图像型垃圾邮件过滤技术研究综述①

宋 文，张明新，彭太乐

(淮北师范大学 计算机科学与技术学院，淮北 235000)

摘 要：首先概述了图像型垃圾邮件的基本概念；其次根据不同的标准对图像型垃圾邮件过滤技术进行了分类，并评述了各种图像型垃圾邮件过滤方法和技术；对已经用于图像型垃圾邮件分类的两类共五种分类算法进行了理论分析与效果比较；最后对图像型垃圾邮件过滤技术的研究方向进行了展望。

关键词：图像型垃圾邮件；垃圾邮件图像；过滤技术；分类算法

Survey on Image-Based Spam Filtering Technology

SONG Wen, ZHANG Ming-Xing, PENG Tai-Le

(School of Computer Science and Technology, Huaibei Normal University, Huaibei 235000, China)

Abstract: This paper firstly introduced the basic concept of image-spam. And then it classified image-based spam filtering technology according to different standards. Meanwhile, it analyzed and evaluated the popular and major image-based spam filtering methods and technologies. Then it discussed and compared two categories five classification algorithms which have been used in image-based spam filtering were outlined. Finally, it gave some future directions of research on the techniques of image-based spam filtering.

Key words: image-based spam; spam images; filtering technology; classification algorithms

随着计算机网络的快速发展，电子邮件凭借其方便、快捷、低成本的优势，迅速成为人们日常工作、学习、生活的一个重要组成部分。然而，不知从何时起，我们的电子邮箱经常会收到不认识的人或地址发来的邮件。这些邮件多以各种广告信息为内容，例如机票打折、某种新版软件、新的销售网站等。这些用户并不想要、但被强塞给用户的邮件就是所谓的垃圾邮件（Spam）。垃圾邮件发送者（spamer）采用将纯文本转换为图像，或者把文本信息嵌入到图像中，让一些基于文本的过滤系统无法识别，这一类的垃圾邮件就是图像型垃圾邮件（I-Spam）。图像型垃圾邮件较一般的垃圾邮件消耗更多的网络资源，严重影响了网络的安全性、稳定性、高效性。

1 图像垃圾邮件（I-Spam）的定义

人们有效地界定一样事物的前提必须对这个事物做一个清楚、准确的定义，很多组织或机构都给邮件下过定义。比如，著名的反垃圾组织 spamhaus (www.spamhaus.com)提出，垃圾邮件具备以下两个特征：

1) 不请自来。用户事先并未提出要求或者同意接收该邮件；

2) 批量性。该邮件的副本在短时间内被大量发送给一个或多个用户。

中国互联网协会定义的垃圾邮件则包括：

1) 收件人事先没有提出要求或同意接收的广告、电子刊物、各种形式的宣传品等宣传性的电子邮件。

2) 收件人无法拒收的电子邮件。

3) 隐藏发件人身份、地址、标题等信息的电子邮件。

4) 含有虚假的信息源、发件人、路由等信息的电子信息。

图像垃圾邮件是垃圾邮件的一个变种，到目前为止，还没有一个统一的权威的定义。图像垃圾邮件是

① 基金项目:安徽省高校自然科学研究重点项目(KJ2010A304);2011 年度淮北师范大学青年科研项目(2011XQXM45)
收稿时间:2011-03-07;收到修改稿时间:2011-04-01

指把垃圾信息嵌入到附件图像的电子邮件，但这种定义不够全面，仅说出了图像垃圾邮件的一方面特性，为此为图像垃圾邮件作了如下定义。

定义 1. 赖均[1]将以下三类图像定义为垃圾邮件图像：具有特定目的的政治、宗教图像；具有商业目的的宣传图像；色情图片。

定义 2. Mark Dredze[2]等人则简单地将垃圾邮件图像定义为：如果一副图像包含在垃圾邮件之中，则认为该图像为垃圾邮件图像。

2 图像垃圾邮件（I-Spam）过滤技术的分类

到目前为止，反图像垃圾邮件过滤的方法主要有文本过滤方法、光学字符识别（Optical Character Recongnition OCR）技术、图像属性分析法、图像内容分析法等。虽然很多方法与技术已经取得很好的使用效果，但是图像垃圾邮件制造者仍然在挖空心思制造更“合理”或者严重干扰过滤器的图像垃圾邮件，所以在图像垃圾邮件过滤技术的研究中，仍然有很多问题要解决。

下面将图像垃圾邮件的过滤方法按照不同角度进行分类，可有以下几种类别：

2.1 根据分类算法的不同划分

根据分类算法的不同，可将图像垃圾邮件的过滤技术分为两类：a)基于规则(集)匹配的图像垃圾邮件过滤；b)基于内容统计学习的图像垃圾邮件过滤。

基于规则(集)匹配的方法：这是一种基于内容的模式匹配，是在邮件内容中寻找特定的模式，例如主题包含“免费”。当一封收到的邮件匹配上某些预定义的规则以后，则系统将其自动分类为垃圾邮件或正常邮件。邮件服务器或用户也可以制订一些复合的规则，拒收或过滤符合规则的邮件。该方法的优点是容易生成人类易于理解的规则，且规则可以共享，因此它的推广性很强。一个人写出的规则可以提供给多个人，多个服务器使用。缺点是要求其应用领域具有明显的规律性，更新速度慢。因为规则一般都是人工编写生成，所以新规则的产生速度跟不上新垃圾邮件出现的速度，换句话说，它的时效性较差。著名的规则方法有 Ripper、决策树方法、粗糙集方法等。

基于内容统计学习的过滤方法：该技术源自机器学习，是使用统计方法解决邮件的二元分类问题，采用机器学习、文本分类技术自动进行邮件分拣。其核心思想是采用某个分类算法(如常用的贝叶斯算法等)对已知的垃圾邮件样本进行学习，提取垃圾邮件的特征，构造过滤器；然后运用此过滤器，分类机对新的邮件进行判断，并自动分拣；过滤的结果提交给用户，用户可对过滤结果进行反馈，系统再根据反馈对过滤器进行调整。该方法的优点就是分类机由程序自动训练出来，只要及时更新样本训练集就可以使分类机更新的速度跟得上垃圾邮件出现的速度，即它的时效性很强。缺点就是分类机不能共享，某个用户用自己的邮件样本集训练出来的分类机对其他用户可能效果不佳，因此该方法的推广性较差。常用的基于统计的分类方法有贝叶斯(Naive Bayes)、支持向量机 SVM (Support Vector Machine)、K 近邻法 KNN(K-nearest Neighbor)、最大熵值法等。

2.2 根据垃圾邮件的检测位置划分

从垃圾邮件的检测位置出发，可将垃圾邮件的过滤技术分为两类：a)基于拦截的过滤方法；b)基于检测的件过滤方法。

基于拦截的过滤方法：该方法采用从源头上发现垃圾邮件，阻止垃圾邮件到达收件方服务器，通常是在收件人端的服务器配置过滤器对邮件做分析，例如检查分析邮件信头、信体及附件。当电子邮件被转发给最终用户时，主题上就会被加上“这可能是垃圾邮件”的字样。尽管过滤器是有用的，但垃圾邮件发送者会不断地审查哪些内容会被过滤、哪些内容不会被过滤，并采取措施躲避过滤器的检查。

基于检测的过滤方法：该方法采用在收件人的服务器端对所有的邮件做检测以确定是否为垃圾邮件。收件人可以采用客户端过滤软件，利用其分拣和过滤功能来设定规则，把接收下来的邮件进行检查和匹配，从发件地址、主题、正文内容中的关键词，对那些符合垃圾邮件特征的邮件执行特定操作。

2.3 根据所使用的特征划分

从使用的特征出发，可将图像垃圾邮件的过滤技术分为三类：a)基于行为特征的过滤方法；b)基于元数据特征的过滤方法；c)基于图像内容特征的过滤方法。

基于行为特征的过滤方法：该方法是以发件人的输出消息作为处理对象，分析发件人近期的邮件行为特征，为发件人建立邮件行为模式，即获取输入消息的特征以达到过滤垃圾邮件的目的。如发件人所发出的邮件中是否使用了 HTML、是否出现脚本标签（用于检

测邮件中是否存在安全隐患)、是否包含超链接(这些超链接是蠕虫病毒传播的有效途径)、是否包含附件(收件人打开一个附件可能会产生蠕虫传播)等,这些发件人的邮件行为特征可作为图像垃圾邮件的过滤特征,同时可以有效地检测各种病毒,较早的发现一些新的病毒。

基于元数据特征的过滤方法:该方法在处理的时候不涉及到邮件的内容,获取的特征仅仅是邮件头信息和图像文件的元数据,而邮件头信息和图像文件的元数据就构成了图像邮件的元数据特征。由于 Internet 上未经授权的商业邮件数量迅猛增长,因此在邮件头中提供虚假信息的行为也日益增长,这也称为欺骗,发件人通过一些软件或者编程手法,随意修改邮件头中的信息,比如冒用他人域名以及账号,或者隐藏自己的域名等。邮件头信息包含收件人、发件人、主题、抄送和秘密抄送等字段,这些信息对于辨别电子邮件的问题或者辨别未经授权的商务邮件的来源非常有用。图像文件的元数据则是存在于图像文件头中的一些相关信息,包括图像的高、宽、面积、帧数、颜色表、索引值(万明成 等.2008)、图像的大小、图像文件的格式、图像压缩比(Sven Krasser,et al.2007)。其中图像文件大小、图像面积、图像压缩比等特征是图像垃圾邮件过滤中区分能力较强的图像文件的元数据特征,因为图像垃圾邮件与正常邮件相比其具有图像文件体积较小、图像面积较大、图像压缩率高等显著特征。从以上看出,元数据特征可作为图像垃圾邮件过滤中的过滤特征,并且提取不涉及图像处理算法,速度快。

基于图像内容特征的过滤方法:该方法是对图像垃圾邮件进行图像内容分析,对图像特征进行提取并表示出来,利用提取出来的图像特征作为判断依据,结合决策树、支持向量机等分类算法对图像垃圾邮件进行识别,从而确定是否为图像垃圾邮件。因图像垃圾邮件与正常邮件相比其文字多且有明显的文字边缘和纹理特征、色彩鲜艳且颜色分布不均匀、含有随机噪声等显著特征,所以图像内容特征可作为图像垃圾邮件的过滤特征。图像内容特征主要包括:图像特征、图像中的文本区域特征以及图像的元数据特征。其中图像特征是图像的底层特征,如:颜色特征、纹理特征、形状特征、噪音特征。图像中的文本区域特征包括:文本区域的面积、文本区域的数量、文本区域的文字数量、文本区域的颜色数量、文本区域的色饱和度。

3 图像垃圾邮件(I-Spam)分类算法

分类算法的选择是图像垃圾邮件过滤中关键的阶段,其选取的好坏直接关系到整个过滤算法的精度和实用性,目前用于图像型垃圾邮件检测算法中的分类算法主要有两类:第一类为基于统计学的分类算法,第二类为基于规则的分类算法。

3.1 基于统计学的分类算法

此类方法在图像型垃圾邮件检测上应用较多,主要有支持向量机、最大熵模型、D-S 证据理论、贝叶斯算法。

1) 支持向量机(Support Vector Machine,简称 SVM,也叫做支撑向量机):是在二十世纪 90 年代以来发展起来的一种统计学习方法,它通过构造最优线性分类面来指导分类,在解决小样本学习、非线性及高维模式识别问题中表现较好。在基于文本内容的垃圾邮件分类中,SVM 是公认的较好的方法之一。在图像型垃圾邮件的分类中,支持向量机分类算法也取得了较好的效果[3,4]。

2) 最大熵模型(Maximum Entropy Model)[5]:当需要对一个随机事件的概率分布进行预测时,预测应当满足全部已知的条件,而对未知的情况不要做任何主观假设。在这种情况下,概率分布最均匀,预测的风险最小。最大熵模型主要应用在基于文本内容的垃圾邮件分类上,并且取得了一定的效果[6,7]。Mark Dredze[8]在 2007 年将最大熵模型应用在图像型垃圾邮件的分类上,取得了较好的效果。

3) D-S 证据理论: Dempster 于 1967 年提出,其学生 Shafer 对 D-S 证据理论进行了深入的研究和分析,形成了一套完整的数学推理理论,成为 Dempster 合成法则。D-S 证据理论主要应用在信息融合、不确定推理、模式识别以及决策系统等领域。也有学者将其引入到垃圾信息过滤和图像识别上,例如传统的基于文本内容的过滤算法和色情图片检测算法[9,10],图像识别算法[11],都取得了较好的效果。

4) 贝叶斯算法:是一种基于概率的算法,由伟大的数学家 Thomas Bayes 所创立的,此算法应用较为广泛。Paul Graham 在 2002 年将贝叶斯算法引入垃圾信息过滤中来[12]。目前主流的基于文本内容的垃圾邮件检测算法大都是基于贝叶斯算法或者有贝叶斯算法参与

的。在智能邮件过滤技术中，贝叶斯算法取得了较大的成功。但在图像型垃圾邮件的检测上，贝叶斯算法效果不佳。

3.2 基于规则的分类算法

此类算法的代表就是决策树算法，决策树是一种对数据进行分类或划分的方法，是用在图像型垃圾邮件检测算法上的基于规则的分类算法的代表。Krasser Sven 等在 2007 年将决策树算法引入到图像型垃圾邮件的分类上，但效果相比支持向量机分类算法较差[13]。

3.3 算法对比

国内外许多学者都提出了针对自己特征集合的图像型垃圾邮件检测算法，提取的特征集合的不同、采用的分类算法的不同等都导致了这些检测算法性能之间的差异。

表 1　算法性能比较

学者	Mark Dredze			Byungki Byun	Zhe Wang
像特征集合	图像的颜色特征、文件属性特征、边缘特征以及随机像素测试结果			颜色距、颜色异质性颜色的显著性以及自相似性	图像的颜色直方图特征、小波变化特征以及边缘特征
类算法	最大熵模型	朴素贝叶斯	ID3 决策树	多类别分类算法	计算待测样本和已有特征集合的距离
精确率	97%	85%	93%	86.6%	82%

可见，在图像垃圾邮件过滤中为满足实时性和精确性的要求，应选出不同的特征集合并筛选出具有最优分类性能的特征子集，确定适合最优特征子集的分类算法。

4　结语

图像垃圾邮件的出现和泛滥，对现有的垃圾邮件过滤方法提出了诸多有待解决的问题，今后还有大量的工作值得去做：a) 构建高质量的图像特征样本库；b)寻求准确快速的图像特征提取方法，并将图像的语义特征结合到图像垃圾邮件过滤中；c) 确定合适的垃圾邮件相似性阈值等。

参考文献

1 赖均.反垃圾邮件技术的研究和原型实现[硕士学位论文].成都:电子科技大学,2005.

2 Dredze M, Gevaryahu R, Elias-Bachrach A. Learning Fast Classifiers for Image Spam.CEAS 2007–Fourth Conference on Email and AntiSpam, 2007.

3 Zuo HQ, Hu WM, Wu O, Chen YF, Luo G. Detecting image spam using local invariant features and pyramid match kernel. Proc. of the 18th International Conference on World Wide Web. 2009.

4 Cheng HR, Qin ZG, Liu Q, Wan MC. Spam Image Discrimination using Support Vector Machine based on Higher-Order Local Autocorrelation Feature Extraction. 2008 IEEE Conference on Cybernetics and Intelligent Systems, 2008.1017–1021.

5 Berger AL, Pietra SAD. A Maximum Entropy Approach to Nature Language Processing. Computational Linguistics, March 1996.

6 Zhang L, Yao TS. Filtering junk mail with a maximum entropy model. Proc. of 20th International Conference on Computer Processing of Oriental Languages (ICCPOL03). 2003.

7 陈文庆,李勤,姚伽华.基于最大熵模型的垃圾邮件过滤方法.网络安全技术应用,2005.

8 Dredze M, Gevaryahu R, Elias-Bachrach A. Learning Fast Classifiers for Image Spam.CEAS 2007-Fourth Conference on Email and AntiSpam, 2007.

9 尹慧琳,王磊.D-S 证据推理改进方法综述.计算机工程与应用,2005.

10 李茹,李弼程.基于 D-S 证据理论的邮件筛选方法.计算机工程与设计,2005,26(10):2833–2836.

11 王嵘,马希荣.基于 D-S 证据理论的表情识别技术.计算机科学,2009.

12 Paul G. Aplan for spam. http://paulgraham.com/spam.html

13 Sven K, Tang YC, Jeremy G, Alpervitch D, Judge P. Identifying Image Spam based on Header and File Properties using C4.5 Decision Trees and Support Vector Machine Learning. IEEE SMC Information Assurance and Security Workshop, 2007.255–261.