则, 就跳出迭代; 否则, 通过选择、交叉或基因重组、变异等操作形成新的个体(子代), 将子代插入到父代中, 重新开始循环, 直到满足优化准则, 使种群进化到包含近似最优解的状态。

找到最优解后, 可运用灰色关联分析(Grey Relational Analysis, GRA)计算父代中可能解与其的灰色关联度。灰色关联度的计算描述如下[5]:

设参考序列

$$x_0 = (x_0(1), x_0(2), \dots, x_0(n)) \quad (1)$$

其中  $n$  表示数组中元素的个数。待比较序列

$$x_i = (x_i(1), x_i(2), \dots, x_i(n)) \tag{2}$$

其中  $i$  表示待比较序列的个数, 待比较序列中元素的个数与参考序列中的相同。  $x_i$  对  $x_0$  在  $k$  点的灰色关联系数为

$$\xi(x_0(k), x_i(k)) = \frac{\Delta_{\min} + \rho\Delta_{\max}}{\Delta_{0i}(k) + \rho\Delta_{\max}} \tag{3}$$

其中  $\Delta_{\min} = \min_k \min_i |x_0(k) - x_i(k)|$ ,  $\Delta_{\max} = \max_k \max_i |x_0(k) - x_i(k)|$ ,  $\Delta_{0i}(k) = |x_0(k) - x_i(k)|$ ;  $\rho \in (0, 1]$  是事先设定的常数, 称为分辨率系数, 一般情况下均取  $\rho = 0.5$ 。灰色关联度为

$$\gamma(x_0, x_i) = \frac{1}{n} \sum_{k=1}^n \xi_{0i}(k) \tag{4}$$

### 3. 基于 GSVM 的密写分析方法

#### 3.1. 特征提取

图像进行 1 层 db8 小波变换后, 利用 M. K. Mihcak 等[6]提出的去噪滤波器可以获得噪声信号。根据文献[3], 噪声信号由其概率分布函数(Probability Distribution Function, PDF)唯一确定。将 PDF 记为  $p$ , 应用指数核  $e^{tx}$ , 并进行 Taylor 展开, 可将  $p$  映射到一个参数空间, 如下式:

$$M(x) = \int_{-\infty}^{\infty} p(t)e^{tx} dt = \int_{-\infty}^{\infty} p(t) \left( 1 + tx + \frac{t^2 x^2}{2!} + \dots + Q_n(t) \right) dt = 1 + m_1 x + \frac{m_2}{2!} x^2 + \dots + R_n(x) \tag{5}$$

上式作为小波域中噪声信号的多项式表示, 在统计上, 被当作矩生成函数, 可通过其系数矩阵求得, 其中  $m_i$  是  $i$  阶矩,  $Q_n$  和  $R_n$  随着  $n$  的增大, 趋近于 0。所以, 可以将矩作为其特征, 而又可避免求  $p$  的复杂性。通过大量的实验, 最终, 在子带内, 选择经二阶矩的标准差  $\sqrt{m_2}$  标准化后的 11 个矩作为特征向量, 即  $m'_i = \frac{m_i}{\sqrt{m_2}}$ ,  $i = 4, 6, \dots, 24$  (经文献[6]中的去噪滤波器所得噪声信号的 PDF 具有奇偶对称性), 在  $LH, HL, HH$  三个子带内计算, 可得一个 33 维的特征向量。

#### 3.2. 最优类特征产生

将训练样本中的正类样本(含密图像样本)  $T_i (i = 1, 2, \dots, n)$  的特征向量, 作为 GA 的初始种群, 采用实数编码方式。设样本  $i$  的代价函数为

$$\Psi_i = 1 - \frac{1}{n} \sum_{j=1}^n \gamma(T_i, T_j) \tag{6}$$

其中  $\gamma(T_i, T_j)$  表示种群中样本  $i$  与样本  $j$  的灰色关联度。  $\Psi_i$  越小, 说明样本  $i$  与种群中的其余样本越相似, 越能代表整个种群。所有样本的代价函数值构成种群的代价函数集合  $\Psi = \{\Psi_1, \Psi_2, \dots, \Psi_n\}$ 。采用 rank-scale [7]方法, 对集合中的代价函数值按升序排列后, 排在第  $j$  位的样本的适应度为

$$\phi_j = \frac{n}{\sqrt{j} \sum_{i=1}^n \frac{1}{\sqrt{i}}} \tag{7}$$

利用上式, 计算种群中所有样本的适应度。根据适应度进行轮盘赌的适应机制, 被选中的个体进行单点交叉和变异的遗传操作, 形成中间种群, 并将其与父代种群合并, 计算合并种群中所有样本的代价函数和适应度, 将适应度大的前  $n$  个样本作为下一代种群, 开始下一代的迭代。当种群中含有  $\Psi_i$  小于 0.05

的样本时, 循环结束, 此样本的特征向量为最优类特征向量; 或者满足别的条件, 循环结束时, 代价函数最小的样本特征为最优类特征。

### 3.3. GSVM 分类器构造

将上面产生的最优类特征记为  $x_0$ , 则其与训练样本  $i$  的灰色关联度为  $\gamma(x_0, x_i)$ , 简记为  $\gamma_0(x_i)$ , 引入  $\gamma_0(x_i)$  后, 训练样本集可表示为:

$$(x_1, y_1, \gamma_0(x_1)), (x_2, y_2, \gamma_0(x_2)), \dots, (x_l, y_l, \gamma_0(x_l)) \quad (8)$$

其中,  $x_j \in R^n$  为样本的特征向量,  $y_j \in \{-1, 1\}$  为类标识,  $0 < \gamma_0(x_j) \leq 1$  是样本特征与类特征的灰色关联度,  $j=1, \dots, l$ ; 并设  $\phi: x \subseteq R^n \rightarrow x \subseteq H$  是从原始特征空间  $R^n$  到一个高维特征空间  $H$  的变换。则文献[8]中非线性分类问题转化为求解如下以  $(w, b, \varepsilon)^T$  为决策变量的二次规划问题:

$$\begin{cases} \min_{w, b, \varepsilon} \frac{1}{2} \|w\|^2 + C \left( \sum_{j=1}^l \gamma_0(x_j) \varepsilon_j \right) \\ \text{s.t. } y_j \left( (w^T, \phi_j) + b \right) + \varepsilon_j \geq 1, j=1, 2, \dots, l \\ \varepsilon_j \geq 0, j=1, 2, \dots, l \end{cases} \quad (9)$$

其中,  $C > 0$  是一个自定义的惩罚因子,  $\varepsilon_j$  是支持向量机目标函数中的分类误差, 由(9)式看出: 小的  $\gamma_0(x_j)$  可以减小  $\varepsilon_j$  在其中的影响, 也可认为相应的  $x_j$  是不重要的样本。

引入核函数  $K(x_j, x_i) = \phi(x_j) \phi(x_i)$ , 则高维空间上的内积运算只需在原空间上进行, (9)式的对偶规划式变为:

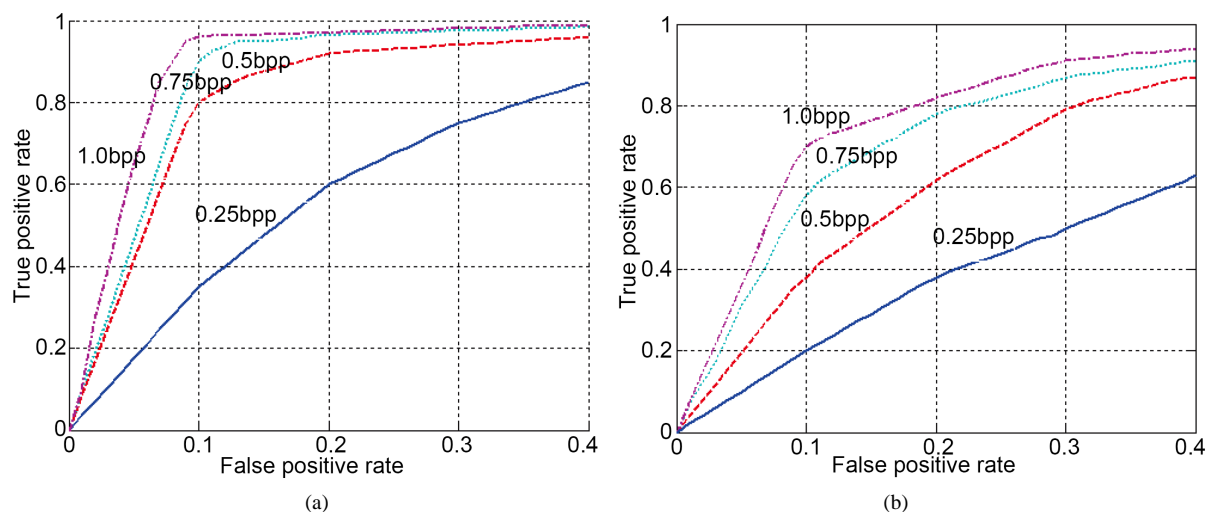
$$\begin{cases} \min_a \frac{1}{2} \sum_{j=1}^l \sum_{i=1}^l a_j a_i K(x_j, x_i) - \sum_{j=1}^l a_j \\ \text{s.t. } \sum_{j=1}^l a_j y_j = 0 \\ 0 \leq a_j \leq \gamma_0(x_j) C, j=1, 2, \dots, l \end{cases} \quad (10)$$

这种将最优类特征与训练样本的灰色关联度  $\gamma_0(x_j)$  引入到样本中, 经训练所得的支持向量分类机称为灰色支持向量分类机。

## 4. 实验结果分析

通过对文中所提算法和文献[3]中算法的结果进行比较, 来分析算法的性能。公平起见, 要在相同的条件下进行。所以与文献[3]一样, 实验图像集也来源于文献[9], 由 2375 幅尺度为  $1500 \times 2100$  像素、存储格式为 TIFF 的 32 位彩色扫描图像组成。同样, 也将样本图像转换为灰度图像。嵌入方法是经典的 LSB 方法, 嵌入率分别为 0.25 bpp (bits per pixel), 0.5 bpp, 0.75 bpp 和 1.0 bpp。图 1(a)和图 1(b)分别给出了文中算法和文献[3]中算法在上述图像集上的 ROC (Receiver Operating Characteristic) 分类性能曲线。从图中可以看出, 文中所提算法的性能要好于文献[3]中算法的性能, 即相同正确接受率(True positive rate)下的错误接受率(False positive rate)较小。这是因为在对分类器进行训练时引入了正类样本(含密图像)的最优类特征与训练样本的灰色关联度, 减小了负类样本(原始图像)中的离群点, 即靠近正类样本的原始图像在支持向量机目标函数中分类误差的影响, 所以使得将负类样本当作正类样本的概率减小。

下面从最终的分类结果来衡量文中所提算法的性能。同样是上述的图像集, 把这些图像按 4:1 分成两组并各复制四份, 用经典的 LSB 方法分别以嵌入率为 0.25 bpp, 0.5 bpp, 0.75 bpp 和 1.0 bpp 给四份复



**Figure 1.** Steganalysis result of different embedded capacity: (a) experiment result of proposed algorithm; (b) experiment result of Holotyak's method

**图 1.** 不同嵌入容量的隐写分析实验结果: (a) 文中方法实验结果; (b) Holotyak 方法实验结果

**Table 1.** True positive rate of original images and stegano images

**表 1.** 原始图像和含密图像的正确分类结果

Embedding rate (bpp)	Original images (%)		Stegano images (%)	
	Proposed algorithm	Holotyak's Algorithm	proposed algorithm	Holotyak's Algorithm
0.25	75%	70%	65%	48%
0.50	82%	75%	89%	68%
0.75	85%	78%	94%	78%
1.00	86%	79%	96%	80%

制图像嵌入秘密信息。实验中, 用没有嵌入秘密信息的第一组图像和一份经复制并嵌入了秘密信息的第一组图像组成训练样本集来训练上面构造的 GSVM, 然后用训练好的 GSVM 对没有嵌入秘密信息的第二组图像和经复制并以相同嵌入率嵌入了秘密信息的第二组图像组成测试集来测试, 其余三份嵌入秘密信息的图像也进行相同的运用, 共得到四组测试数据。表 1 给出了上述实验中原始图像和含密图像各自的正确检测率与文献[3]中方法正确检测率的对比。从表中同样可以看出, 文中所提算法的性能较文献[3]中的要好。原因是在对分类器训练时削弱了离群点的作用。

## 5. 结论

文中提出了一种新的基于 GA 和 GSVM 的图像通用密写分析算法。在对 SVM 训练时, 引入了样本与类特征的灰色关联度, 构造了 GSVM, 减小了离群点的作用。另外, 在计算灰色关联度时, 采用 GA 来搜索最优类特征作为参考序列, 使得计算出的灰色关联度能真正反映出样本与类的关联程度。最后, 用训练好的 GSVM 进行检测。实验结果表明, 在相同条件下, 文中提出的通用密写分析算法的错误接受率更低, 具有更高的正确检测率。但文中仅对 Holotyak 的噪声特征进行了测试, 下一步研究的重点是寻求更好的分类特征, 对所构造的 GSVM 进行训练, 形成更有效的通用密写分析算法。

## 参考文献 (References)

- [1] Avcibas, I., Memon, N., Sankur, B., *et al.* (2005) Image Steganalysis with Binary Similarity Measures. *EURASIP*

---

*Journal on Advances in Signal Processing*, **17**, 1-9.

- [2] Siwei, L. and Farid, H. (2004) Steganalysis Using Color Wavelet Statistics and One-Class Support Vector Machines. *Proceedings of SPIE*, **5306**, 35-45. <https://doi.org/10.1117/12.526012>
- [3] Holotyak, T., Fridrich, J. and Voloshynovskiy, S. (2005) Blind Statistical Steganalysis of Additive Steganography Using Wavelet Higher Order Statistics. *9th IFIP TC-6 TC-11 Conference on Communications and Multimedia Security*, **3677**, 273-274. [https://doi.org/10.1007/11552055\\_31](https://doi.org/10.1007/11552055_31)
- [4] Holland, J.H. (1985) *Adaptation in Nature and Artificial Systems*. University of Michigan Press, Ann Arbor, 68-72.
- [5] Wu, J.H. and Chen, C.B. (2009) An Alternative Form for Grey Relational Grades. *The Journal of Grey System*, **11**, 7-12.
- [6] Mihcak, M.K., Kozintsev, I., Ramchandran, K. *et al.* (2009) Low-Complexity Image Denoising Based on Statistical Modeling of Wavelet Coefficients. *IEEE Signal Processing Letters*, **6**, 300-303. <https://doi.org/10.1109/97.803428>
- [7] Schmitt, L.M. (2011) Theory of Genetic Algorithms. *Theoretical Computer Science*, **259**, 1-61. [https://doi.org/10.1016/S0304-3975\(00\)00406-0](https://doi.org/10.1016/S0304-3975(00)00406-0)
- [8] Vapnik, V. (2008) *Statistic Learning Theory*. J.Wiley & Sons, New York.
- [9] NRCS Photo Gallery[EB/OL]. <http://photogallery.nrcs.usda.gov>

#### 知网检索的两种方式:

1. 打开知网页面 <http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD>  
下拉列表框选择: [ISSN], 输入期刊 ISSN: 2329-1273, 即可查询
2. 打开知网首页 <http://cnki.net/>  
左侧“国际文献总库”进入, 输入文章标题, 即可查询

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [jisp@hanspub.org](mailto:jisp@hanspub.org)