基于半监督学习的JPEG图像通用隐写检测方法

陈东泳，钟尚平
（福州大学 数学与计算机科学学院，福建 福州 350002）

摘 要：目前，JPEG图像的通用隐写检测是基于监督学习的，其关键技术包括图像特征提取和分类器的设计。首次提出了运用半监督学习中的EM算法来进行分类器的设计，该方法利用大量未标记样本辅助少量有标记样本进行分类器的学习。针对经典的JPEG隐写方法：Outguess和F5，用监督学习与文中半监督学习方法进行实验对比，结果表明，在缺少大量标记样本的情况下，文中方法能得到较好的分类性能，从而提高了JPEG图像通用隐写检测方法的实用性。

关键词：通用隐写检测；JPEG图像；半监督学习；EM算法
中图分类号：TP391.41 **文献标识码**：A **文章编号**：1673-629X(2009)02-0169-04

A Universal Steganalysis Method for JPEG Images Based on Semi-supervised Learning

CHEN Dong-yong, ZHONG Shang-ping
(Dept. of Mathematic and Computer Science, Fuzhou University, Fuzhou 350002, China)

Abstract: At present, the universal steganalysis methods for JPEG images are based on supervised learning, the key technologies of these methods inclue image feature extraction and classifier design. Proposes a novel classifier which based on semi-supervised learning EM algorithm, the classifier makes use of a great quantity of unlabeled samples combined with a small quantity of labeled samples. Aimed at the popular JPEG steganography technologies: Outguess and F5, compared with this paper's method and the supervised learning method, the experimental results show that the proposed method can obtain good performance when a large number of labeled samples can't be obtained. Therefore, our method can improve the practicability for JPEG image universal steganalysis.

Key words: universal steganalysis; JPEG image; semi-supervised learning; EM algorithm

0 引 言

JPEG是一种Internet上使用非常广泛的图片格式，因此关于JPEG图像的数据隐写及其隐写分析已经成为当前研究的热点。隐写分析的目的是从载体（如图像、音频等）中发现是否存在隐藏信息，甚至只是指出载体中存在隐藏信息的可能性。隐写分析技术根据达到的效果可以分成攻击技术、检测技术和破解技术。目前的隐写分析技术主要集中在检测技术上，文中提到的隐写分析技术指的就是检测技术。隐写检测技术可以分为通用的隐写分析和针对某一种具体的隐写技术的隐写分析。文中讨论的是通用的隐写分析。隐写分析是基于学习的方法，实际上就是模式识别问题，即按照图像是否嵌入隐藏信息将图像分为两类：嵌入信息前和嵌入信息后。和一般模式识别问题相同，隐写分析的关键技术包括图像特征的提取和分类器的设计。特征的选取指的是从载体图像中提取对数据隐藏过程中较为敏感的特征，这是当前研究的热点。其中，比较具有代表性有：Avcibas[1]等人提出的基于图像质量统计量的检测算法；Farid[2]等人提出的基于小波域的统计特征的检测算法；Frdrich[3]等人提出了基于DCT域和DCT边界统计量的检测算法。而在文献[4]中，作者提出了基于DCT域共生矩阵的隐写分析方法，在实验中也取得了一定的效果。

关于分类器的设计目前大多采用传统的监督学习方法，此方法通过对大量有标记的训练样本进行学习，从而建立分类模型，然后运用此分类模型对待识别样本进行分类。虽然利用大规模有标记的样本可以提高学习算法的准确度，但是标记需要由人工完成，这是一项费时费力的工作（例如文献[4]，实验所用到的1096幅测试图像，需要拿其中的896幅对图像进行标记训

收稿日期：2008-06-24
基金项目：国家"863"高技术开发项目(2005AA142110)；福建省青年人才基金项目(2006F3076)
作者简介：陈东泳(1982-)，男，福建厦门人，硕士研究生，研究方向为网络与信息安全；钟尚平，博士，副教授，研究方向为网络信息安全。

练),同时也忽略了未标记样本的作用。而半监督学习是一种结合了少量标记样本和大量未标记样本的学习算法,近年来国内外对半监督学习已开展了大量研究,并取得了很多研究成果。例如 Nigam[5]等人用半监督学习的 EM 算法来进行文本分类研究;Zhou[6,7]等人将半监督学习中的协同训练方法引入基于内容学习的图像检索,提出了基于主动半监督学习的相关反馈方法。

文中首次将半监督学习算法引入 JPEG 图像通用隐写分析中,在文献[4]上的基础上,提出了利用半监督学习中的 EM(Expection Maximization)算法来设计分类器。实验结果表明,该方法在只有少量标记样本的情况下,可以利用大量未标记标本来改善分类性能。

1 基于 EM 算法的 JPEG 图像隐写分析

1.1 图像隐写分析过程

图 1 展示了一个典型的图像隐写分析过程。该过程主要由待识别图像获取、预处理、特征提取、分类决策和分类器设计五个单元组成。一般分为上下两大部分,上部分为分类器的训练过程,利用样本进行训练,确定分类器的具体参数,完成分类器的设计;下部分完成对待识别样本的分类。

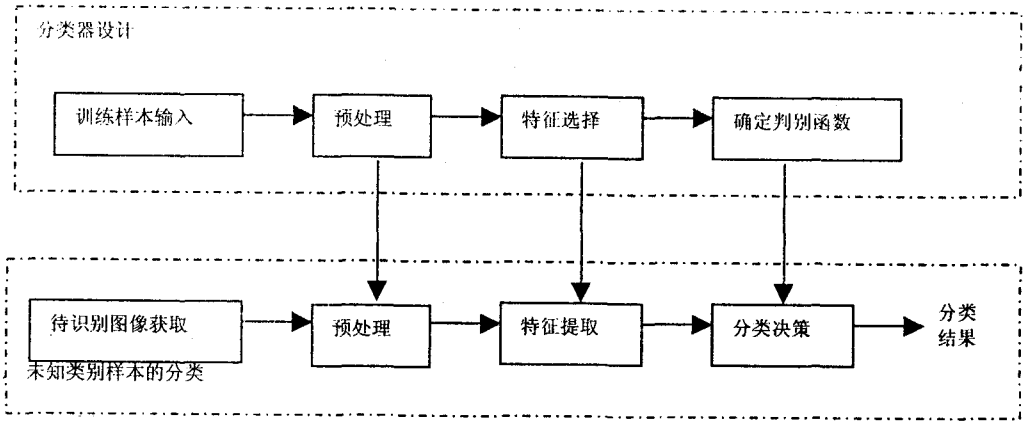

图 1 图像隐写分析过程

各个单元功能如下:

* 待识别图像获取:获取未知类别的图像样本,文中所研究的图像类型是 JPEG 图像;

* 预处理:对输入图像进行初步处理,提取有用的信息;

* 特征提取:对预处理后的图像数据进行变换,得到最能反映分类本质的特征;

* 分类决策:在特征空间将待识别的样本归为某一类别;

* 分类器设计:利用收集到的训练样本集来确定判别函数。

1.2 JPEG 图像特征提取

目前,大多数 JPEG 图像的隐写算法如 Jsteg[8],F5[9],Outguess[10]等都是根据一定的规则将隐藏数据嵌入在量化后的 DCT 系数上,而这些方法通常会改变载体图像的 DCT 系数直方图统计特征和分块特征,这为隐写分析提供了研究方向。文中采用了文献[4]的特征提取方法,直接在 DCT 域进行隐写分析,主要步骤如下:

Step1:直接读取 JPEG 图像的 DCT 系数,这里只取亮度通道 Y,没有取 UV。

Step2:对每个 8×8 的 DCT 分块以 ZigZag 顺序展开成一维向量 $\boldsymbol{V}_i(0,1,2,3,\cdots,63)$,取各分块的一维向量的低频部分 $\boldsymbol{V}_j(1,2,3,\cdots,20)$。

Step3:统计各分块一维向量的共生矩阵 $\boldsymbol{G}_i$,数值 $[-7,7]$ 外的系数不做统计。

Step4:计算全局平均共生矩阵 $\boldsymbol{G}=\frac{1}{n}(\sum_{i=1}^{n}\boldsymbol{G}_i)$,其中 n 为图像的分块数。因为共生矩阵具有对称性,因此,取 $\boldsymbol{G}$ 上三角的 120 维作为特征。

1.3 基于半监督学习 EM 算法的分类器设计

1.3.1 未标记样本对于分类性能的影响

由于隐写分析实际上属于模式识别问题,即可按照载体图像是否嵌入隐藏信息分为两类。隐写分析过程就是根据类别决策函数来对未知类别的图像进行分类。目前的分类器的训练均依赖于大量的人工标记样本,而忽略了未标记样本的作用。事实上,真实世界中通常存在大量的未标记样本,而有标记的样本则比较少。同时,随着技术的发展,收集大量未标记样本已经变得相当容易。半监督学习算法就是利用已标记样本和未标记样本来改善机器学习性能,并逐渐成为研究的热点。

关于利用未标记样本改善学习性能方面的研究,D. J. Miller 和 H. S. Uyar[11]从数据分布估计的角度给出了分析。他们假设所有数据服从于某个由 L 个高斯分布混合而成的分布,即

$$f(x\mid\theta)=\sum_{l=1}^{L}\alpha_l f(x\mid\theta_l) \tag{1}$$

其中 $\sum_{l=1}^{L}\alpha_l=1$ 为混合系数,$\theta=\{\theta_l\}$ 为参数。这样,标

记就可视为一个由选定的混合成分 m_i 和特征向量 x_i 以概率 $P(c_i \mid x_i, m_i)$ 决定的随机变量。根据最大后验概率假设，最优分类由下式给出：

$$h(x) = \arg\max_k \sum_j P(c_i = k \mid m_i = j, x_i) P(m_i = j \mid x_i) \quad (2)$$

其中 $P(m_i = j \mid x_i) = \dfrac{a_j f(x_i \mid \theta_j)}{\sum_{l=1}^{L} a_l f(x_i \mid \theta_l)}$

这样一来，学习的目标变成利用训练样本来估计 $P(c_i = k \mid m_i = j, x_i)$ 和 $P(m_i = j \mid x_i)$。显然，第一项和类别标记有关，而第二项并不依赖样本的标记。因此，如果有大量的未标记样，则 $P(m_i = j \mid x_i)$ 的估计会更准确，从而使得式(2) 更加准确。

1.3.2　基于 EM 算法的分类器设计

EM算法又称为最大期望算法，是一种被广泛使用的半监督学习算法。该算法由 Dempster[12] 等人提出，当某一数据模型丢失了某些数据的时候，EM 算法利用当前的模型的不完整数据，通过反复计算，对缺失数据获得最大的后验概率的估计，从而提高模型的性能。EM 算法主要分成两步：先用当前模型估算不完整数据缺少的值(E－step)，然后用这些数据改进模型(M－step)。

设 $Z^{(k)} = E[Z/D;\theta^{(k)}]$ 是第 k 代缺失的数据 z 的估计，$\theta^{(k)}$ 是第 k 代模型参数 θ 的估计，EM 算法过程：

1)E－step：计算 $Z^{(k+1)} = E[Z/D;\theta^{(k)}]$

2)M－step：计算 $\theta^{(k+1)} = \arg\max_\theta P(\theta/D;Z^{(k+1)})$

EM算法的底层分类模型为贝叶斯分类器。当把 EM 算法运用于 JPEG 图像分类的时候，模型参数 θ 为分类器的参数，缺失的数据 z 为未标记图像样本的类别，θ 的初值由已记的图像样本计算得到。设已标记的训练集合为 L，未标记的集合为 U，而且 $|U|$ 远大于 $|L|$，这里 $|L|$ 和 $|U|$ 分别为 L 和 U 的大小。EM 算法在训练模型参数的时候使用了 L 和 U。EM 算法将未标记的图像的类别视为不完整数据，通过迭代将 U 中图像转换为 L 中的图像，从而增大训练集 L 的规模，使得原来的训练集合中的数据 $L = \{L\}$ 扩大为 $L = \{L, U\}$。具体步骤如下：

(1) 输入标记数据集合 L 和未标记数据集合 U；

(2) 用已标记的数据集作为初始数据，初始化贝叶斯分类器；

(3) 重复下列两步，当分类器趋于稳定的时候终止循环：

E－step：使用分类器对未标记数据 U 进行软分类；

M－step：用所有已标记数据集合 L 和未标记数据集合 U 来重新估算分类器参数。

2　实验结果和分析

实验中使用的载体图像为从 Internet 上下载的动物主题和风景主题的 JPEG 图像各 1100 幅。图像的分辨率为 640 * 512 像素，并且使用质量因子 0.75 进行压缩。实验选择 F5 和 Outguess 作为信息嵌入算法，嵌入信息量选择 2kB(2048Bytes)，对于 Outguess，由于存在嵌入失败的问题，动物类和风景类图像成功嵌入的数目分别为 1045 幅和 1057 幅。

在采用文献[4]的图像特征提取方案的基础上，采用监督学习中的贝叶斯分类与基于 EM 算法的贝叶斯分类进行实验对比。

对于前者，分别随机抽取 200、400、600、800、1000 幅标记图像作分类器的训练样本，每次原图和嵌入图数量各占一半，同时随机抽取 100 幅原图和 100 幅嵌入图作测试。对于后者，在已有 100 幅标记图像(原图、嵌入图各 50 幅)基础上，随机抽取 400、600、800、1000 幅图像作为分类器的训练样本，同样随机抽取 100 幅原图和 100 幅嵌入图做测试。为了实验稳定性，程序重复运行 10 次，每次都重新随机选择训练和测试图像，求得平均识别率。实验结果分别显示于图 2～图 5。

从图中可以看到，虽然在具有相同训练样本数量的情况下，传统的贝叶斯识别效果优于基于 EM 算法的贝叶斯。但是它所采用的训练样本均为标记样本，在现实中，获取大量已标记数据是相当困难的，而获取大量未标记数据却是越来越容易。基于 EM 算法的贝叶斯在只需要少量标记样本的情况下，有效地利用了未标记样本，而且随着未标记样本数量的增加，它的识别率也逐步上升。

3　结束语

首次提出了利用半监督学习中的 EM 算法来进行

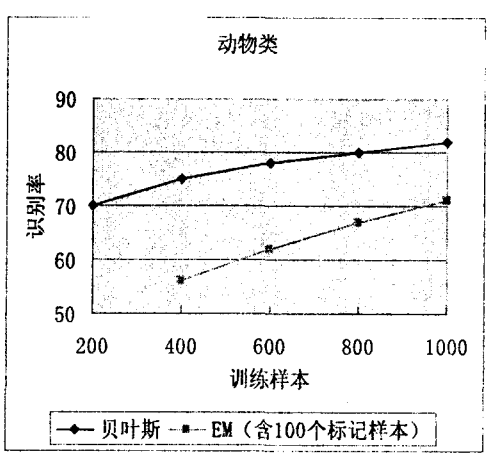

图 2　F5 测试结果(a)

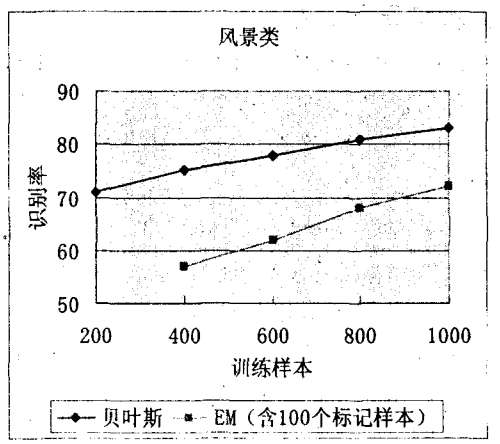

图3 F5测试结果(b)

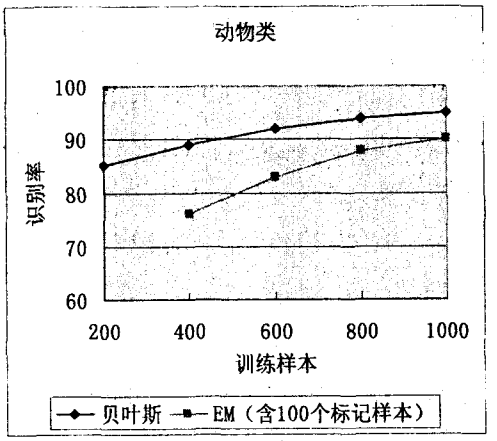

图4 Outguess测试结果(a)

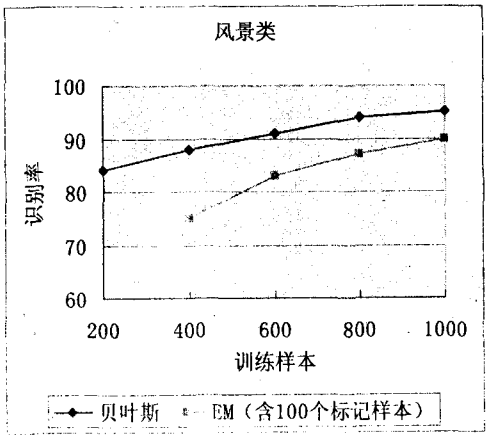

图5 Outguess测试结果(b)

JPEG图像隐写分析中的分类器设计。该方法在缺少大量标记样本的情况下，通过有效地利用未标记样本，在一定程度上提高分类性能，也更符合实际情况。

参考文献：

[1] Avcibas I, Memon N, Sankur B. Steganalysis using image quality metries[J]. IEEE Trans on Image Processing, 2003, 12(2):221-229.

[2] Farid H. Detecting hidden message using higher-order statistical models[C]// In Proc. IEEE Int. Conf. on Image Processing. New York:[s.n.], 2002:905-908.

[3] Fridrieh J. Feature-based steganalysis for JPEG images and its implications for future design of steganographic schemes [C]// In:6th International Hiding Workshop. Toronto, Ontario, Canada:Springer-Verlag, 2005:67-81.

[4] 黄 聪，宣国荣，高建炯，等. 基于DCT域共生矩阵的JPEG图像隐写分析[J]. 计算机应用，2006，26(12):2863-2865.

[5] Nigam K, McCallum A. Text Classification from Labeled and Unlabeled Documents using EM[J]. Machine Learning, 2000, 39(2):103-134.

[6] Zhou Y, Goldman S. Democratic co-learning[C]// In: Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'04). Boca Raton, FL:[s.n.], 2004:594-602.

[7] Zhou Z-H. Learning with unlabeled data and its application to image retrieval[C]// In: Proceedings of the 9th Pacific Rim International Conference on Artificial Intelligence (PRICAI'06). Guilin, China:[s.n.], 2006:5-10.

[8] Hsu C T, Wu J L. Hidden Digital Watermarks in Images[J]. IEEE Trans. On Image Processing, 1999, 8:58-68.

[9] Westfeld A. F5 a steganographic algorithm:High capacity despite better steganalysis[C]// In:4th International Workshop on Information Hiding. Pittsburgh, PA, USA:[s.n.], 2001:289-302.

[10] Provos N. Defending against statistical steganalysis[C]// In: 10th USENIX Security Symposium. Washington DC, USA:[s.n.], 2001:323-336.

[11] Miller D J, Uyar H S. A mixture of experts classifier with learning based on both labelled and unlabelled data[C]// In: Mozer M, Jordan M I, Petsche T, eds. Advances in Neural Information Processing Systems 9. Cambridge, MA: MIT Press, 1997:571-577.

[12] Despster P, Laird N M, Rubin D B. Maximum likeihood from incomplete data via the EM algorithm[J]. Royal Stat. Soc., 1997, 39(1):1-38.

(上接第168页)

of PKC 2005 (LNCS 3386), Jan 23-26, 2005, Les Diablerets, Switzerland. Berlin, Germany: Springer-Verlag, 2005:398-415.

[16] Shi Y, Li J. Provable efficient certificateless public key encryption. Cryptology ePrint Archive, Report 2005/287[EB/OL]. 2005. http://eprint.iacr.org/2005/287.

[17] Cheng Z H, Comley R. Efficient Certificateless Public Key Encryption. Cryptology ePrint Archive, Report 2005/012, 2005 [EB/OL]. 2005-12. http://eprint.iacr.org/2005/012/.