

文章编号:1007-2780(2017)07-0560-07

基于贝叶斯分类器的图像隐写分析

张兴春¹, 孙寿健^{2,3*}

- (1. 武警黑龙江总队司令部, 黑龙江 哈尔滨 150028;
2. 武警黑龙江总队 佳木斯支队司令部, 黑龙江 佳木斯 154000;
3. 网络与信息安全武警部队重点实验室, 陕西 西安 710086)

摘要:集成分类器是目前用于图像隐写分析的主流分类器。为提高集成分类器的检测精度, 针对集成分类器基分类器组合方法过于简单, 无法体现基分类器之间的内在联系, 不能从整体上对结果进行判定的缺点, 依据图像特征在集成分类器分类超平面上的投影值服从多维正态分布这一特性, 提出了一种基于贝叶斯分类器的图像隐写分析算法。首先基于随机森林算法生成若干基分类器, 然后计算类条件概率密度函数与先验概率并训练贝叶斯分类器, 最后使用经过训练的贝叶斯分类器代替简单投票方法进行分类判决。算法的检测错误率比以往算法平均降低了 1.6%, ROC 曲线比简单投票方法更接近于左上角, 即具有更高的检测率, AUC 值平均增长约 2.12%, 并且训练时间仅有少量提高, 最大提高约 2.610 s。可以有效提高集成分类器的检测精度。

关键词: 隐写分析; 集成分类器; 组合方法; 多维正态分布; 贝叶斯分类器

中图分类号: TP391 **文献标识码:** A **doi:** 10.3788/YJYXS20173207.0560

Image steganalysis based on bayesian classifier

ZHANG Xing-chun¹, SUN Shou-jian^{2,3*}

- (1. *Command of Heilongjiang Crops of PAP, Harbin 150028, China;*
2. *Command of Jiamusi Detachment, Heilongjiang Crops of PAP,*
Jiamusi 154000, China; 3. *Network and Information Security Key Laboratory of PAP,*
Engineering University of the PAP, Xi'an 710086, China)

Abstract: Recently, ensemble classifier is predominantly used for steganalysis of digital media. For increasing the detection accuracy of ensemble classifier and focused on the defects that the fusion method of ensemble classifier is too simple to reflect the correlation of base learners and can't give an overall result, an algorithm based on Bayesian Classifier is proposed according to a characteristic that the projection of feature on the hyper planes of ensemble classifier obeys the framework of the multivariate Gaussian distribution. At first, some base learners were generated based on the random forest, then the probability density function and prior probability were calculated for training a Bayesian classifier. At last, the algorithm uses a Bayesian Classifier instead of the majority voting rule to fuse the deci-

收稿日期: 2017-01-06; **修订日期:** 2017-03-23.

基金项目: 国家自然科学基金(No.61403417)

Supported by National Natural Science Foundation of China(No.61403417)

* 通信联系人, E-mail: zynwj08@163.com

sions of base learners. The error detection rate lower than before by an average of 1.6%, ROC curve is closer to the top left corner than simple voting method, which has higher detection rate, the growth of AUC value is about 2.12%, and the training time increase about 2.610 s. The proposed method is able to increase steganalysis performance of ensemble classifier.

Key words: steganalysis; ensemble classifier; fusion technique; multivariate Gaussian distribution; Bayesian classifier

1 引言

隐写术是信息隐藏技术的一个重要分支,其主要是利用数字载体的冗余性,在载体中嵌入隐秘信息,从而达到隐秘通信的目的^[1]。图像由于其结构简单、冗余空间较大、编码方便等特点^[2-3],成为隐写术中最常用的载体。隐写分析作为隐写技术的反技术,其目的是检测载体中是否存在隐秘信息。目前,隐写分析主要是通过有监督的机器学习来实现的,特征提取与分类器设计是其中两个重要方面^[4-5]。在早期的隐写分析中,由于特征维数较低,神经网络分类器与支持向量机被广泛使用。但随着基于最小失真代价函数^[6]等新型隐写算的出现,低维特征已不能满足隐写分析的需求。为能够对新型隐写算法进行有效的检测,Fridrich团队开发出基于共生矩阵的“富模型”^[7](Spatial Rich Model, SRM)特征,该特征由于维数过高造成了传统分类器的失效。为了克服“维数灾难”问题,Kodovsky在文献[8]中设计了基于随机森林理论^[9]的集成分类器,该分类器通过集成各基分类器的分类结果来获的最终的分类结果。在训练过程中,每个基分类器通过选取部分特征而不是全部特征进行训练,因此与传统分类器相比,集成分类器具有更低的复杂度与更高的检测精度。

集成分类器中,一个重要的步骤是如何将各基分类器的输出结果按照某种方式组合在一起^[10]。文献[8]使用的是简单投票方法,此方法把各基分类器看成平等独立的个体,赋予每个基分类器相同的投票权重,按照“少数服从多数的原则”对分类结果进行投票。但简单投票方法无法体现基分类器之间的内在联系并从整体上给出判决结果,而造成部分误判,降低了检测精度^[11]。为解决这一问题,本文根据Cogranne在文献[12]中指出的特征在集成分类器分类超平面上的投影

值服从多维高斯分布这一特性,提出了基于贝叶斯分类器的图像隐写分析,并给出了先验概率与类条件概率密度的计算方法。实验结果表明,针对不同的隐写算法,本文算法在分类精度上,均优于简单投票方法。

2 集成分类器

2.1 基分类器的生成

文献[8]中所设计的集成分类器由 L 个独立的费舍尔线性分类器(Fisher Linear Discriminate FLD)组成。该算法从特征空间中随机抽取部分特征生成特征子空间用于训练基分类器,通过基分类器对样本进行分类判决,最后使用简单投票方法获得最终判决结果。

假设特征总维数为 d , N^{tm} 代表训练样本集,训练样本集中存在成对的原始特征与隐写特征。 B_l 代表每一个独立的基分类器。集成分类器训练过程如下:

输入:训练样本集 N^{tm}

For $l = 1$ to L

{

(1) 从 d 维特征空间中随机选取特征子空间 d_{sub} ,其中 $d_{\text{sub}} \ll d$

(2) 使用自助法产生随机样本 $N_l = \text{Bootstrap}(N^{\text{tm}})$,其中 $|N_l| = N^{\text{tm}}$

(3) 取 N_l 和 d_{sub} 对应的特征集合训练基分类器 B_l

}

输出 L 个基分类器 $\{B_1, B_2, B_3 \cdots B_L\}$

2.2 FLD分类器

FLD分类器是Kodovsky在文献[8]中所用的基分类器。与支持向量机相比,该分类器具有较低的训练复杂度。FLD分类器的分类原理是

Fisher 判别法, Fisher 判别法是一种模式识别的方法, 其基本思想是将高维数据点投影到一个合适的低维空间上。这个低维空间能够保证较大的类间散度与较小的类内散度。

假设 $C^{\text{trn}} = (c_1^{\text{trn}}, c_2^{\text{trn}}, \dots, c_N^{\text{trn}})$ 代表原始图像训练样本集, $S^{\text{trn}} = (s_1^{\text{trn}}, s_2^{\text{trn}}, \dots, s_N^{\text{trn}})$ 代表隐写图像训练样本集。 c_i^{trn} 与 s_i^{trn} 分别代表 d 维的原始特征与隐写特征。 f 代表一幅图像的 d 维特征向量, 现要使用 FLD 分类器判断该图像是否被隐写。

FLD 分类器会根据训练样本集生成权值向量 w 与判断阈值 b 。权值向量 w 按照公式(1)进行计算。

$$w = (\Sigma_c + \Sigma_s)^{-1}(\mu_c - \mu_s), \quad (1)$$

其中: μ_c 代表原始特征的均值向量, μ_s 代表隐写特征的均值向量, Σ_c 代表原始特征的协方差矩阵, Σ_s 代表隐写特征的协方差矩阵。 $\mu_c, \mu_s, \Sigma_c, \Sigma_s$ 计算公式如下:

$$\mu_c = \frac{1}{N_{\text{trn}}} \sum_{n=1}^{N_{\text{trn}}} c_n^{\text{trn}}, \quad (2)$$

$$\mu_s = \frac{1}{N_{\text{trn}}} \sum_{n=1}^{N_{\text{trn}}} s_n^{\text{trn}}, \quad (3)$$

$$\Sigma_c = \frac{1}{N_{\text{trn}} - 1} \sum_{n=1}^{N_{\text{trn}}} (C_n^{\text{trn}} - \mu_c)(C_n^{\text{trn}} - \mu_c)^T, \quad (4)$$

$$\Sigma_s = \frac{1}{N_{\text{trn}} - 1} \sum_{n=1}^{N_{\text{trn}}} (S_n^{\text{trn}} - \mu_s)(S_n^{\text{trn}} - \mu_s)^T, \quad (5)$$

求出投影向量 w 后, 求出判断阈值 b 。在文献[6]中, b 是选取该 FLD 分类器在分类错误率最低时所取到的判断阈值。每个 FLD 分类器按照以下的方法对图像进行判别:

$$\begin{cases} \text{未被隐写 } w^T f - b < 0 \\ \text{被隐写 } w^T f - b > 0 \end{cases} \quad (6)$$

2.3 简单投票方法

Kodovsky 在文献[8]中使用的是简单投票方法来集成各基分类器的分类结果, 该方法的基本思想是“少数服从多数”。简单投票方法把每个基分类器看成平等独立的个体, 在投票时赋予他们相同的投票权重。

假设 y 代表需要进行判别样本的特征, $B_l(y)$ 代表第 l 个基分类器的判别结果。若判定为隐写图像, 则 $B_l(x) = 1$, 若判定为载体图像,

则 $B_l(x) = 0$ 。使用简单投票方法将各基分类器结果进行集成, 得到最终的判别结果为:

$$B(y) = \begin{cases} 1 & \text{当 } \sum_{l=1}^L B_l(y) > \frac{L}{2} \\ 0 & \text{当 } \sum_{l=1}^L B_l(y) < \frac{L}{2} \\ \text{随机, 其它} & \end{cases} \quad (7)$$

但是, 简单投票方法过于简单。它将每个基分类器看出了相互独立的个体, 忽略了基分类器之间的内在联系与基分类器之间的内在差异。基于此, 本文使用贝叶斯分类器代替简单投票方法。

3 贝叶斯分类器

3.1 贝叶斯基本原理

贝叶斯分类器是基于最大后验概率准则来实现的, 即利用某对象的先验概率计算其后验概率, 并选择具有最大后验概率的类作为该对象所属的类^[13]。令 f 代表需要进行判定的特征, $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$ 代表类别空间, 则根据最大后验概率原理, f 所属的类别 ω_i 需要满足:

$$P(\omega_i | f) = \max_{j=1,2,\dots} P(\omega_j | f), \quad (8)$$

即满足最大后验概率准则。将贝叶斯表达式

$$P(\omega_i | f) = \frac{P(f | \omega_i)P(\omega_i)}{P(f)}, \quad (9)$$

进行推导可得:

$$\max P(f)P(\omega_i | f) = \max P(f | \omega_i)P(\omega_i), \quad (10)$$

其中 $P(\omega_i)$ 代表先验概率, $P(f | \omega_i)$ 代表类条件概率密度。由于 $P(f)$ 通常是一个定值, 所以:

$$\max P(\omega_i | f) = \max P(f | \omega_i)P(\omega_i), \quad (11)$$

即构成了贝叶斯分类器。

3.2 贝叶斯分类器的生成

训练贝叶斯分类器, 即要求出先验概率 $P(\omega_i)$ 与类条件概率密度 $P(f | \omega_i)$ 。根据 Cognanne 在文献[12]中指出, 原始图像特征与隐写图像特征在 FLD 分类器分类超平面上的投影服从多维正态分布。假设基分类器数量为 L , f 代表 ω_i ($i = 1$ 为原始图像, $i = 2$ 为隐写

图像)类图像中一幅图像的 d 维特征向量, p 代表一个 $L \times d$ 的维的稀疏矩阵,矩阵的每一行包括该基分类器特征空间对应的权值向量 w ,空余位置用 0 代替。计算向量 $v = pf$,则向量 v 满足多维正态分布,多维正态分布的概率密度函数为 Σ :

$$P(v) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left[-\frac{1}{2} (v - \mu)^T \Sigma^{-1} (v - \mu) \right], \quad (12)$$

公式(12)即作为 w_i 类的类条件概率密度函数。

假设在 L 个基分类器中,有 l_1 个基分类器判定 f 为 ω_1 类,有 l_2 个基分类器判定 f 为 ω_2 类,根据古典概率原理,求得先验概率为

$$\begin{cases} P(\omega_1) = l_1/L \\ P(\omega_2) = l_2/L \end{cases}, \quad (13)$$

所以贝叶斯分类器为 $\max P(\omega_i/f) = \max P(v)P(\omega_i)$ 。

为方便计算,对其进行对数变换,

$$g_i(v) = v^T W_1 v + W_2^T v + W_3, \quad (14)$$

其中:

$$\begin{cases} W_1 = -\frac{1}{2} \Sigma_i^{-1} \\ W_2 = \Sigma_i^{-1} \mu \\ W_3 = -\frac{1}{2} \mu^T \Sigma_i^{-1} \mu - \frac{1}{2} \log(|\Sigma_i|) + \log(P(\omega_i)) \end{cases}. \quad (15)$$

如果 $g_1(v) > g_2(v)$,则判定为原始图像,如果 $g_1(v) < g_2(v)$,则判定为隐写图像。

3.3 算法流程

本文首先使用随机森林的方法训练出 L 个基分类器,然后使用训练样本计算出先验概率与类条件概率密度函数并训练贝叶斯分类器,最后对测试样本进行分类。

4 实验结果与分析

4.1 实验准备

为验证本文所提出算法的有效性,从 BOSS-Base1.01 图像库中选取 10 000 幅图像进行实验。所有图像均裁剪为 512×512 ,质量因子为 75。从 10 000 幅图像中随机选取 5 000 幅图像组成训练

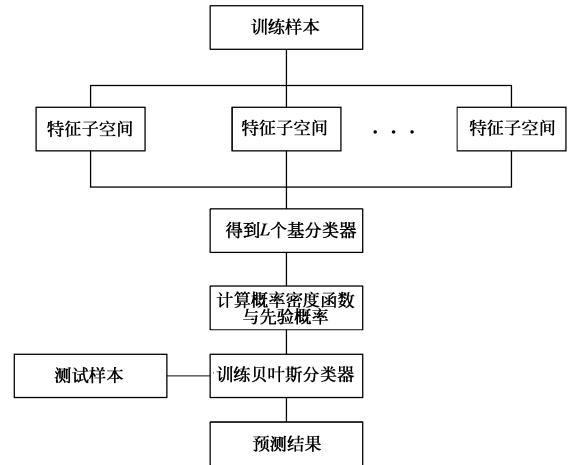


图1 算法流程图

Fig.1 Flow diagram of our algorithm

样本,剩余图像组成测试样本。分别采用嵌入率为 0.4 bpp 的 S-UNIWARD^[14]、HILL^[15] 与 WOW^[16] 三种隐写算法对图像进行隐写。对每幅图像,提取 SRM(34671 维)特征用于隐写分析。实验环境为 Windows Server 2012,仿真软件为 MATLAB_R2012a。

4.2 检测错误率比较

为了能定量地比较检测效果,使用 ER (Error Rate, ER) 值作为衡量检测性能的定量指标,并与使用简单投票方法的集成分类器进行比较。计算公式为:

$$ER = \frac{1}{2} (P_{FA} + P_{MD}), \quad (16)$$

其中, P_{FA} 代表虚警率,即将载体图像判定为隐写图像的概率, P_{MD} 代表漏检率,代表将隐写图像判定为载体图像的概率。实验所用的训练子集的特征维数为 300 维, ER 值取 10 次实验的平均值。

从表 1 中可以看出,与简单投票方法相比,贝叶斯分类器在检测错误率上平均降低 1.6%,最大降低 2.57%。且随着分类器数量的增多,降低的幅度越大,当基分类器数量达到一定时,检测错误率不再降低。这是由于随着分类器数量的增多,特征在超平面上的投影越符合多维正态分布,所以类条件概率密度与先验概率计算也就越准确,所训练的贝叶斯分类器也就越精确。但当基分类器数量过多时,贝叶斯分类器会出现过拟和现象,所以检测错误率不再进一步降低。

表 1 ER 值对比表

Tab.1 Comparison of ER values

隐写算法	分类器数量	贝叶斯分类器	简单投票方法
S-UNIWARD	50	0.259 1	0.263 0
	100	0.240 6	0.249 0
	150	0.233 4	0.248 0
	200	0.215 8	0.243 0
	250	0.216 1	0.243 2
	300	0.221 2	0.243 5
HILL	50	0.278 0	0.284 0
	100	0.275 3	0.281 0
	150	0.265 6	0.280 0
	200	0.251 2	0.287 0
	250	0.250 9	0.276 0
	300	0.251 1	0.277 4
WOW	50	0.255 3	0.263 0
	100	0.236 6	0.251 0
	150	0.227 2	0.247 0
	200	0.215 5	0.244 0
	250	0.222 9	0.249 0
	300	0.223 1	0.249 1

4.3 ROC 曲线图比较

为能定性地比较两种方法,本文对比了两种分类器在分类器数量为 200 情况下,3 种隐写算法的 ROC 曲线图。

从图 2,图 3,图 4 看,贝叶斯分类器的 ROC 曲线要缓于简单投票方法的 ROC 曲线,并且更接近于左上角。说明在相同的虚警率的情况下,贝叶斯分类器具有更高的检测率,即在分类效果上,贝叶斯分类器要优于简单投票方法。

为了能够从数值上对 ROC 曲线有一个定性的分析,本文又计算了 ROC 曲线的 AUC(Area Under Curve)值。

AUC 值代表 ROC 曲线下面积,AUC 值越大,说明分类效果越好。从表 2 中可以看出,贝叶斯分类器的 AUC 值比简单投票方法平均高出 2.12%,说明贝叶斯分类器的分类精度要高于简单投票方法。

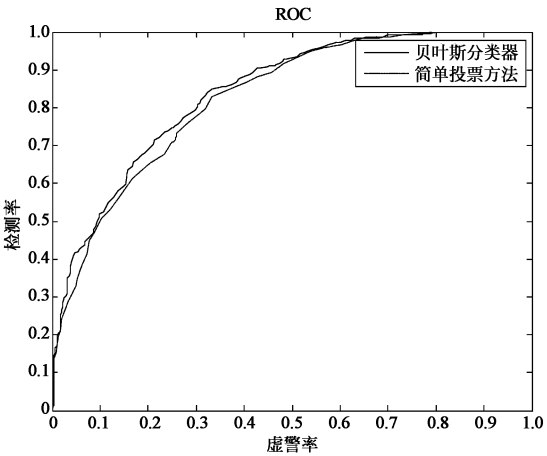


图 2 S-UNIWARD 检测 ROC 曲线图
Fig. 2 ROC of the detection of S-UNIWARD

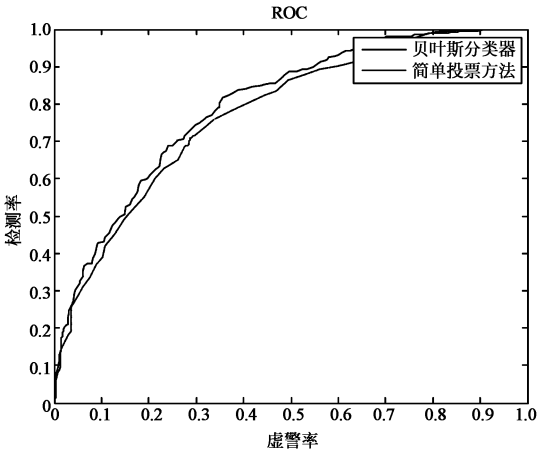


图 3 HILL 检测 ROC 曲线图
Fig.3 ROC of the detection of HILL

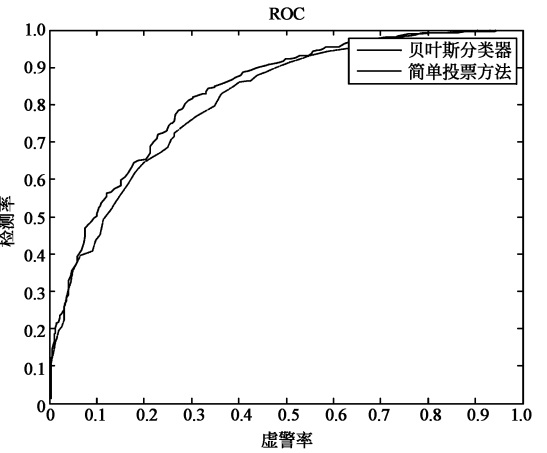


图 4 WOW 检测 ROC 曲线图
Fig.4 ROC of the detection of WOW

表2 AUC值对比表

Tab.2 Comparison of the value of AUC

隐写算法	贝叶斯分类器	简单投票方法
S-UNIWARD	0.858 7	0.836 5
HILL	0.809 9	0.788 3
WOW	0.846 8	0.827 0

4.4 训练时间比较

训练时间也是评价分类器性能的重要因素,分类器的运算时间主要来自与训练阶段。本文比较了两种方法针对 S-UNIWARD 隐写算法进行检测时分类器的训练时间。

表3 训练时间对比表

Tab.3 Comparison of the training time

分类器数量	贝叶斯分类器/min	简单投票方法/min
50	0.362	0.359
100	0.381	0.374
150	0.412	0.398
200	0.446	0.423
250	0.496	0.462
300	0.535	0.491

参 考 文 献:

- [1] TANG W X, LI H D, LUO W Q, *et al.* Adaptive steganalysis based on embedding probabilities of pixels [J]. *IEEE Transactions on Information Forensics and Security*, 2016, 4(11): 734-745.
- [2] 柯洪昌,孙宏彬.图像序列的显著性目标区域检测方法[J].*中国光学*,2015,8(5):768-774.
KE H C, SUN H B. A saliency target area detection method of image sequence [J]. *Chinese Optics*, 2015, 8(5): 768-774. (in Chinese)
- [3] 陈燕芹,段锦,祝勇,等.基于纹理特征的图像复杂度研究[J].*中国光学*,2015,8(3):407-414.
CHEN Y Q, DUAN J, ZHU Y, *et al.* Research on the image complexity based on texture features [J]. *Chinese Optics*, 2015, 8(3): 407-414. (in Chinese).
- [4] LI F Y, WU K, LEIJ S, *et al.* Steganalysis over large-scale social networks with high-order joint features and clustering ensembles [J]. *IEEE Transactions on Information Forensics and Security*, 2016, 11(2): 344-357.
- [5] 卢洋,王世刚,赵文婷,等.基于人脸姿态估计的虚拟眼镜试戴技术[J].*中国光学*,2015,8(4):582-588.
LU Y, WANG S G, ZHAO W T, *et al.* Technology of virtual eyeglasses try-on system based on face pose estimation [J]. *Chinese Optics*, 2015, 8(4): 582-588. (in Chinese)
- [6] FILLER T, JUDAS J, FRIDRICH J. Minimizing embedding impact in steganography using trellis-coded quantization [C]//*Proceedings of SPIE 7541, Media Forensics and Security II*. San Jose, California: SPIE, 2010: 18-20.
- [7] FRIDRICH J, KODOVSKY J. Rich models for steganalysis of digital images [J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(3): 868-882.
- [8] KODOVSKY J, FRIDRICH J, HOLUB V. Ensemble classifiers for steganalysis of digital media [J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(2): 432-444.

从表3中可以看出,贝叶斯分类器的训练时间要多于简单投票方法,且随着分类器数量的增多而增大。这是由于与简单投票方法相比,本文所用算法还需训练贝叶斯分类器对未知样本进行分类,因此增加了训练时间,并且基分类器数量越多,计算类条件概率密度函数与先验概率所花的时间也就越长。但是,本文算法增加训练时间的幅度并不高。与简单投票方法相比,最大增加2.610 s,增加的时间在可接受的范围内。

5 结 论

本文提出了一种基于贝叶斯分类器的图像隐写分析算法。在生成 L 个 FLD 分类器的基础上,计算类条件概率密度函数与先验概率并训练贝叶斯分类器进行分类判决。实验表明,与简单投票方法相比,使用贝叶斯分类器进行分类判决的检测错误率比以往算法平均降低了 1.6%,ROC 曲线比简单投票方法更接近于左上角,即具有更高的检测率,AUC 值较平均增长约 2.12%,并且训练时间仅有少量提高,最大提高约 2.610 s。下一步将继续改进分类器的训练方法,减少训练时间。

- [9] KODOVSKÝ J, FRIDRICH J. Steganalysis in high dimensions: fusing classifiers built on random subspaces [C]// *Proceedings of SPIE 7880, Media Watermarking, Security, and Forensics III*. San Francisco, CA, USA: SPIE, 2011: 78800L.
- [10] XU X K, DONG J, WANG W, *et al.* An adaptive ensemble classifier for steganalysis based on dynamic weighted fusion [C]// *Proceedings of the 2015 International Conference on Communications, Signal Processing, and Systems*. Berlin, Heidelberg: Springer, 2016: 639-647.
- [11] 张敏情, 狄富强, 刘佳. 基于选择性集成分类器的通用隐写分析[J]. *四川大学学报(工程科学版)*, 2015, 47(1): 36-41.
ZHANG M Q, DI F Q, LIU J. Universal steganalysis based on selective ensemble classifier [J]. *Journal of Sichuan University (Engineering Science Edition)*, 2015, 47(1): 36-41. (in Chinese)
- [12] COGRANNE R, FRIDRICH J. Modeling and extending the ensemble classifier for steganalysis of digital images using hypothesis testing theory [J]. *IEEE Transactions on Information Forensics and Security*, 2015, 10(12): 2627-2642.
- [13] 曹玲玲, 潘建寿. 基于 Fisher 判别分析的贝叶斯分类器[J]. *计算机工程*, 2011, 37(10): 162-164.
CAO L L, PAN J S. Bayesian classifier based on fisher discriminant analysis [J]. *Computer Engineering*, 2011, 37(10): 162-164. (in Chinese)
- [14] DENEMARK T, FRIDRICH J, HOLUB V. Further study on the security of S-UNIWARD [C]// *Proceedings of SPIE 9028, Media Watermarking, Security, and Forensics 2014*. San Francisco, California, United States: SPIE, 2014: 902805.
- [15] LI B, WANG M, HUANG J W, *et al.* A new cost function for spatial image steganography [C]// *Proceedings of 2014 IEEE International Conference on Image Processing (ICIP)*. Paris, France: IEEE, 2014: 4206-4210.
- [16] HOLUB V, FRIDRICH J. Designing steganographic distortion using directional filters [C]// *Proceedings of 2012 IEEE International Workshop on Information Forensics and Security (WIFS)*. Tenerife: IEEE, 2012: 234-239.

作者简介: 张兴春(1980—), 男, 黑龙江哈尔滨人, 主要从事信息安全、隐写分析以及图像处理的工作。

孙寿健(1990—), 男, 山东临邑人, 硕士, 主要从事信息隐藏以及信号处理等工作。E-mail: zynwj08@163.com