一种基于线性规划特征选择和集成分类器的图像隐写分析方法

关晴骁[1,2], 朱 杰[1,2], 赵险峰[1,2], 于海波[1,2], 刘长军[1,2]

[1] 中国科学院信息工程研究所信息安全国家重点实验室, 北京 中国 100093
[2] 中国科学院大学网络空间安全学院, 北京 中国 100093

摘要 隐写分析是防范由隐写术进行信息隐藏所带来危害的有效方法。图像隐写分析方法主要用于检测图像是否被隐写术嵌入隐秘信息。通用型图像隐写分析能够针对广泛类型的隐写术进行检测, 该类方法一般采用从图像提取的统计特征和分类器模型进行。当前的高性能隐写分析一般采用高维特征和集成分类器进行。高维特征能够较好地表达图像统计特性中被隐写术扰动的成分, 但另一方面, 高维特征具有较多的冗余和无效成分, 因此进行特征选择能较好的提升效率。本文提出一种使用线性规划的特征选择模型, 该模型可与集成分类器协同使用, 同时考虑集成分类器中子分类器的检测精度和多个子分类器使用特征的多样性。实验证明, 本文提出的方法对多个隐写术的检测性能有较好的提升。

关键词 隐写分析; 特征选择; 隐写术; 信息隐藏
中图法分类号: TP37 DOI 号 10.19363/j.cnki.cn10-1380/tn.2018.01.006

Image Steganalysis Based on Linear Programing Feature Selection and Ensemble Classifier

GUAN Qingxiao[1,2], ZHU Jie[1,2], ZHAO Xianfeng[1,2], YU Haibo[1,2], LIU Changjun[1,2]

[1] State Key Laboratory of Information Security, Institute of Information Engineering,Chinese Academy of Sciences, Beijing 100093, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100093, China

Abstract steganalysis is an effective method to prevent the vicious usage of steganography. Image steganalysis can detect the presence of secret message embedded by steganography in digital image. Universal image steganalysis method is designed to detect various kinds of steganography, such methods usually use ensemble classifier and high dimensional feature which can capture the disturbance introduced by steganography embedding. On another hand, there are many ineffective and redundant components in high dimensional feature, thus feature selection methods can enhance its detection accuracy. In this paper we propose a feature selection method for high dimensional feature and ensemble classifier based image stegnalysis. In this method, we consider the accuracy of base classifiers in ensemble classifier and the diversity of subsets of feature used by them. Experimental result shows that our method can improve detection performance on many kinds of steganography.

Key words Steganalysis, Feature selection, Steganography, Information hiding

1 引言

1.1 隐写术简介

隐写术(Steganography)[1]是一种将信息隐藏至数字媒体文件中的技术, 该类技术对数字媒体文件的内容数据进行少量的极其轻微的修改, 从而可以将任意的计算机文件或信息嵌入到媒体文件内容的数据中, 同时不改变该媒体文件的格式信息、视觉外观、媒体可理解的内容等因素, 具有不可见性。因此, 隐写术是一种可对信息进行伪装的信息隐藏技术。

与其他信息隐藏技术相比, 例如应用于版权保护或内容认证的信息隐藏技术相比, 例如数字水印[2]、可逆隐藏[3]等, 隐写术主要注重抗检测性及嵌入容量, 即在嵌入一定量的信息条件下, 尽可能地提升媒体文件的抗检测性。因此, 隐写术在嵌入信息时, 尽可能减少对媒体文件内容数据的修改次数或统计

通讯作者: 关晴骁, Email: guanqingxiao@iie.ac.cn。
本课题得到国家自然科学基金(U1536105, U1636102), 国家重点研发计划(2016QY15Z2500, 2016QY15Z2500)资助。
收稿日期: 2016-06-14; 修改日期: 2016-10-20; 定稿日期: 2017-12-05

特性的扰动, 并可以嵌入相对较长的信息。

隐写术与传统的加密方法相比, 也具有较大的差别。加密方法尽管能够保护信息的内容不被他人获取, 但加密后的数据具有异于常规明文数据的特性, 该类数据在传输过程中, 将暴露信息传递者发送加密数据这一行为,因此难以避免地会引起监视方的怀疑和警觉。而隐写术可使用正常媒体文件传输隐秘信息, 因此可使得数据的发送者以正常用户的行为传输数据, 不被引起怀疑, 可同时保护信息和信息传输的行为。值得一提的是, 隐写和加密技术是并存关系, 传输者可使用任意加密方法将信息加密后嵌入至隐写媒体, 接受者提取密文后进行相应的解密即可得到明文。

数字图像是当前互联网流传最为广泛的媒体信息载体之一, 大量的网站、社交平台等用户使用数量巨大的数字图像传递信息, 普通的电子邮件中也有为数较多的图像文件附件。互联网上的图像在传递信息的同时, 也可能被不法分子、敌对势力等利用, 进行失泄密、传递不法信息等。使用隐写术在网络空间传递隐写图像具有较高的隐蔽性, 信息传递双方无需建立联通关系, 接收方以普通用户身份在公开的数据分享平台下载含有隐秘信息的隐写图像, 即完成信息的获取。当前, 已经有一些国内外使用隐写术传递情报、进行犯罪活动的报道。例如在美国被捕获的俄罗斯间谍安娜查普曼, 采用隐写将情报信息隐藏在图像中, 并发布在俄罗斯的某社交网站, 以此传递情报信息, 为使用隐写进行隐秘通信的典型案例。

1.2 隐写分析简介

为应对隐写术可能带来的危害, 隐写分析(Steganalysis)方法应用而生, 另一方面, 进行隐写术的研究也依赖于隐写分析方法对其进行安全性验证, 因此不少的研究机构都开展相应的工作。近些年来, 针对数字图像的隐写方法层出不穷, 因此本文主要论述的隐写分析方法主要针对数字图像进行隐写检测。一般而言, 针对数字图像的隐写分析方法具有以下几方面的难点:

1) 隐写嵌入的低扰动性。隐写嵌入引起的修改对图像扰动极小, 不仅不会破坏图像的视觉内容, 其修改幅度甚至低于某些图像在成像时的噪声幅度。

2) 图像内容的掩盖。数字图像本身的内容相对隐写嵌入引入的修改较强, 且图像的拍摄内容、场景环境、光照的条件等都具有极大的复杂性、多样性和不可控性, 无法用特定模型精确描述, 因此给隐写检测带来较大的不确定性。

3) 数字图像内容的部分随机性。自然图像多个像素值的最低比特位组成的比特串本身具有较强的随机性, 即使在已知嵌入方法和嵌入参数的条件下, 也无法通过从图像中主动提取可能存在的隐写信息来判别该图像是否为隐写图像。

由于以上几点原因, 隐写分析通常采用统计判别的方法进行, 即采用从图像中计算统计信息, 并据其进行判决该图像是否为隐写图像。采用统计判别方法检测隐写图像的基本前提为: 尽管隐写嵌入信息在图像视觉层面和局部信息不会造成可察觉的扰动, 但在一定程度上将破坏图像数据的全局统计特性, 即对统计特性造成扰动, 因此可据此进行检测。

根据适用范围不同, 图像隐写分析可分为两种, 即针对特定隐写术的分析方法和针对多种隐写术的通用型分析方法。两者具有较大的区别, 前者针对特定性隐写术嵌入的分析方法在针对该隐写术的特点进行设计, 能够较为精确地进行检测, 且该类方法一般提取的统计信息较少, 通常不需要采用分类器等方法, 直接以阈值进行判别。而后者不针对特定隐写术方法设计, 而是在图像中提取多个统计特征进行分析, 因此可针对广泛类型的隐写术进行检测。由于通用型方法采用多个统计特征, 且无法显式地对隐写样本和正常样本两类特征分布建模, 因此必须借助于分类器将特征映射至决策结果。分类器模型通过两类样本训练得到, 即通过输入大量正常图像样本和隐写嵌入后图像样本的特征, 根据分类器训练的方法训练。因此通用型分析方法在训练阶段, 需要人为准备大量的训练样本。

在现实应用环境中, 所面临的隐写使用者可以动态选用多种不同的隐写方法, 因此对于待测图像, 一般而言通常无法获取其相关的隐写术方法先验知识, 即无法具体得知被检测的图像所使用的隐写术种类, 因此通常只能采用通用型方法进行检测。本文主要讨论通用型图像隐写分析方法。

1.3 本文的主要内容

由于通用型隐写分析方法通常采用特征和分类器进行, 因此其训练分类器的过程至关重要, 其训练过程中使用特征的有效性、维度等方面的因素均对隐写分析的精度具有较大影响。针对当前主流的通用型隐写分析方法特征维度高, 冗余特征多等问题, 本文提出一种特征选择方法进行改进, 该方法主要基于线性规划的特征选择, 能够较好地与当前基于高维隐写分析特征使用的分类器方法进行结合, 将基于线性规划的特征选择引入集成分类器的子分类器训练中。该方法在减少子分类器所使用的特征

维度, 提升其性能精度的同时, 也将每个子分类器的相互差异考虑在内。通过特征选择模型的参数, 均衡各个子分类器最终选用特征的差异程度。实验证明该方法能够有效提升通用型隐写分析的性能。

本文的以下章节内容按照以下组织: 第二章主要介绍通用型隐写分析方法的一般性流程, 并回顾了隐写分析特征和分类器的相关工作。第三章主要介绍本文方法涉及到的两种高维特征以及用于隐写分析的集成分类器。第四章是本文主要章节, 回顾了针对于通用型隐写分析方法的特征选择方法, 并提出基于线性规划的隐写分析特征选择方法及其与集成分类器结合的方法。第五章是实验部分, 主要对比了采用特征选择策略前后的检测性能。第六章是结论, 总结了本文的工作并展望未来的工作。

2 通用型隐写分析方法介绍及相关工作

2.1 通用型隐写分析检测的过程

通用型隐写分析主要依赖与特征和分类器, 因此隐写分析特征和分类器的设计是通用型隐写分析的核心问题。本章的以下小节介绍相关的工作。

2.2 隐写分析特征相关工作

隐写分析特征是通用型隐写分析所以来的基本信息。分析隐写特征通常是从图像中计算得到的一组数据, 因此可视为一个包含多个数值的向量。由于图像本身内容的复杂性和多样性对隐写特征概率分布具有影响, 理想的隐写分析特征是能够表达隐写图像和正常图像的差距, 即正常图像集与隐写图像集提取的特征分布差异尽可能较大。为此, 隐写分析的特征需要对隐写引入的修改较为敏感。

当前, 已经有一些隐写分析特征, 根据其针对的图像格式不同, 可分为两种, 即针对空域图像的特征和针对 JPEG 图像的特征。从技术路线的方式划分, 特征可分为早期的低维特征和近期的高维特征。以下分别对具有典型性的特征进行介绍。

1) 空域图像包含 BMP、PNG、PGM 等格式, 该类格式图像直接存储像素的灰度值。自然空域图像的相邻像素之间具有一定的相关性, 因此针对空域图像隐写分析特征主要通过表达像素间相关性被隐写破坏的情况, 对隐写嵌入进行表达。针对空域图像的特征早期采用信号分解的方式, 如利用小波分解等, 在各个频带提取峰度、偏度、方差等作为特征[4]。随后, 以马尔可夫转移模型为代表的特征成为主流, 该类方法将图像相邻像素的变化建模为马尔可夫过程, 在差分图像上统计马尔可夫概率转移矩阵作为特征, 马尔可夫类特征以 SPAM 特征[5]较为具备代表性。SPAM 特征在差分图像上统计二阶马尔可夫转移矩阵。空域图像的特征随后被扩展至高维特征, 与低维特征使用单一的差分方法滤波相比, 高维特征使用更多的滤波器并计算高阶共生矩阵, 得到多组特征并进行组合, 该类特征以 SRM[6]和 PSRM[7]为代表。高维特征在检测性能方面相比之前的特征具有较大的提高。

2) JPEG 图像是一种常见的压缩格式图像, 其存储的基本信息为 JPEG 分块 DCT 变换并经过量化后的 DCT 系数。JPEG 隐写分析特征随着 JPEG 图像隐写术的发展而出现。早期的 JPEG 在 JPEG 图像存储的 DCT 系数中提取统计特征, 与空域特征有一定的类似性, 如 Chen 等人提出的 JPEG 马尔可夫特征[8], 将块内相邻的 DCT 系数变化建模成马尔可夫过程。基于 JPEG 图像分块 DCT 变换的特点, Pevny 等人提出了一组著名的 JPEG 隐写分析特征 PEV274[9], 该特征共 274 维, 由多种统计量组成, 包括 DCT 系数的全局直方图和某些频段的直方图、多种 DCT 系数的共生矩阵、块间的差分统计值、马尔可夫转移矩阵等, 此外, 该特征还采用了校正技术(Calibration), 即将 JPEG 图像解压缩至空域后, 切除左边四列和上方的四行, 然后重新压缩成为 JPEG 图像, 得到校正图像。该特征分别在原图像和校正图像提取特征, 然后相减得到校正后的特征。实验表明校正过程有利于增强特征对隐写嵌入的敏感性。该特征随后还被推广至 PEV548 特征[10], 采用拼接代替相减, 使其维度增加一倍。校正的方法在随后也出现的多种 JPEG 高维隐写分析特征方面有所体现, 如 CC-JRM[11]、DCTR[12]、GFR[13]等, 均可以视为采用了校正及其扩展的技术, 该类特征在检测隐写术方面精度也有较多的提升。

在本文的工作中, 主要使用了空域图像的 SRM 特征和 JPEG 图像的 DCTR 特征, 因此在第 3 章中将介绍这两种特征的计算方法。

2.3 隐写分析的分类器相关工作

通用性隐写分析利用多个统计特征表达隐写术扰动情况, 与专用型隐写分析方法使用简单的统计量相比, 其工作原理难以直接显式地表达出来, 因此需要借助训练得到分类器模型。通用型隐写分析在这方面的工作主要以借鉴和应用在统计学习领域中的分类器工作为主要途径。早期的隐写分析特征维度较低, 一般不超过 1000 维, 对此使用的分类器有支持向量机、神经网络等。其中以使用高斯核和多项式核的支持向量机较为普遍, 成为通用型隐写分析的主流分类器。随着新型的高安全性隐写术出

现, 传统的特征难以对图像进行有效表达, 因此出现了高维特征。高维特征的维度较传统特征具有大幅增加, 因此支持向量机难以胜任。因此, 出现了集成 FLD 分类器的方法[14], 该方法利用多个子分类器进行投票决策, 每个子分类器只使用部分特征, 因此克复了对高维特征容易过学习的问题, 此外还具有较高的训练速度, 可处理大量的数据。本文的第 3 章将对集成分类器进行详细介绍。

3 高维特征与集成分类器介绍

当前, 高维特征和集成分类器已经成为主流的通用型隐写分析方法, 该种方法不仅对传统的隐写术具有更高的检测性能, 对于多种新型的隐写术, 其检测性能也大幅度超越传统的方法。本文论述的方法主要基于高维特征和集成分类器, 因此该方法是本文论述的方法的基础, 为此, 本章对高维特征和集成分类器进行详细介绍, 其中, 高维特征包含针对空域图像的 SRM 特征以及针对 JPEG 图像的 DCTR 特征。

3.1 SRM 特征介绍

SRM(Spatial Rich Model)特征集由 Kodovsky 等人提出, 是当前针对空域图像隐写分析特征中最为成功的特征集之一。SRM 特征较之传统的空域隐写分析特征, 如 686 维的 SPAM 特征等, 均都具有较大的检测性能优势。SRM 具有以下几个特点: 使用了多种滤波器、多种量化因子提取特征并进行组合, 因此维度较高, 高达 32000 维。在计算特征的过程中, SRM 采用了四阶共生矩阵作为统计量, 在多个方向上计算四阶共生矩阵, 并根据多种对称性合并准则合并共生矩阵元素多个元素。

SRM 特征包含了多种统计像素邻域的特征子集, 因此 SRM 特征能够检测更多的隐写嵌入算法。SRM 每种特征子集的计算过程都遵循相同的计算模型, 即空域图像滤波后的经过量化(Quantization)、截断(Truncation), 最后统计共生矩阵并对其进行缩并和合并后得到。以下介绍分别 SRM 的滤波、量化和截断及其统计共生矩阵的过程。

3.1.1 SRM 的滤波过程

SRM 特征具有较多的特征子集, 能够对各类内容复杂的图像进行表达, 以反映其被隐写术嵌入的情况。SRM 特征的丰富性和多样性主要由多种类型的滤波器产生。其滤波器包括两种, 第一种是线性滤波器, 包括一阶和高阶差分、特定的滤波核滤波等, 另外一种是非线性滤波器, 主要采用最大和最小滤波, 其滤波方式是首先采用同类型不同方向的多个线性滤波核进行滤波, 然后取多个滤波值的最大或者最小值。

3.1.2 SRM 的量化与截断

SRM 对滤波后的图像进行量化和截断, 量化和截断操作主要是使滤波后的图像更加适用于高阶共生矩阵计算, 即使用截断操作控制滤波后像素值的取值范围, 以控制高阶共生矩阵的维度, 使用量化以使得不同的变化程度的滤波值都能反应在特征中。

SRM 对于同一张滤波后的图像, 采用多个量化因子, 得到多张量化后的图像, 分别进行截断操作后计算共生矩阵, 在增加特征丰富程度的同时控制特征总体维度。SRM 采用的量化因子 q 与滤波器阶数 c 有关, 例如采用一阶差分滤波及其构造的最大最小滤波器。对于不同的滤波器, SRM 采用三种量化因子, 即取滤波器阶数的 1 倍、1.5 倍、2 倍。因此, 对于每一种滤波器, SRM 可均提取三组特征。

截断操作是指对量化后的滤波图像中的像素值, 进行阈值截断操作。SRM 的截断操作为双边截断操作, 其截断函数把大于阈值T的像素值置为T, 小于$-T$的像素值置为$-T$, 因此在截断操作完成后, 截断后滤波图像的像素值取值范围为$[-T,T]$之间的$2T+1$个整数。截断操作控制了像素值取值范围, 能够在后续计算共生矩阵过程中控制共生矩阵的规模, 从而控制整体特征的维度。

3.1.3 SRM 的共生矩阵计算

SRM 通过计算滤波后图像的四阶共生矩阵来表达像素之间的依赖性和相关性, 从而反应隐写术对图像的扰动。共生矩阵反映的是由像素组成的结构元类型在图像中出现的频率。SRM 中的四阶共生矩阵的计算在截断后的滤波图像上进行, 其基本结构元为截断后滤波图像中的四个相邻的像素, 按照结构元像素排列的方向, 共生矩阵分为水平方向和垂直方向两种。由于滤波后图像的像素值经过截断后取值为$[-T,T]$, 因此 SRM 中四阶共生矩阵结构元共有$(2T+1)^4$种类型。本文以下以结构元包含的四个像素的像素值(d_1,d_2,d_3,d_4)表示结构元的类型, 其中$d_1,d_2,d_3,d_4 \in [-T,T]$。SRM 统计每一种结构元在截断后的滤波图像中出现的频次, 组成四阶共生矩阵, 并将其归一化, 用$\mathrm{C}(d_1,d_2,d_3,d_4)$表示结构元$(d_1,d_2,d_3,d_4)$在共生矩阵中对应的数值。

SRM 在得到各个滤波后图像的共生矩阵以后, 对每个共生矩阵均进行对称化合并。合并共生矩阵的方式主要包含两个操作步骤:

第一个步骤是沿对称方向合并, 把从左至右(从上到下)和从右至左(从下到上)出现的四个元素相同

的结构元视为同一类, 将其在共生矩阵中对应的元素合并, 得到新的共生矩阵, 第二步骤是按照符号对称合并, 即对共生矩阵中, 将其对应的结构元只存在符号相差的元素进行合并。SRM 通过合并共生矩阵中的一些元素, 缩减了特征维度, 使得在同样的维度下, 可容纳更多的特征子集, 增强特征的丰富程度。另一方面, 统计特征通过大量的统计量来反映图像被隐写术修改的状况, 因此统计量的单元数目越多, 特征越能够稳定地反应隐写修改的情况。

3.2 DCTR 特征介绍

DCTR(Discrete Cosine Transform Residual)特征集是一种针对 JPEG 图像进行隐写分析的新型特征, 该特征也属于高维特征。对当前包括新型自适应隐写术在内的多种隐写术, 该特征均具有较好的检测性能。DCTR 特征由 64 个滤波后图像上的直方图统计得到。滤波图像分别由 64 个 DCT 卷积核对 JPEG 图像解压缩后的空域图像卷积所得。随后对滤波后的图像的数据还进行了量化、取整、截断操作。与之前的 JPEG 高维特征 CC-JRM [6]相比, DCTR 特征的计算复杂度和特征维数均更低, 总共 8000 维, 但是隐写分析准确率并没有下降, 且针对某些隐写嵌入算法时, DCTR 特征的隐写分析准确率要高于 CC-JRM, 因此是当前较为先进的 JPEG 高维隐写分析特征。

DCTR 特征从 JPEG 图像 Y 通道的空域数据中提取, 使用多个 2 维 DCT 滤波器获取滤波后的图像, 随后进行量化和截断, 并进行降采样, 最后并计算降采样图像的直方图并合并。具体过程如以下小节所示。

3.2.1 DCTR 的滤波过程

滤波时 DCTR 特征计算过程的第一个步骤。由于滤波是在空域的亮度通道(Y 通道)进行, 因此需要先将 JPEG 图像解压到空域, 即首先根据图像的量化表将图像的 Y 通道每个 8×8 分块的 DCT 系数进行反量化, 随后反变换为空域像素灰度值。与普通的 JPEG 图像解压缩过程不同, 在此处的 DCT 逆变换后, 对像素灰度值不进行取整, 仍然保持为浮点数。对图像 Y 通道数据, 分别使用 64 个二维 8×8 的 2 维 DCT 变换基进行滤波, 得到 64 张滤波后的图像。

3.2.2 DCTR 的量化和截断过程

滤波完成后, 对滤波后的图像的像素取绝对值, 随后采用一个量化因子对其进行量化。量化因子根据该 JPEG 图像量化表对应的品质因子决定。

对量化后的滤波图像中各个像素进行截断操作。截断操作如下所示, 当像素的数值大于阈值T时, 将其设为T。截断操作如下所示:

$$R_{i,j} = trunc_T(R_{i,j}),$$

其中 $trunc_T(x) = \begin{cases} x\ if\ x < T \\ T\ if\ x \geqslant T \end{cases}$

其中, $R_{i,j}$为滤波后图像在(i,j)位置像素的像素值, T为截断阈值。在 DCTR 中, $T = 4$, 由于滤波后的图像像素数值做了取绝对值操作, 因此滤波图像各个像素数值在截断后取值范围为$[0,T]$。

3.2.3 DCTR 的直方图统计

对截断后的滤波图像, 进行直方图计算。直方图计算不直接计算全图的直方图, 而是进行以 8 为间隔降采样后进行。对于降采样, 由于可以在纵向和横向各偏移不同个数的像素后进行降采样, 因此根据偏移不同, 以 8 为间隔的降采样具有 64 种方式, 即每个滤波图像都有 64 个降采样图像。对每个降采样图像都计算直方图, 并对直方图进行归一化。

对于每个滤波后图像提取的 64 直方图, 按照一定的规则进行合并。注意到对于滤波后图像中任意一个像素的数值, 是通过 8×8 的 DCT 基与原图像中一个 8×8 块的像素计算得来, 该 8×8 块像素块相对于 JPEG 本身的 8×8 分块的相对位置呈现一定的关系。对于任意一个降采样的图像, 其像素数值计算过程中涉及到的 8×8 块像素块与 JPEG 的 8×8 分块的相对关系呈周期性出现。另一方面, 对于(x,y)偏移的降采样图像, 另外三张(某些位置为一张或者两张)降采样图像$(8-x,y)$, $(x,8-y)$, $(8-x,8-y)$的数值计算涉及到的8×8块像素块与 JPEG 的 8×8 分块的相对关系与(x,y)相同。因此将$(x,y),(8-x,y),(x,8-y),(8-x,8-y)$四张图像计算得到的直方图进行合并。由此将 64 个直方图合并后得到共 25 个直方图。

3.3 集成分类器介绍

隐写分析依赖于分类器根据特征进行检测, 分类器是隐写分析重要的组成部分。传统的隐写分析采用核支持向量机(Kernel Support Vector Machine, KSVM)[9]等作为分类器, 随着特征维度的增大, 单独使用支持向量机已经难以应对, 会出现训练困难、过学习等问题。为适应高维特征, 当前基于高维特征的隐写分析一般采用 FLD 集成分类器, FLD 集成分类器[10]采用 Fisher 线性判决分类器(Fisher Linear Discriminant, FLD)作为子分类器, 多个子分类器分别预测后, 对结果进行投票融合, 得到最终的预测结果。集成分类器中的每个子分类器只使用随机抽取的部分特征, 因此避免了过学习的问题。在训练阶段, 每个投票融合阶段, 所有的子分类器的判决结果联合预测, 增强了判决的精度, 如图 1 所示。

集成分类器依次训练多个子分类器, 即每次随机选择使用部分特征, 将训练样本的该部分特征抽出, 训练子分类器。由于每个子分类器使用的特征均是随机从高维特征中随机抽取的不同子集, 因此训练多个子分类器后, 高维特征中所有的特征均被使用。集成分类器一般具有两个参数, 即子分类器所使用的特征子集的维度d以及子分类器的个数L。d和L的取值主要由搜索得到, 搜索的依据主要是训练样本的错误率, 使用训练样本错误率最小时的参数。d取合适的值时, 能够充分地利用高维特征同时避免过学习, d的最优值随特征不同而不同, 一般而言, 最优值取值范围为 600~1200。对于d的某个取值, 当L增大到一定程度时分类器性能指标趋向于稳定, 选取此时的L作为当前d取值下的最优L参数, 随后继续搜索计算d不同的取值。

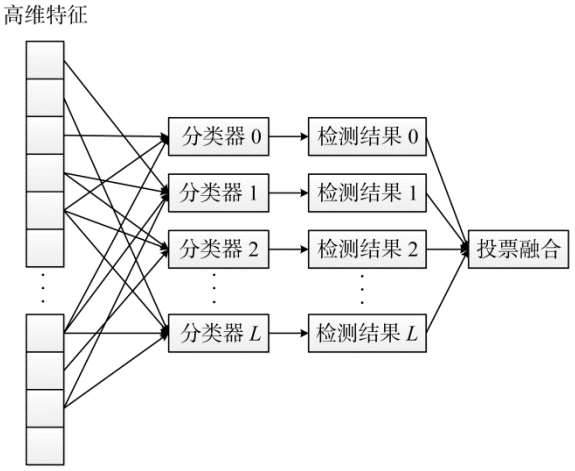


图 1　集成分类器检测过程示意图

Figure 1　Steganographic detection with ensemble classifier

4　基于线性规划的隐写分析特征选择方法

4.1　特征选择简介

当前的隐写分析特征一般采用高维特征, 增加特征丰富程度是增强检测精度的有效方法。但另一方面, 在增加特征数量的同时, 也不可避免地增加一些无效特征、或是具有冗余信息的特征, 这些特征在分类器训练阶段, 将对其造成一定的影响, 如训练的分类器模型精度降低、分类器容易过学习等。为此, 需要从原始的特征集合中进行特征选择, 将有效性高、冗余性小的特征子集选择出来。尽管当前已有一些特征选择的工作在隐写分析方面得到应用[15], 但该类工作基本集中于针对传统的低维特征进行, 无法应用于当前的高维特征。文献[16]中的方法是为数不多的针对高维隐写分析特征的方法, 但该方法较少的考虑了特征之间的冗余性, 且没有针对集成分类器进行优化。针对高维特征和集成分类器的特征选择具有以下几个难点:

1) 特征选择方法一般采用监督学习方法进行, 因此在样本数量有限的条件下, 采用该类方法直接从高维特征中选择特征子集难以避免高维特征带来的过学习的问题。

2) 难以与集成分类器结合。由于集成分类器进行检测时通过多个子分类器的检测结果投票决定, 因此根据集成学习的理论, 子分类器在保证精度的同时, 还需要使得多个子分类器对不同的样本检测结果具有多样性。因此传统的特征选择方法只考虑选择特征的有效性, 不考虑多个子分类器使用的特征多样性和有效性, 也没有对集成分类器方法进行设计。

为解决以上两个问题, 本文针对高维特征的特征选择方法进行研究, 主要采用线性规划方法对集成分类器中的各个子分类器所使用的特征集进行特征选择, 通过调节每次特征选择的参数更新保证不同子分类器选择的特征的有效性和多样性。本章以两小节将介绍线性规划特征选择模型及其与集成分类器和结合方法。

本章 4.2~4.3 节论述线性规划特征选择模型及其计算方法, 4.4 节论述该特征选择模型与集成分类器的多个子分类器训练结合的方法。

4.2　基于线性规划的特征选择模型

基于线性规划方法的特征选择方法通过两类样本构建具有一定意义的选择模型进行特征选择, 其选择的过程为求解特征选择模型的优化问题, 求解模型后得到每个特征对应的权值参数, 随后据此进行特征选择。该方法的选择模型问题如下:

$$\begin{aligned} &\min_{w,b,\xi^c,\xi^s} \lambda f^T|w| + \sum_{i=1}^{N} \xi_i^c + \xi_i^s \\ &\text{s.t. } w^T x_i^c - b \leqslant -1 + \xi_i^c, \quad i = 1,2,\cdots,N \\ &w^T x_i^s - b \geqslant 1 - \xi_i^s, \quad i = 1,2,\cdots,N \\ &\xi_i^c \geqslant 0, \xi_i^s \geqslant 0, \quad i = 1,2,\cdots,N \end{aligned} \tag{1}$$

以上模型中, $x_i^c \in \Re^d$和$x_i^s \in \Re^d$分别为第i个正常图像样本的d个备选特征组成的d维向量, $w = [w_1, w_2 \cdots w_d]^{\mathrm{T}}$为分界面加权权值向量, b为偏置值, w和b均为需要求解的变量, N为两类样本的数目, $|w|$为w的元素取绝对值组成的向量, 即$|w| = [|w_1|, |w_2|, \cdots, |w_d|]^T$, $f^T = [f_1, f_2, \cdots, f_d]$为预设的特征选择权值参数向量, 其元素大于等于 0, 表示选择特征的先验倾向性, $\lambda > 0$为预设的选择参数,

$\xi^c=[\xi_1^c,\xi_2^c,\cdots,\xi_N^c]^T$和$\xi^s=[\xi_1^s,\xi_2^s,\cdots,\xi_N^s]^T$分别为正常和隐写样本偏离分界面间隔的距离组成的向量。以上参数中, w,b,ξ^c以及ξ^s均为模型需要求解的变量, 该模型的特征选择结果由w决定。

由以上模型可以观察到, 该模型的目标由两部分组成, 即分类误差项$\sum_{i=1}^N\xi_i^c+\xi_i^s$和特征选择惩罚项$\lambda f^T|w|$, 通过约束条件可知, 分类误差项是在分类加权参数w对两类样本特征进行投影后偏离分界面界限α或β的距离误差, $\lambda f^T|w|$对w中元素的数值进行惩罚, 由于f中的元素大于等于 0 且$\lambda>0$, 因此该项对目标函数的影响由三方面决定, 其中, λ决定该项对目标函数的整体影响, f中的元素数值越大则w中相应的元素影响越大, 该两项为设定的参数, 而w为求解的结果, $|w|$表示绝对值越大的元素对目标函数影响越大。通过以上的模型可以看出, 增加λ值, 则模型求解的w中的元素趋向于 0, 因此调节λ至不同程度可使得求解的结果中w具有不同数量的元素等于 0, 据此模型进行特征选择, 即将w中不为零的系数所对应的特征作为特征选择的结果。

以上模型中, 分类误差项$\sum_{i=1}^N\xi_i^c+\xi_i^s$的表示训练样本中的某些正样本或者负样本偏离其相应的分解边界的距离之和, 因为特征选择模型中约束条件$\xi_i^c\geqslant0,\xi_i^s\geqslant0,\ i=1,2,\cdots,N$的存在, 因此求解过程中, 在当前的$w$下, 只有当样本位于向其相反的类别一侧才会对该分类误差项产生影响。由于分类误差项的存在, 该模型可以理解为在挑选特征的同时, 尽可能地保证被挑选得到的特征的判决能力。由于两类样本判决界的边界具有一定距离, 因此与大边界模型的学习方法类似, 即该模型特征选择方法具有一定的泛化能力。

需要说明的是, 类似于支持向量机的原理, 尽管在模型(1)中边界值为设定的数值–1 和 1, 但对该边界值任意进行缩放或平移并不影响选择结果, 因为使用值平移边界值时, 因为模型中偏置值b的存在, 在求解的过程中, 求解得到的模型最优解中的b值会自行进行相应的平移补偿。由于该模型选择的特征的个数通过调节λ并求解模型得到, 因此当边界值按照某个尺度缩放后, 只要将λ按照同样的因子缩放, 模型在求解后选择的特征与之相同, 该过程论述如下, 考虑将问题(1)边界值 1 和–1 替换成a和$-a$, 则问题(1)修改为以下模型问题(1)*:

$$\begin{aligned}&\min_{w,b,\xi^c,\xi^s}\lambda f^T|w|+\sum_{i=1}^N\xi_i^c+\xi_i^s \quad (1)^*\\ \text{s.t. }&w^Tx_i^c-b\leqslant-a+\xi_i^c,\qquad i=1,2,\cdots,N\\ &w^Tx_i^s-b\geqslant a-\xi_i^s,\qquad i=1,2,\cdots,N\\ &\xi_i^c\geqslant0,\xi_i^s\geqslant0,\qquad i=1,2,\cdots,N\end{aligned}$$

注意到a大于 0, 问题(1)*可写为

$$\begin{aligned}&\min_{w,b,\xi^c,\xi^s}a\lambda f^T\left|\frac{w}{a}\right|+a\sum_{i=1}^N\frac{\xi_i^c}{a}+\frac{\xi_i^s}{a} \quad (1)^*\\ \text{s.t. }&\frac{w^T}{a}x_i^c-\frac{b}{a}\leqslant-1+\frac{\xi_i^c}{a},\qquad i=1,2,\cdots,N\\ &\frac{w^T}{a}x_i^s-\frac{b}{a}\geqslant1-\frac{\xi_i^s}{a},\qquad i=1,2,\cdots,N\\ &\xi_i^c\geqslant0,\xi_i^s\geqslant0,\qquad i=1,2,\cdots,N\end{aligned}$$

令$w'=\frac{w}{a},b'=\frac{b}{a},\xi_i^{c'}=\frac{\xi_i^c}{a},\xi_i^{s'}=\frac{\xi_i^s}{a}$并注意到约束中的不等式两成乘以大于零的数值, 不等式仍成立, 且目标函数乘以一个大于零的因子, 目标函数求解的结果不变, 因此, 问题(1)*可写为:

$$\begin{aligned}&\min_{w',b',\xi^{c'},\xi^{s'}}a\lambda f^T|w'|+a\sum_{i=1}^N\xi_i^{c'}+\xi_i^{s'}\\ \text{s.t. }&w'^Tx_i^c-b'\leqslant-1+\xi_i^{c'},\qquad i=1,2,\cdots,N\\ &w'^Tx_i^s-b'\geqslant1-\xi_i^{s'},\qquad i=1,2,\cdots,N\\ &\xi_i^{c'}\geqslant0,\xi_i^{s'}\geqslant0,\qquad i=1,2,\cdots,N\end{aligned}$$

即与原问题(1)相同。

该特征选择的模型的目标函数中, 采用w的绝对值加权和, 这种方法对于特征选择具有较好的稀疏特性, 尽管也有其他的方法使得w中不为 0 的系数对目标将函数产生惩罚项的影响, 例如使用加权的平方和$\lambda f^T\|w\|^2,\|w\|^2=[w_1^2,w_2^2,\ldots,w_K^2]^T$, 尽管使用加权的平方和作为惩罚项在求解时具有便利性, 但当w中的元素小于 1 时, 取平方后的该数值较小, 对目标函数产生的惩罚影响较小, 因此使用二次方作为惩罚项时, 求解的w中多个元素难以趋向于 0, 即加权的平方和惩罚项不具备较好的特征选择特性。

4.3 特征选择模型的计算方法

本节主要介绍 4.2 节中(1)式描述的特征选择模型求解方法。由于(1)中具有$\lambda f^T|w|$项的存在, 因此该问题的目标函数不具备高阶光滑性。为此, 本文的方法中采用线性规划方法求解, 即将该问题转换为一个对偶的线性规划求解的问题, 通过一定的方法构造线性规划问题的目标函数和线性约束条件, 保证该线性规划问题的解与原特征选择问题(1)的解一致。观察(1)式, 其主要困难在于目标函数中对w的元素取绝对值, 因此, 对原问题的w中的每个元素分解为两个非负变量的相减, 并构建对偶的问题求解。为此, 引入$2d$个辅助变量$w_1^+,w_1^-,w_2^+,w_2^-,\cdots,w_d^+,w_d^-$, 令$w^+=[w_1^+,w_2^+,\cdots,w_d^+]^T$以及$w^-=[w_1^-,w_2^-,\cdots,w_d^-]^T$, 并构建以下线性规划问题:

$$\begin{aligned}&\min_{w^+,w^-,b,\xi}\lambda f^T(w^++w^-)+\sum_{i=1}^N\xi_i^c+\xi_i^s\\ \text{s.t. }&(w^+-w^-)^Tx_i^c-b\leqslant-1+\xi_i^c,\qquad i=1,2,\cdots,N\\ &(w^+-w^-)^Tx_i^s-b\geqslant1-\xi_i^s,\qquad i=1,2,\cdots,N\\ &w_1^+\geqslant0,\cdots,w_d^+\geqslant0\\ &w_1^-\geqslant0,\cdots,w_d^-\geqslant0\end{aligned}$$

$$\xi_i^c \geqslant 0, \xi_i^s \geqslant 0,\ i = 1,2,\cdots,N \qquad (2)$$

注意到, 在以上线性规划问题(2)中, 需要求解的变量w^+, w^-, b, ξ关于目标函数和约束条件均为线性关系或线性不等式关系, 因此可以使用线性规划的方法求解, 求解得到以上问题(2)的最优解w^+和w^-后, 令$w = w^+ - w^-$既是原特征选择问题(1)的最优解。本小节以下部分对该结论进行证明。

首先证明以下结论: 如果$w^{+*}, w^{-*}, \xi^{c*}, \xi^{s*}$是问题(2)的最优解, 则对任意的$1 \leqslant l \leqslant d, w_l^{+*}$和$w_l^{-*}$中不可能同时大于 0。假设在$w^{+*}$和$w^{-*}$中存在$w_l^{+*} > 0$且$w_l^{-*} > 0$, 则可据此构建另一个符合问题(2)约束条件但使其目标函数取值更小的可行解$\widetilde{w^+}, \widetilde{w^+}, \widetilde{\xi^c}, \widetilde{\xi^s}$, 其构建方法如下所示:

$$\widetilde{w_n^+} = w_n^{+*}, \widetilde{w_n^-} = w_n^{-*}, \quad \forall n \neq l$$
$$\widetilde{w_l^+} = \max(0, w_l^{+*} - w_l^{-*}),$$
$$\widetilde{w_l^-} = \min(0, w_l^{-*} - w_l^{+*})$$
$$\widetilde{\xi^c} = \xi^{c*}, \widetilde{\xi^s} = \xi^{s*}$$

在以上过程中, $\widetilde{w_l^+}$和$\widetilde{w_l^-}$与w_l^+和w_l^-不同, 且$\widetilde{w_l^+} - \widetilde{w_l^-} = w_l^+ - w_l^-$, 因此$\widetilde{w^+}, \widetilde{w^+}, \widetilde{\xi^c}, \widetilde{\xi^s}$均满足问题(2)的约束条件, 但另一方面, 由于$w_l^{+*} > 0$且$w_l^{-*} > 0$, 可得$\widetilde{w_l^+} + \widetilde{w_l^-} < w_l^{+*} + w_l^{-*}$, 因此, 将$\widetilde{w^+}, \widetilde{w^-}, \widetilde{\xi^c}, \widetilde{\xi^s}$代入问题(2)的目标函数, 可得以下结论:

$$\lambda f^T\left(\widetilde{w_l^+} + \widetilde{w_l^-}\right) + \sum_{i=1}^{N} \widetilde{\xi_i^c} + \widetilde{\xi_i^s}$$
$$= \lambda\left(\sum\nolimits_{i,i\neq l} f_i(w_i^{+*} + w_i^{-*}) + f_l\left(\widetilde{w_l^+} + \widetilde{w_l^-}\right)\right)$$
$$+ \sum_{i=1}^{N} \xi_i^{c*} + \xi_i^{s*}$$
$$< \lambda f^T(w^{+*} + w^{-*}) + \sum_{i=1}^{N} \xi_i^{c*} + \xi_i^{s*}$$

以上与$w^{+*}, w^{-*}, \xi^{c*}, \xi^{s*}$是问题(2)的最优解矛盾, 因此结论得证。

由以上论证可得知, 问题(2)求解的结果中, $w_i^{+*} + w_i^{-*} = |w_i^{+*} - w_i^{-*}|, \forall i$, 其所优化的目标函数实际为$\lambda f^T|w^+ - w^-| + \sum_{i=1}^{N} \xi_i^c + \xi_i^s$, 因此$w^{+*} - w^{-*}$与原问题(1)的解相同。

通过将问题(1)的求解转换为线性规划问题避免了求解过程中的困难, 使其可使用标准的线性规划方法求得精确解, 保证了求解的精度和效率。

4.4 结合特征选择的集成分类器方法

基于高维特征的隐写分析方法均需要使用集成分类器, 因此本文论述的特征选择方法需要与训练集成分类器的过程进行结合。在特征选择的过程中, 可采用有两种特征选择方案:

1) 从高维特征中选择一个特征子集, 作为集成分类器使用的特征, 集成分类器在训练时, 每个子分类器仅在该特征子集中随机抽取部分特征使用。

2) 集成分类器在训练时, 对于每个子分类器, 均在高维特征中随机抽取部分特征后, 再经过特征选择留下部分特征中的一个子集使用。

以上两种方法中, 第一种方法在特征选择时需要使用高维特征包含的所有特征, 因此容易引起过学习等问题, 且无法控制每个子分类器选择特征的差异性。因此本文采用第二种方法, 在第二种方法中, 由于会经过特征选择过程, 因此集成分类器的子分类器在此之前抽取的高维特征的部分特征维度相对较大, 以保证其抽取的特征子集中具有足够多的有效且多样的特征供特征选择。

根据集成学习的理论, 集成分类器的精度不仅取决于子分类器的决策精度, 还依赖于不同子分类器决策结果的多样性, 增加多样性可以显著提升检测精度。在高维特征隐写分析的集成分类器中, 子分类器的多样性程度主要与其抽取的特征子集大小有关。在未使用特征选择时, 子分类器抽取的特征子集越大, 则可利用的特征越多, 但与此同时, 不同的子分类器使用的特征子集相互重叠的特征也较多, 从而导致其多样性降低。因此子分类器在使用具有较好检测精度的特征的同时, 还需要保证不同子分类器所选择的特征的差异性。

本文的特征选择方法主要适应集成分类器在训练子分类器时的多样性和精度要求, 利用选择模型的误差项和惩罚项分别保证子分类器的检测精度和不同子分类选择特征的多样性。如图 2 所示, 结合了本文提出的特征选择方法的集成分类器在训练时, 仍然依次训练多个子分类器, 且每个子分类器仍从高维特征中随机抽取部分特征集, 但子分类器在训练前, 先进行特征选择, 即从抽取的特征中选择特征使用。利用模型惩罚项的中的特征选择权值f控制每个子分类器选择的特征的多样性, 当高维特征中某个特征被子分类器选择后, 其惩罚权值将增高, 因此该特征在随后的子分类器训练时, 如果被抽中, 则在其随后的子分类特征选择中被选择的可能性将降低。本文的方法中, 采用一个权值增长比例因子$\alpha > 1$进行控制, 当某个特征被选中后, 该特征的权值将会乘以α, 因此当该特征被下一个分类器在训练时被抽选到时, 在其特征选择阶段, 被选择的可能性将降低。

设高维特征总体维度为$K, F^t = \{F_1^t, F_2^t, \cdots, F_K^t\}$为

高维特征在训练第为t个子分类器时所有的特征的权值, 其中f^t为训练该分类器时, 在高维特征中抽取的部分特征所对应的权值组成的向量, 即f^t中的元素为F^t中的子集。第t个子分类器选择特征时, 在特征选择模型(2)中使用f^t作为其惩罚项的权值。设X_i^c和X_i^s分别为第i个正常训练样本和隐写训练样本提取的高维特征, 相应地, $x_{t,i}^c, x_{t,i}^s$和w^t分别为训练第t个子分类器时在X_i^c和X_i^s中抽取的特征子集组成的向量, 以及根据特征选择模型(2)求解得到的表示特征选择结果的权值向量。算法 1 为在集成分类器训练中采用特征选择的具体过程。

在算法 1 中, α的数值决定各个子分类器特征选择的差异性, 该参数越大, 则差异性越大, 反之越小。以上的参数中, Seed 可任意选择, 对分类器最终训练的结果影响较小。由于存在特征选择步骤, 因此在子分类器训练时, 一般将初次抽取的特征子集维度设置较大。在算法 1 中, 子分类器的个数、子分类器抽取的特征维数d、以及子分类器最终选择的特征子集维度d^*为输入参数, 该参数也可使用自动搜寻获取, 根据文献[14]中描述的方法, 在每次训练子分类器时均采用 bootstrap 策略以抽取放回的方式抽取部分样本训练子分类器, 留下未使用的样本评估其错误率, 将错误率达到最小且稳定时的参数作为最优参数。

算法 1: 基于线性规划的集成分类器特征选择方法

输入: 训练样本特征$X_i^c, X_i^s\ i = 1,2,\cdots,N$

特征抽取随机数生成种子 Seed

特征选择权值系数增长比例$\alpha, \alpha > 1$

子分类器抽取的特征维度d, 最终选择的特征维度d^*

输出: 集成分类器, 多个子分类器特征选择结果权值向量$w^t\ t = 1,2,\cdots,M$

初始化$F^1 = \{\frac{1}{K}, \frac{1}{K}, \cdots, \frac{1}{K}\}$, 根据 Seed 生成随机数序列$\{\text{seed}_t\}$, 每个元素作为初始化子分类器特征抽取种子。

For $t = 1$ to M

1. 根据seed_t抽取第t个分类器所使用的d维特征子集, 从$X_i^c, X_i^s\ i = 1,2,\cdots,N$抽取$x_{t,i}^c, x_{t,i}^s\ i = 1,2,\cdots,N$,在$F^t$中取出特征子集所对应的权值组成$f^t$。

2. 根据$f^t, x_{t,i}^c, x_{t,i}^s\ i = 1,2,\cdots,N$以及特征选择模型(2), 调节参数求解特征选择结果w^t, 使得w^t中不为 0 的元素个数为d^*个。

3. 根据特征选择结果w^t选择特征子集中的部分特征, 训练第t个子分类器。

4. 更新权值, 将最终被选择的特征在F^t中对应的权值乘以增长因子α, 得到$\widetilde{F^t}$

5. 将$\widetilde{F^t}$归一化, 结果作为F^{t+1}

end

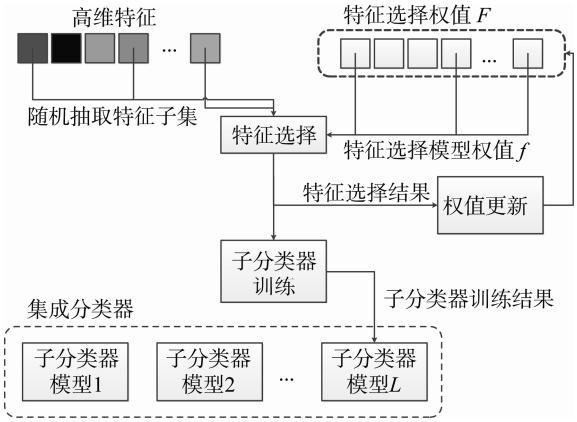


图 2 子分类器训练中特征选择权值更新示意图

Figure 2 Renewing weight for feature selection in base classifier training

5 实验结果

本章主要论述对本文提出方法的试验结果, 包含对实验环境和数据的说明、实验评价指标、对比试验数据等, 以验证本文方法的有效性。

5.1 实验数据

本文实验采用标准的图像数据库BOSS V1.02 [17]进行。BOSS V1.02 数据库是当前隐写术和隐写分析研究中评价算法性能的标准实验图像数据库, 该数据库包含 10000 张未经压缩的正常灰度图像, 格式为存储灰度空域图像的专用格式 PGM。BOSS 图像数据库中的图像由 6 款相机拍摄, 拍摄后通过统一的算法和裁剪处理成为统一尺寸为512×512。BOSS 图像数据库包含广泛题材的图像内容, 包含不同光照条件和地点的室内室外场景、建筑、动物、人物等。在本章实验中, 分别验证本文方法针对空域图像和 JPEG 图像进行隐写分析的性能, 因此需要准备空域格式和 JPEG 格式的正常图像。其中, 空域格式的正常图像直接使用 BOSS 图像数据库所包含的图像, JPEG 格式的正常图像由 BOSS 图像数据库图像通过 JPEG 压缩得到, 压缩质量因子分别取 95 和 85, 以验证本文的方法对不同品质因子的 JPEG 图像的隐写分析性能。因此, 正常样本一共有三组, 即空域图像、品质因子分别为 95 和 85 的 JPEG 图像。以上三组每组均包含 10000 张图像。

5.2 实验包含的隐写方法

本文实验所涉及隐写方法包含 5 种, 分别是针对空域图像的 WOW[18]、S-Uniward[19]、HILL[20]以及

针对 JPEG 图像的 UED[21]、J-Uniward[19]。以上 5 中隐写方法均为当前先进的自适应隐写方法中的代表性方法, 其安全程度较高。为生成实验使用的隐写图像样本, 分别采用以上 5 种方法, 对空域正常图像和 JPEG 正常图像进行嵌入, 并得到隐写图像。在嵌入的过程中, 对每一种算法均生成不同嵌入率的隐写图像, 以测试本文方法在不同嵌入率时的。嵌入率代表嵌入的消息的相对长度, 本文实验中, 嵌入率取值均为 0.1、0.2、0.3 三种。空域图像的嵌入率表示嵌入的信息比特数与图像尺寸之比, JPEG 图像的嵌入率表示嵌入的信息比特数与 JPEG 图像亮度通道非零交流系数个数之比。因此, 本文实验中包含的隐写图像共有 21 组, 即空域图像共 9 组(3 种算法、3 种嵌入率), JPEG 图像共 12 组(2 种算法、3 种嵌入率、2 种品质因子)。

5.3 实验设定和评价指标

本文的对比试验包括两种方法, 分别为仅使用集成分类器的方法, 以及本文提出的使用了特征选择集成分类器的方法。在同一组实验中, 两种方法采用的特征相同, 针对空域图像, 采用的特征均为 32000 维的 SRM 特征, 针对 JPEG 图像, 采用的特征均为 8000 维的 DCTR 特征。

由于分类器模型需要通过训练得到, 因此对于隐写样本和正常样本, 需要划分为训练样本和测试样本两个不相交的样本集, 测试样本用于测试隐写检测方法的精度。本文的实验中, 每一组实验均采用从样本集中随机抽取的正常样本和隐写样本各 5000 张, 剩余的正常样本和隐写样本各 5000 张作为测试样本。

实验的评价指标为检测正确率, 正确率越高即表示检测精度越好, 算法性能越高。正确率计算方法为对隐写和正常两类样本检测的正确率的平均值。每类样本的正确率即为该类测试样本中被正确检测样本个数的占比, 例如, 在用于测试的样本中, 共有正常样本和隐写样本各 5000 个, 其中正常样本中有 4800 个检测结果为正常样本, 隐写样本中有 4500 个被检测为隐写样本, 则对与正常样本和隐写样本的正确率分别为 96%、90%, 该算法的正确率为 93%。

在本方法中, 涉及控制不同子分类器特征选择模型中f^T更新的参数α, 如 4.4 节所示。参数α越大, 则越倾向于增加不同子分类器特征选择的多样性。为研究该参数的影响, 本实验测试对比不同α参数下的性能。

5.4 实验结果及分析

本节实验结果分为两部分, 第一部分为本算法与普通时用集成分类器的算法的对比, 对比试验结果如下表 1 所示, 表中最后两行分别为本文方法机器对比方法的检测准确率:

表 1 对比试验结果

Table 1 Comparison experimental results

隐写方法	品质因子	嵌入率	集成分类器方法	本文方法
WOW	\	0.1	63.6%	64.2%
		0.2	66.4%	67.6%
		0.3	72.6%	74.8%
S-Uniward	\	0.1	60.9%	61.6%
		0.2	66.0%	67.3%
		0.3	71.0%	72.6%
HILL	\	0.1	61.6%	62.5%
		0.2	64.4%	65.7%
		0.3	70.5%	72.2%
UED	85	0.1	77.0%	78.7%
		0.2	92.9%	93.3%
		0.3	98.3%	98.7%
	95	0.1	58.8%	60.2%
		0.2	75.8%	77.1%
		0.3	91.5%	92.8%
J-Uniward	85	0.1	70.3%	71.3%
		0.2	88.3%	89.9%
		0.3	96.1%	97.4%
	95	0.1	64.0%	65.3%
		0.2	79.7%	80.6%
		0.3	89.0%	90.4%

第二部分为本文方法在不同的参数α时的性能对比, 如表 2 所示:

从实验结果可知, 对于以上多种算法和图像格式的隐写分析中, 采用本文方法比不使用特征选择方法性能具有一定的提升, 因此验证了本文提出的特征选择方法的有效性, 在实验过程中, 子分类器抽取的特征子集维度以及从中进行特征选择后留下的特征集维度是关键参数, 该参数可以使用参数搜索的方法进行选择, 即根据文献[14]中所论述的方法, 每次在训练子分类器使用 bootstrap 策略以放回的方式抽取部分训练样本训练子分类器, 并利用余下的训练样本评估丢包错误率(Out-of-bag error), 取丢包错误率最低时候的参数。对于本文的实验中涉及到的两种检测方法, 在没有特征选择的原方法中, 特征抽取子集的维度和子分类器个数的最优参数由搜索得到, 在本文方法中, 特征抽取子集的维度d^*设定为原集成分类器方法搜索d最优值的 50%, 即特征选择后余下的特征维度为抽取子集的维度的一半, 采用该策略尽管并非为最优结果, 但节约了特征选

表 2 不同α参数检测性能

Table 2 Detection performance with different settings of parameter α

隐写方法	品质因子	嵌入率	α = 1.5	α = 2	α = 2.5	α = 3
WOW	\	0.1	64.0%	64.2%	63.9%	63.7%
		0.2	67.4%	67.6%	67.2%	66.8%
		0.3	74.2%	74.4%	74.8%	74.5%
S-Uniward	\	0.1	61.0%	61.2%	61.6%	61.1%
		0.2	67.1%	67.3%	66.8%	67.0%
		0.3	72.4%	72.6%	72.5%	72.3%
HILL	\	0.1	62.3%	62.5%	62.1%	62.2%
		0.2	65.4%	65.7%	65.5%	65.1%
		0.3	72.1%	72.2%	72.0%	72.0%
UED	85	0.1	76.7%	76.8%	77.0%	76.9%
		0.2	92.5%	92.7%	92.9%	92.6%
		0.3	98.0%	98.2%	98.3%	98.1%
	95	0.1	60.1%	60.2%	60.1%	60.0%
		0.2	77.0%	77.1%	76.8%	76.9%
		0.3	92.4%	92.6%	92.7%	92.8%
J-Uniward	85	0.1	71.3%	71.2%	71.0%	71.0%
		0.2	89.8%	89.9%	89.7%	89.6%
		0.3	96.8%	97.4%	97.0%	96.9%
	95	0.1	65.1%	65.3%	64.8%	64.6%
		0.2	80.1%	80.6%	80.5%	80.3%
		0.3	90.2%	90.4%	90.1%	90.0%

择的时间, 且实验结果证明该策略能够较好地提升性能, 因此也充分表明了该特征选择方法的合理性。

在训练阶段的耗时方面, 与不使用特征选择的集成分类器的隐写分析方法相比, 本文提出的方法训练时间较长, 其主要原因是求解过程中多次使用不同的参数进行特征选择以使得最终留下的特征为设定的维度。但在隐写检测过程中, 本文方法最终训练的集成分类器在形式上仍然是多个线性分类器的组合, 因此在实际应用过程中其效率与原方法相同, 具有较高的实时性。

6 结论

本文针对当前使用高维特征和集成分类器的隐写分析方法, 提出了一种特征选择方法, 提升了对多种图像格式、多种隐写术的隐写检测精度。本文的主要贡献主要为以下两点:

1) 针对集成分类器提出了特征选择的模型, 同时考虑了集成分类器训练子分类器的精度和多样性, 并将其融合在特征选择模型中。

2) 采用了高效的求解方法, 将选择模型转化为线性规划问题求解, 并在多个子分类器的特征选择过程中, 提出了权值更新的策略, 控制多个子分类器特征选择的多样性程度。

当前, 隐写分析仍然面临一些挑战, 如失配问题(Miss-matching)等, 因此该方面未来的工作将各方面多种问题考虑在内, 提升隐写分析的通用性。此外, 探索不同的特征选择准则也是发展方向之一。

致谢 本文工作得到国家自然科学基金(U1536105, U1636102), 国家重点研发计划(2016QY15Z2500, 2016QY15Z2500)资助

参考文献

[1] Y. J. Chanu, T. Tuithung and K. Manglem Singh, “A short survey on image steganography and steganalysis techniques,” Emerging Trends and Applications in Computer Science (NCETACS), 2012 3rd National Conference on, Shillong, pp. 52-55,2012.

[2] U. H. Panchal and R. Srivastava, “A Comprehensive Survey on Digital Image Watermarking Techniques,” Communication Systems and Network Technologies (CSNT), 2015 Fifth International Conference on, Gwalior, pp. 591-595, 2015.

[3] Y. Q. Shi; X. Li; X. Zhang; H. Wu; B. Ma, “Reversible Data Hiding: Advances in the Past Two Decades,” in IEEE Access , vol.PP, no.99, pp.1-1.

[4] H. Farid and S. Lyu, “Higher-order Wavelet Statistics and their Application to Digital Forensics,” Computer Vision and Pattern Recognition Workshop, 2003. CVPRW'03. Conference on, Madison, Wisconsin, USA, pp. 94-94, 2003.

[5] T. Pevny, P. Bas and J. Fridrich, “Steganalysis by Subtractive Pixel Adjacency Matrix,” in IEEE Transactions on Information Forensics and Security, vol. 5, no. 2, pp. 215-224, June 2010.

[6] J. Fridrich and J. Kodovsky, “Rich Models for Steganalysis of Digital Images,” in IEEE Transactions on Information Forensics and Security, vol. 7, no. 3, pp. 868-882, June 2012.

[7] V. Holub and J. Fridrich, “Random Projections of Residuals for Digital Image Steganalysis,” in IEEE Transactions on Information Forensics and Security, vol. 8, no. 12, pp. 1996-2006, Dec. 2013.

[8] Y. Shi, C. Chen, and W. Chen. “A markov process based approach to effective attacking jpeg steganography,” Information Hiding: 8th International Workshop, IH 2006, pp.249–264, 2007.

[9] T. Pevny and J. Fridrich. “Merging Markov and DCT Features for Multi-Class JPEG Steganalysis,” in Proc. SPIE Electronic Imaging, Photonics West, pp. 03-04 January 2007.

[10] J. Kodovský and J. Fridrich. “Calibration revisited,” in Proceedings of the 11th ACM workshop on Multimedia and security (MM&Sec '09). ACM, New York, NY, USA, pp.63-74. 2009.

[11] J. Kodovsky and J. Fridrich. “Steganalysis of JPEG Images Using Rich Models,” in Proc. SPIE, Electronic Imaging, Media Watermarking,

Security, and Forensics XIV, vol. 8303, San Francisco, CA, January 22-26, pp. 0A 1-13, 2012.

[12] V. Hulob and J. Fridrich. "Low-Complexity Features for JPEG Steganalysis Using Undecimated DCT," Information Forensics and Security, IEEE Transactions on, vol.10, no.2, pp.219-228, Feb. 2015.

[13] Xiaofeng Song, Fenlin Liu, Chunfang Yang, Xiangyang Luo, and Yi Zhang. "Steganalysis of Adaptive JPEG Steganography Using 2D Gabor Filters," in Proceedings of the 3rd ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec'15). ACM, New York, NY, USA, pp.15-23, 2015.

[14] J. Kodovsky, J. Fridrich, and V. Holub. "Ensemble classifiers for steganalysis of digital media," Information Forensics and Security, IEEE Transactions on, 7(2):432–444, April 2012.

[15] J. Dong, X. Chen, L. Guo and T. Tan, "Fusion Based Blind Image Steganalysis by Boosting Feature Selection," Digital Watermarking: 6th International Workshop, IWDW 2007 Guangzhou, China, December 3-5, pp.87-98, 2008.

[16] Y. Tan and F. Huang and J. Huang, "Feature Selection for High Dimensional Steganalysis," Digital-Forensics and Watermarking: 14th International Workshop, IWDW 2015, Tokyo, Japan, October 7-10, pp. 134-144, 2016.

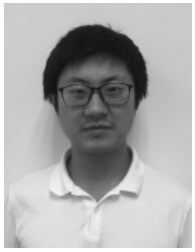
[17] J. Fridrich, J. Kodovsky, Holub. Vojtch and M. Goljan, "Breaking HUGO -- The Process Discovery" Information Hiding: 13th International Conference, IH 2011, Prague, Czech Republic, May pp.18-20, 2011.

[18] V. Holub and J. Fridrich, "Designing Steganographic Distortion Using Directional Filters," IEEE Workshop on Information Forensic and Security, Tenerife, Canary Islands, December 2–5, 2012.

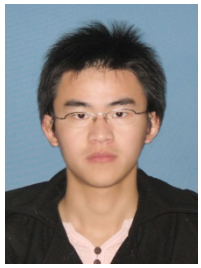
[19] V. Holub and J. Fridrich, "Digital image steganography using universal distortion," in Proceedings of the first ACM workshop on Information hiding and multimedia security (IH&MMSec'13). ACM, New York, NY, USA, pp.59-68, 2013.

[20] B. Li, M. Wang, and J. Huang. "A new cost function for spatial image steganography," in Proceedings IEEE, International Conference on Image Processing, ICIP, Paris, France, October 27–30, 2014.

[21] L. Guo, J. Ni and Y. Shi, "An efficient JPEG steganographic scheme using uniform embedding," 2012 IEEE International Workshop on Information Forensics and Security (WIFS), Tenerife, pp. 169-174, 2012.



关晴骁 于 2013 年在中国科学技术大学模式识别与智能系统专业获得博士学位。现任中国科学院信息工程研究所第一研究室助理研究员。研究领域为信息隐藏、模式识别。研究兴趣包括: 隐写术、隐写分析、图像取证检测等。Email: guanqingxiao@iie.ac.cn



朱杰 于 2011 年在华北电力大学软件工程专业获得学士学位。现在中国科学院信息工程研究所第一研究攻读博士学位。研究领域为信息安全、信息隐藏。研究兴趣包括: 隐写分析、深度学习。Email: zhujie@ iie.ac.cn

赵险峰 于 2003 年在上海交通大学计算机系统结构专业获得博士学位, 现为中国科学院信息工程研究所信息安全国家重点实验室研究员、博士生导师。主要研究领域为信息保密与内容安全防护, 包括: 信息隐藏、隐蔽通信及其检测, 内容伪造取证, 数字水印与数字版权保护, 内容安全标识及其管控, 多媒体特定内容与目标识别, 对抗情况下的机器学习, 隐私保护等。Email: zhaoxianfeng@iie.ac.cn

于海波 博士, 中国科学院软件研究所出站博士后,目前为中国科学院信息工程研究所正高级工程师, 博士生导师、硕士生导师。研究领域是网络空间信息对抗技术, 主要包括面向网络环境, 入侵检测、访问控制等, 主持或作为骨干成员参加国家级项目 9 项, 发表论文 10 余篇。Email: yuhaibo@iie.ac.cn

刘长军 2002 年毕业于中国人民解放军国防科学技术大学, 工学硕士。2003 年至 2013 年在总参第五十五研究所信息安全研究室工作, 高级工程师; 2013 年起至今在中国科学院信息工程研究所第一工程部工作, 高级工程师, 硕士生导师。主要从事网络与系统安全领域技术研究.
Email: liuchangjun@iie.ac.cn