基于改进的局部敏感哈希算法实现图像型垃圾邮件过滤*

曹玉东，刘艳洋，贾 旭，王冬霞
（辽宁工业大学 电子与信息工程学院，辽宁 锦州 121001）

摘 要：提出一种快速的图像型垃圾邮件过滤方案，结合半监督机器学习技术改进局部敏感哈希（LSH）算法，基于改进的 LSH 算法构建垃圾图像特征库索引，提高图像的查找速度。搜集并构造了 60 000 个垃圾图像样本，实验结果表明，利用改进的 LSH 算法能有效地提高垃圾图像的过滤速度。

关键词：垃圾图像过滤；局部敏感哈希；图像特征提取；高维数据索引
中图分类号：TP393.098　　**文献标志码**：A　　**文章编号**：1001-3695(2016)06-1693-04
doi:10.3969/j.issn.1001-3695.2016.06.021

Image spam filtering with improved LSH algorithm

Cao Yudong, Liu Yanyang, Jia Xu, Wang Dongxia
(*College of Electronics & Information Engineering, Liaoning University of Technology, Jinzhou Liaoning* 121001, *China*)

Abstract: This paper presented a fast image-based spam filtering scheme. It improved LSH(locality sensitive hashing) algorithm based on semi-supervised technology, which was used to build an index of feature collection of 60 000 spam images. The experimental results show that the improved LSH can speed up the image spam filtering because the space complexity of improved LSH algorithm is lower than that of original LSH.

Key words: image spam filtering; LSH; image feature extraction; high dimension data index

0 引言

随着互联网技术的飞速发展，电子邮件以其方便、迅捷的特点逐渐成为人们日常工作和生活中的重要沟通工具，但是部分用户利用电子邮件发送违规广告、欺诈信息、谣言和反动言论。如何有效检测并过滤掉这些含有垃圾信息的电子邮件是电子邮件服务商面临的一个难题。为了逃避常规检测，垃圾邮件制造者把需要传播的文字信息嵌入图像中，然后再以正文或附件的形式发送出去。图像型垃圾邮件会消耗大量的网络带宽资源，侵占收件人信箱空间，骚扰邮件用户的正常工作和生活，甚至造成邮件服务器拥塞。如果暴力、恐怖分子通过图像型邮件传送或散布信息，将会危及社会和国家安全。近年来，垃圾邮件制造者刻意地对图像进行各种变化，增加了邮件监管的难度，同时垃圾邮件的数量呈现增长趋势，如何快速有效地检测出图像型垃圾邮件是一项有意义的研究课题。国内外很多学者一直在开展垃圾邮件过滤技术研究，文献[1,2]总结了最近十几年来图像型垃圾邮件过滤技术的发展情况；Liu 等人[3]使用扩展的角点和边缘特征实现垃圾邮件文本区域定位；罗常泳[4]提出一种基于邮件内容特征的结合正交质心特征选择算法，降低图像特征维数；刘芬等人[5]用梯度和颜色特征描述图像邮件，利用 SVM 算法实现分类；冯兵等人[6]利用灰度共生矩阵检测图像型垃圾邮件；刘艳洋等人[7]对比了图像梯度特征、颜色特征和 LBP 特征，探索了 SVM 算法中惩罚函数对垃圾邮件过滤效果的影响；林海卓等人[8]引入个性化推荐方法分析用户对垃圾邮件的感兴趣程度；王瑛等人[9]利用文本挖掘方法提升邮件自动分类效率；秦伟[10]利用光学字符识别技术实现图像型垃圾邮件过滤；Wang 等人[11]归纳了图像型垃圾邮件的各种构造方法。检测效果和检测速度很难兼顾，大规模的数据查询会降低检测速度，本文基于改进的局部敏感哈希（locality sensitive hashing, LSH）算法构造数据索引，在保证算法性能不降低的前提下，能提高过滤垃圾图像的速度。

1 垃圾邮件过滤技术方案

垃圾邮件图像具有如下特征：基于某一模板生成，并在某一段时间内重复发送，这些垃圾图像具有视觉相似性；垃圾图像在色彩的分布上不同于正常图像；垃圾邮件图像通常都包含人为的干扰，即在模板的基础上增加一些随机噪声[12]。

基于文本的邮件检测技术已经很成熟，邮件服务商可以通过分析邮件头的某些特征判断哪些邮件属于垃圾邮件，或者基于敏感词作出判定。如果根据上述方法判定该邮件是垃圾邮件，该邮件正文或者附件中携带的图像将极有可能是垃圾图像。如果图像型邮件没有文本型正文，同时也不能由邮件头信息分析和判定该邮件是否属于垃圾邮件，只能根据携带的图像作出判断。显而易见，如果查询图像是垃圾图像，同时垃圾图

收稿日期：2015-02-11；**修回日期**：2015-03-27　**基金项目**：国家自然科学基金资助项目(61272214)；辽宁省自然科学基金资助项目(201302022)；辽宁工业大学教师科研启动基金资助项目(X201216)

作者简介：曹玉东(1971-)，男，辽宁昌图人，副教授，博士，主要研究方向为模式识别与图像处理(cyd9229@163.com)；刘艳洋(1989-)，女，吉林长春人，硕士研究生，主要研究方向为图像识别；贾旭(1983-)，男，辽宁开原人，副教授，博士，主要研究方向为机器学习；王冬霞(1975-)，女，辽宁海城人，教授，博士，主要研究方向为多媒体信息处理.

像库规模足够大，容易找到相似的垃圾图像，进而判断查询图像也是垃圾图像。图 1 总结了图像型垃圾邮件的过滤方案。规模大的垃圾图像库会为查询的准确性提供保证，但线性查找过程是十分耗时的，最好的办法就是建立索引。

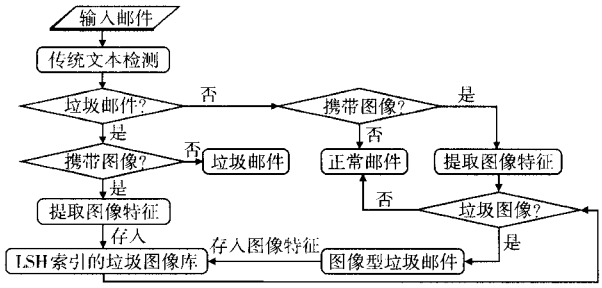

图 1 图像型垃圾邮件过滤框架图

1.1 高维数据索引

索引是一种用来查询记录的数据结构。由于图像的特征表示通常是几十维甚至上百维的高维数据，诸如 k-D-B tree 和 quad-tree 等传统的低维索引方法通常无法直接用于高维数据空间，即所谓的维数灾难问题。众多研究者已经给出了大量的索引方法，快速、准确地索引信息是非常重要的，哈希索引算法可以很好地解决这类问题[13]。Indyk 等人[14]于 1998 年提出了 LSH 算法，基于欧氏空间的 LSH 算法(Euclidean locality sensitive hashing，E2LSH)是应用非常广泛的近邻搜索算法[15,16]。随后众多学者对其作了深入研究[17,18]，并得到了广泛的应用[19~21]。

1.1.1 LSH 算法概述

LSH 算法的基本思想就是用随机的哈希函数值保证相似的数据点以很高的概率发生冲突而能够被检测到。先将高维图像特征点投影到某个特征空间，如果高维图像特征点与哈希函数矢量的内积，经量化后输出相同的值(也称为哈希码值)，则会被认为可能是近邻而散列到同一个哈希桶中，构造好数据索引结构以后，对查询图像也作相同的散列，穷举搜索与查询图像发生冲突的哈希桶，将会以较高的概率得到查询点的近邻。由于缩小了查找范围，LSH 算法能够在较短的时间内找到近邻。

哈希函数是单向的映射函数，在 LSH 算法中，哈希函数满足：

$$Pr_{h\in H}[h(u)=h(v)]=\mathrm{sim}(u,v) \tag{1}$$

其中：H 是哈希函数簇，哈希函数 h 从 H 中均匀选取，它把一个图像特征矢量映射为一个数；矢量 u 和 v 是两个数据点，$\mathrm{sim}(u,v)\in[0,1]$ 是相似度函数。

基于随机投影法[22]定义的哈希函数如下：

$$h(u)=\mathrm{sgn}(r\cdot u)=\begin{cases}1 & \text{if } r\cdot u\geqslant 0\\ 0 & \text{if } r\cdot u<0\end{cases} \tag{2}$$

矢量 r 的每一个元素服从标准正态分布，哈希函数的值取决于数据点 u 位于超平面的哪一侧。对于数据点 u_1 和 u_2 有

$$Pr(h(u_1)=h(u_2))=1-\frac{\theta(u_1,u_2)}{\pi} \tag{3}$$

基于 p 稳定分布的哈希函数[23]把数据库中的矢量点投影到随机的方向矢量 a_i 上，a_i 的每个元素服从 p 稳定分布(当 $p=1$ 时是柯西分布，$p=2$ 时就是标准高斯分布)。p 稳定分布具有如下的性质，如果两个变量服从 p 稳定分布，那么这两个变量的线性组合也服从 p 稳定分布。基于 p 稳定分布的哈希函数 $\{h:\mathrm{R}^d\to\mathrm{Z}\}$ 定义如下：

$$h(v)=\left\lfloor\frac{a_i\cdot v+b}{w}\right\rfloor \tag{4}$$

其中：内积 $a_i\cdot v$ 就是数据点 v 在 a_i 上的投影；w 表示投影窗口的量化宽度；b 的取值服从 $[0,w]$ 的均匀分布，参数 b 增强了哈希函数的随机性，有利于消除桶(bucket)边界的影响；符号 $\lfloor\ \rfloor$ 表示取整操作。对于数据点 v_1 和 v_2 有

$$p(c)=Pr(h_{a,b}(v_1)=h_{a,b}(v_2))=\int_0^w\frac{1}{c}f_p\left(\frac{x}{c}\right)\left(1-\frac{x}{r}\right)\mathrm{d}x \tag{5}$$

其中：$c=\|v_1-v_2\|_p$，$f_p(x)$ 是 p 稳定分布绝对值的概率密度函数。当数据库的规模很大时，LSH 的优势十分明显。

单个哈希函数 h_i 的区分性较弱，因此需要构建第二级哈希函数，其形式为

$$g_j(v)=\{h_{j,1}(v),\cdots,h_{j,k}(v)\}\qquad j=1,\cdots,l \tag{6}$$

式中由 k 个整数构成的矢量可以看做是数据点 v 的低维表示(有些文献称做哈希码)。局部敏感哈希函数簇表示为 $G=\{g:\mathrm{R}^d\to\mathrm{Z}^k\}$。

1.1.2 LSH 算法中的参数

在 LSH 索引结构中，参数 k 表示哈希码的维数，k 的取值越大，平均查全率(avera-ge recall rate)越低，平均查准率(average pricision rate)越高。k 到底取多大合适？如果不希望漏掉相似的查找对象可以取较小的 k 值，其代价是增加一些查找时间。参数 l 表示哈希表的构造数量，该参数取值的变化情况对平均查准率和平均查全率的影响正好与 k 相反。基于 p 稳定分布的哈希函数中还包含窗口参数 w，w 取值较大时查询结果的平均均值精度(mean average precision，mAP)较高[24]，当 w 的值大到一定程度时，查询性能没有明显的改善。对于不同规模的数据库和不同维数的图像特征矢量表示，参数的最优取值往往是不同的，量化宽度 w 和第二级哈希函数的数量 k 对哈希函数的性能影响也比较大。

1.2 结合半监督学习技术改进 LSH 算法

空间复杂度可以用哈希表的个数来表示，或者用内存总的使用情况来衡量[25]，用公式描述如下：

$$\text{内存消耗量}=O(l\times n) \tag{7}$$

其中：n 表示图像库规模，减少哈希表的数量能有效地降低算法的空间复杂度。提高 LSH 算法的性能可以从查询机制[26,27]和划分数据存储空间两方面考虑，设计或选择合理的哈希函数可以更好地划分数据空间。

随着因特网的迅速发展，比较容易获得大量的无标记样本，在仅有少量标记样本的情况下，是否可以再利用无标记样本提升学习性能[28]？半监督学习[29]就是利用大量无标记样本和少量有标记样本训练学习系统，让学习器自动地对大量未标记数据进行利用，辅助少量有标记数据进行学习，整个过程不需要人为干预。例如，在图 2 中有两类数据点需要分类，显然基于投影法的哈希函数 h_1、h_2 和 h_3 要比 h_4、h_5 和 h_6 好。基于 p 稳定分布的哈希函数也有类似结论。

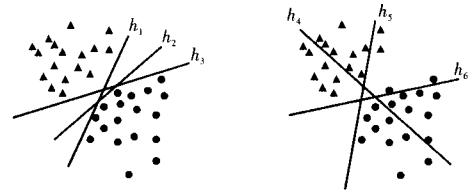

图 2 哈希函数分割数据空间示意图

结合半监督学习技术，利用标记样本，通过迭代方法选择更合适的哈希函数。如果随机生成的哈希函数足够多，标记的

训练样本又具有足够的代表性，那么最终的空间划分是很理想的。改进的 LSH 算法（improved LSH）描述如下：

```
1 input
  Ω = {(u_i, v_i)}：被标记的相似样本；
  w：未标记样本；
  Th：距离度量的阈值；
  q：查询数据点。
2 随机生成足够多的哈希函数 h（利用式(2)或(4)）；
3 初始化每个哈希函数，令 score(h) = 0；
4   for i = 1,2,… do
5     for j = 1,2,… do
6       if h_j(u_i) = h_j(v_i) then
7         score(h_j) = score(h_j) + 1；
8       end if
9     end for
10  end for
11 由大到小排序 score(h_j)，保留 l × k 个，构造 l 个二级哈希函数 g，得到 l 个哈希表 T_1, T_2, …, T_l；
12  for i = 1,2,… do
13    输入 w_i 作为查询点，w_i ∈ w；
14    for j = 1,2,…,l do
15      if w_i ∈ T_j 中第 m 个桶，在该桶中线性搜索近邻 x
16        if ‖w_i − x‖_2 < = Th
17          Ω = Ω ∪ {(w_i, x)}；
18        end if
19      go to 12；
20      end if
21    end for
22  end for
23 重复执行步骤 3 ~ 11，直到满足预定条件，构建新的哈希表 T_1, T_2, …, T_l；
24 输入查询点 q，在新的索引结构中查找最近邻 p；
25  if ‖q − p‖_2 < = Th
      判断 q 是垃圾图像；
26  end if
27  stop
```

在以上步骤中，生成哈希索引结构的过程可以离线完成。步骤 23 中的停止条件可以设置为迭代次数，或规定检测效果。

2　实验仿真

利用真实的电子邮箱收集了约 2 000 张垃圾图像，从Image spam hunter 数据集[30]下载到 928 幅垃圾图像。大量获取垃圾邮件图像比较困难，利用百度和谷歌的图片搜索功能从互联网下载了 57 000 多幅图像，然后人工在图像上添加不同类型的垃圾文字，并将文字作不同程度的变形处理，如改变字体颜色、字体倾斜、添加下划线和改变线型等，总计构造了60 000幅垃圾图像，最初的相似样本集 Ω 包含 100 对相似图像。

图 3 给出了部分垃圾图像的示意图。通过随机方式构造了 6 个规模不等的数据集，实验就是在这 6 个数据集上完成的。利用电子邮箱重新收集了 100 幅图像作为测试图像，包括 80 幅垃圾图像和 20 幅正常图像，测试图像完全独立于垃圾图像数据集。

图 3　垃圾图像库中的图像

使用 Gist 特征[31]描述两类图像之间的差异。Gist 将图像划分为均匀的网格，在不同的尺度和方向上对每个网格进行滤波，利用各方向和尺度滤波的平均值描述图像块的方向和轮廓信息，利用 RGB 通道的三维颜色均值表示图像块的表面信息，得到对该图像块的完整描述。Gist 描述了一系列感知维度，包括自然度（naturalness）、开放度（openness）、粗糙度（roughness）、扩展度（expansion）和光滑度（ruggedness）。这些感知维度代表了场景的空间结构，可以通过谱变换和粗略的局部定位信息估计出来。每幅图像用 320 维的 Gist 特征描述。

将垃圾图像定义为正样本，正常图像定义为负样本，给定一幅查询图像，返回的结果可以被标记为如下四种类型：实际为垃圾图像，被判定为垃圾图像（true positive，TP）；实际为正常图像，被判定为垃圾图像（false positive，FP）；实际为正常图像，被判定为正常图像（true negative，TN）；实际为垃圾图像，被判定为正常图像（flase negative，FN）。实验中使用准确率（accuracy，ACC）作为图像视觉相似性评价测度[32]。ACC 的计算公式定义为

$$\mathrm{ACC} = \frac{\#\mathrm{TP} + \#\mathrm{TN}}{\#\mathrm{TP} + \#\mathrm{TN} + \#\mathrm{FP} + \#\mathrm{FN}} \tag{8}$$

ROC 性能（receiver operating characteristic）[33]曲线反映检测率（true positive rate，TPR）和虚警率（false positive rate，FPR）之间的关系，TPR 和 FPR 的计算公式为

$$\mathrm{TPR} = \frac{\#\mathrm{TP}}{\#\mathrm{TP} + \#\mathrm{FN}},\quad \mathrm{FPR} = \frac{\#\mathrm{FP}}{\#\mathrm{FP} + \#\mathrm{TN}} \tag{9}$$

ROC 性能曲线对比如图 4 所示。

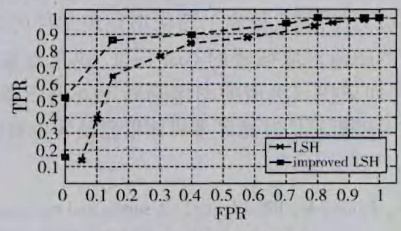

图 4　ROC 性能曲线对比

从图 4 中可以看出，改进的 LSH 的性能要好于 LSH，利用该 ROC 曲线确定阈值 Th 为 0.45。

表 1 ~ 3 分别给出了参数 l、k、w 的变化对标准 LSH 算法和改进 LSH 算法的准确率性能的影响，哈希函数基于式（4）的形式生成。从表 1 可以看出，当 l 的取值相同且均比较小时，改进算法的性能优于标准算法，如果维持相同的性能，标准 LSH 算法需要更多的哈希表，内存消耗量明显提高。

表 1　改变 l 的值对性能的影响（$k=3$，$w=0.1$）　/%

l	LSH_ACC	improved LSH_ACC
1	59.0	74.0
3	66.0	81.0
5	67.0	83.5
9	73.0	84.0
15	75.0	87.2

表 2　改变 k 的值对性能的影响（$l=3$，$w=0.1$）　/%

k	LSH_ACC	improved LSH_ACC
1	62.5	63.0
2	67.5	75.0
3	72.5	82.5
4	71.0	73.0
5	57.5	67.5

表 3　改变 w 的值对性能的影响（$l=3$，$k=3$）　/%

w	LSH_ACC	improved LSH_ACC
0.05	78.0	80.0
0.10	69.0	82.0
0.15	72.0	73.0
0.20	74.0	74.0
0.25	75.0	78.0

线性搜索（linear search）算法就是将查询图像数据和数

据库中的所有数据都比对一次，所以线性搜索算法可以作为理想的参照基准。图 5 比较了 LSH、改进的 LSH 和线性搜索算法的 accuracy 性能。LSH 算法和改进的 LSH 算法参数均设置为 $l=5, k=3, w=0.1$，可以看出改进的 LSH 算法的性能优于原来的 LSH 算法。

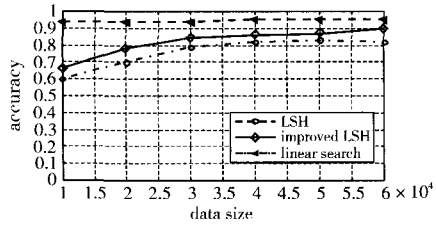

图 5 算法的 accuracy 性能比较

表 4 给出了三种算法在 60 000 幅垃圾图像库中查询一幅图像时的平均消耗时间，LSH 和改进 LSH 算法的参数仍设置为 $k=3, w=0.1$，可以看出改进后 LSH 算法的性能比标准 LSH 算法要好。

表 4 三种算法的查询时间比较

算法	t/ms	ACC /%
LSH($l=25$)	201.4	78.5
improved LSH($l=5$)	19.2	83.0
linear search	208	91.7

3 结束语

提出完整的图像型垃圾邮件过滤方案，基于改进的 LSH 算法构建垃圾图像索引，提高了图像型垃圾邮件的过滤速度。对邮件携带的图像提取稠密的 Gist 特征，快速搜索垃圾图像特征索引库，如果找到类似的图像则断定查询图像是垃圾图像。该方案对手机彩信过滤等多媒体信息监管也有借鉴意义。

参考文献：

[1] Biggio B, Fumera G, Pillai I, *et al*. A survey and experimental evaluation of image spam filtering techniques[J]. Pattern Recognition Letters, 2011, 32(10): 1436-1446.

[2] 李鹏, 崔刚. 图像型垃圾邮件过滤技术研究进展[J]. 智能计算机与应用, 2013, 3(3): 28-36.

[3] Liu Qiao, Qin Zhiguang, Chen Hongrong, *et al*. Efficient modeling of spam images[C]//Proc of the 3rd International Symposium on Intelligent Information Technology and Security Informatics. [S. l.]: IEEE Press, 2010: 663-666.

[4] 罗常泳. 基于内容的垃圾邮件检测方法研究[D]. 杭州：浙江大学，2014.

[5] 刘芬, 帅建梅. 基于梯度和颜色特征的图像垃圾邮件过滤[J]. 人工智能及识别技术, 2010, 36(16): 157-160.

[6] 冯兵, 李芝棠, 花广路. 基于灰度—梯度共生矩阵的图像型垃圾邮件识别方法[J]. 通信学报, 2013, 34(Z2): 2-4.

[7] 刘艳洋, 曹玉东, 贾旭. 基于内容的图像型垃圾邮件过滤技术研究[J]. 辽宁工业大学学报, 2014, 34(2): 86-90.

[8] 林海卓, 王继龙, 吴建平, 等. 高校误判垃圾邮件自动召回系统的研究与实现[J]. 通信学报, 2013 (S2): 121-132.

[9] 王瑛, 王勇. 电子邮件动态分类系统的研究与应用[J]. 自动化与信息工程, 2014(3): 7-13.

[10] 秦伟. 基于 OCR 的图像型垃圾邮件过滤系统研究[J]. 机械工程与自动化[J]. 2013(6): 184-185.

[11] Wang Zhe, Josephson W, Lyu Qin, *et al*. Filtering image spam with near-duplicate detection[C]// Proc of the 4th Conference on Email and Anti-Spam. 2007: 1-10.

[12] Mehta B, Nangia S, Gupta M, *et al*. Detecting image spam using visual features and near duplicate detection[C] // Proc of the 17th International Conference on World Wide Web. 2008: 497-506.

[13] 毛晓蛟, 杨育彬. 一种基于子空间学习的图像语义哈希索引方法[J]. 软件学报, 2014, 25(8): 1781-1793.

[14] Indyk P, Motwani R. Approximate nearest neighbor: towards removing the curse of dimensionality[C]//Proc of the 30th Annual ACM Symposium on Theory of Computing. 1998: 604-613.

[15] PauleveL, Jegou H, Amsaleg L. Locality sensitive hashing: a comparison of hash function types and querying mechanisms[J]. Pattern Recognition Letters, 2010, 31(11): 1348-1358.

[16] Gionis A, Indyk P, Motwani R. Similarity search in high dimensions via hashing[C]//Proc of the 25th International Conference on Very Large Data Bases. 1999: 1-19.

[17] Dasgupta A, Kumar R, Sarlos T. Fast locality-sensitive hashing [C]//Proc of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2011: 1073-1081.

[18] Andoni A. Nearest neighbor search: the old, the new, and the impossible[D]. Cambridge: MIT, 2009.

[19] 赵永威, 李弼程, 彭天强, 等. 一种基于随机化视觉词典组和查询扩展的目标检索方法[J]. 电子与信息学报, 2012, 34(5): 1154-1161.

[20] 高毫林, 彭天强, 李弼程, 等. 基于多表频繁项投票和桶映射链的快速检索方法[J]. 电子与信息学报, 2012, 34(11): 2574-2581.

[21] 魏晖, 杨高波, 夏明. 一种基于取证哈希的数字视频篡改取证方法[J]. 电子与信息学报, 2013, 35(12): 2934-2941.

[22] Charikar M S. Similarity estimation techniques from rounding algorithms[C]// Proc of the 34th Annual ACM Symposium on Theory of Computing. 2002: 380-388.

[23] Datar M, Immorlica N, Indyk P. Locality-sensitive hashing scheme based on p-stable distributions[C]// Proc of the Symposium on Computational Geometry. 2004: 253-262.

[24] 曹玉东, 刘福英. 基于局部敏感哈希算法的图像高维数据索引技术研究[J]. 辽宁工业大学学报, 2013, 33(1): 1-4.

[25] 蔡衡, 李舟军, 孙健, 等. 基于 LSH 的中文文本快速检索[J]. 计算机科学, 2009, 36(8): 201-204.

[26] Jegou H, Amsaleg L, Schmid C, *et al*. Query-adaptive locality sensitive hashing[C]// Proc of IEEE International Conference on Acoustics, Speech, and Signal Processing. [S. l.]: IEEE Press, 2008: 825-828.

[27] Joly A, Buisson O. A posteriori multi-probe locality sensitive hashing [C]//Proc of the 16th ACM International Conference on Multimedia. New York: ACM Press, 2008: 209-218.

[28] 周志华. 基于分歧的半监督学习[J]. 自动化学报, 2013, 39(11): 1871-1878.

[29] Chapelle O, Scholkopf B, Zien A. Semi-supervised learning[M]. Cambridge: MIT Press, 2006.

[30] Gao Yan, Yang Ming, Zhao Xiaonan, *et al*. Image spam hunter [C]//Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. [S. l.]: IEEE Press, 2008: 1765-1768.

[31] 高隽, 谢昭. 图像理解理论与方法[M]. 北京：科学出版社, 2009.

[32] Al-Duwairi B, Khater I, Al-Jarrah O. Detecting image spam using image texture features[J]. International Journal for Information Security Research, 2012, 2(3/4): 344-353.

[33] Fawcett T. An introduction to ROC analysis [J]. Pattern Recognition Letters, 2006, 27(8): 861-874.