基于免疫机制的智能反垃圾邮件过滤器的研究*)

王 磊 董 菡 银彩燕

(西安理工大学计算机科学与工程学院 西安 710048)

摘 要 针对传统的反垃圾邮件过滤技术不能有效识别未知特征及变异特征的问题,本文借用人工免疫系统的一些原理,通过引入诸如自体耐受、免疫识别、免疫学习、免疫记忆和协同刺激等机制,结合智能决策的思想,运用粗糙集理论提取特征集,更新特征库,提出了一个基于免疫功能的智能反垃圾邮件过滤器。仿真试验表明,该方法能够有效地降低误报率,具有较好的使用效果。

关键词 邮件过滤,垃圾邮件,粗糙集,人工免疫系统,自适应

Study on Immune Mechanism Based Intelligent Spam Filter

WANG Lei DONG Han YIN Cai-Yan

(College of Computer Science and Engineering, Xi'an University of Technology, Xi'an 710048)

Abstract With regard to the problem that the traditional anti-spam filtering mechanism is incapable of recognizing unbeknown and mutational character, a novel anti-spam filter based on immunity is constructed with employing multiple principles and mechanisms of the artificial immune system, such as the self tolerance, immune recognition, the immune learning, the immune memory and coordinated stimulating, combine ideology of intelligent decision support system based on extension of classical rough set theory, we give an efficiency algorithm of attributes reduction, find minimal attributes extraction and update the attributes base. With the simulation research and many experiments, it demonstrates that the technique is able to reduce and improve the availability of anti-spam system.

Keywords E-mail filtering, Junk E-mail, Rough set, Artificial immune system, Self-adaptive

互联网上垃圾邮件泛滥,给人们的学习和工作带来很大的不便并极大地消耗了网络资源。据著名网络安全公司 Sophos 发布的报告显示,世界上约 70%的电子邮件是垃圾邮件。2007 年 4 月至 6 月,全球垃圾邮件较去年同期上升了九个百分点,中美两国并列全球垃圾邮件的"老大"位置。垃圾邮件是传播病毒、木马程序的主要途径,邮件中涉及到的诈骗、色情、暴力等内容给社会带来很大的危害。1999 年 2 月,Internet 协会 ISOC (Internet Society)发布了 RFC 2502(Anti-Spam Recommendations for SMTP MTAs),标志着垃圾邮件已经成为信息安全领域的重要研究课题。面对越来越猖狂的垃圾邮件,不论从经济、政治还是法律的角度,如何有效识别和拦截垃圾邮件成为当今网络安全技术领域迫在眉睫的研究课题。

目前,常见的反垃圾邮件技术可分为 4 类:过滤器[1]、验证查询、挑战(challenges)和密码术(cryptography)。验证查询、挑战、密码术都需要邮件服务器的介入,用户是被动接受方,不能根据个人观点来决定邮件的性质。过滤器是一种相对来说最简单却很直接的处理垃圾邮件的技术,主要用于客户端,常用的过滤技术有黑名单、白名单、HASH 技术、基于规则的过滤技术、Bayesian 过滤技术和神经网络技术等[2],这些技术沿用了类似杀病毒系统的原理,实现容易,对于已知的垃圾邮件特征具有较好的性能,但还是存在特征变种绕过、误报、过滤器复查等问题,往往存在维护困难、实时性低、缺乏反馈机制等缺陷。

另一方面,人工免疫系统作为目前人工智能研究的一个新领域,提供了一种强大的信息处理和问题求解的方式,在信息安全、机器学习、故障诊断、数据挖掘等领域有着广阔的研究应用前景[3]。该系统采用智能决策支持的框架,运用粗糙集理论来实现抗原特征核值的提取,生成规则,更新成熟细胞库,借鉴区分矩阵算法确定特征的权重。提出一种基于多个免疫机制的智能反垃圾邮件过滤器(Intelligent Immunity-based Spam Filter, IISF)。文中在介绍该模型所应用的免疫学原理及机制的基础上,描述了模型的实现过程,通过仿真实验的分析,发现该模型具有良好的动态性和自适应性,为该模型的推广使用创造了条件。

1 基于免疫的智能反垃圾邮件过滤器

1.1 IISF 的提出

传统的邮件过滤器以经验知识为依据,采用适当的技术手段,代替用户对未知类别的邮件进行分类,这样的过滤器都是静态的。而用户对于垃圾邮件的定义处在不断的变化中,例如对于关键词"化妆品",一般来讲用户将包含该词的邮件看作具有商业性质的垃圾邮件而拒收。然而一旦用户需要购买化妆品或者查询相关情况的时候,含有该词的邮件就不再被用户定义为垃圾邮件,因此邮件过滤器必须具有根据用户反馈进行动态学习的功能。再比如针对关键字列表,垃圾邮件可以随机更改一些单词的拼写,比如("强悍","弓虽悍","强-悍"),传统的技术就很难识别。针对这些情况,IISF 将自

*)基金项目:国家自然科学基金(the National Natural Science Foundation of China under Grant No. 60603026)。王 磊 教授,博士生导师,研究方向:人工免疫理论及其应用技术、计算机网络信息安全系统等;董 菡 硕士研究生,研究方向:智能决策支持系统。

体耐受、免疫识别、免疫学习、免疫记忆、协同刺激等机制应用到垃圾邮件过滤器中，有效地识别、拦截新的垃圾邮件。

1.2 自体、非自体、抗原、抗体的定义

生物免疫系统的自体和非自体是形态不同的蛋白质链。在 IISF 中，S 为自体，是正常的邮件特征向量；T 为非自体，是垃圾邮件的特征向量。抗原 ag 是所有邮件的特征向量，免疫细胞(即抗体)ab 是垃圾邮件的特征向量。简单起见，本文中的免疫细胞融合了 B 细胞、T 细胞和抗体的性质，用于检测和识别垃圾邮件。

设 $Ag=\{ag\}$，$Ab=\{ab\}$，β 为成熟细胞激活阈值，定义成熟细胞集合 $Tb=\{x|x\in Ab, x.\text{count}<\beta\}$，记忆细胞集合 $Mb=\{x|x\in Ab, x.\text{count}>=\beta\}$，$Ab=Mb\cup Tb$。

定义未成熟免疫细胞集合 Ib，$|Tb|+|Ib|=\delta$，其中 δ 为常数，N 为自然数，$|Tb|$ 表示成熟细胞数量，$|Ib|$ 表示未成熟细胞数量。

1.3 抗体的产生

初始抗体特征库通过提取已明确定义为垃圾邮件的训练样本的特征字段来建立。在 IISF 中，不能把正常的邮件识别成垃圾邮件，所以免疫细胞必须要经历一个耐受的过程。如果在耐受期内与自体发生匹配，就会死亡并被新的免疫细胞代替。免疫细胞经过耐受期成为成熟免疫细胞。此后，成熟免疫细胞若遇抗原产生匹配，且匹配次数如果超过 β，则被激活转变为记忆细胞。如成熟免疫细胞在其生命周期内未能累积足够的亲和力，则走向死亡，并被新的成熟免疫细胞代替。这样可以确保免疫细胞的多样性，保证了其对抗原空间的持续搜索能力，并能保留那些最好的免疫细胞[1]。对于记忆细胞，在再次匹配抗原后就会被再次激活并克隆自己，克隆生成的新细胞加入成熟细胞集合。一部分符合条件的免疫细胞可以进行变异，变异产生的细胞是未成熟的免疫细胞，需要耐受才能成熟。

1.4 智能决策支持系统

在决策过程中运用人工免疫思想，模拟生物免疫过程，在决策阶段的全过程为反垃圾邮件的判断提供有力的支撑，并且运用免疫学习和克隆选择变异对知识库的信息更新以及相似信息的模糊识别提供了更加有效的实现手段。智能垃圾邮件过滤器的总体结构如图 1 所示。其中垃圾邮件特征库模拟抗原特征库，邮件特征库模拟抗原。

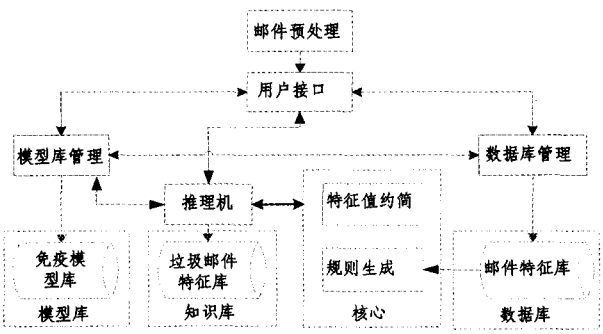

图 1 IISF 结构图

在邮件特征库中获得疑似垃圾邮件，通过对其特征进行分析，匹配，约简，生成规则，更新垃圾邮件特征库即抗体特征库，对抗体特征库的疑似特征样本经过细胞克隆变异，生成为成熟细胞集，对经过自体耐受的特征细胞放入成熟细胞集。

1.5 IISF 模型实现

人体对病原体的防御包括能抵抗和消灭入侵病原体的两道防线，第一道防线主要指皮肤和粘膜，第二道防线指吞噬细胞或巨噬细胞。IISF 模拟这种机制，它包含了一个可信任地址列表(经过用户授权的邮件地址)和一个黑色地址列表(臭名昭著的垃圾邮件地址)，地址属于可信任邮件地址列表中的邮件被认为是正常邮件，地址属于黑色地址列表中的邮件被认为是垃圾邮件，其它邮件均被作为可疑邮件进行进一步的识别。IISF 模型如图 2 所示。

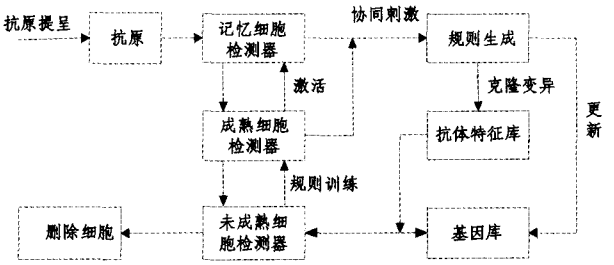

图 2 IISF 免疫模型

1.5.1 抗原提呈

抗原提呈是将邮件文本转化为可供免疫细胞识别的抗原格式，是整个系统进行的前提。信件由信头、信件体和它们之间的空行组成。TCP/IP 报文格式标准规定了邮件头的格式与语义，只需提取所有词语并消除重复关键词，将其放入抗原的特征向量即可。

表 1 邮件头关键字

关键字	解释
From	发件人的邮件地址
To	收件人的邮件地址
Subject	邮件主题
Reply-To	发件人期望收件人回复的地址
Message-ID	邮件唯一标志，该字段由 MUA 或者第一个 MTA 产生
Received	信件的 MTA 处理纪录，用于跟踪信件的路径和处理过程

而对于信体，由于其长度可能较大，如果将所有关键词都放入特征向量，可能造成基因库过于臃肿，并且识别操作和种群更新操作的复杂度过大。我们引入粗糙集理论，对关键词进行知识约简，即在保持知识库分类能力不变的条件下，删除其中的不相关或不重要的知识。

对于决策系统，即信体关键字系统 $S=(U,R,Y,f)$，$R=C\cup D$，其中 $C=\{a_i|i=1,2,\cdots,m\}$ 是条件属性，$D=\{d\}$ 是决策属性，根据约简定义，令 $P=\{C-\{a_i\}\}$ 和 Q 为等价关系族，Q 的 P 正域记为 $\text{POS}_P(Q)=\bigcup_{X\in U/Q}\underline{P}X$ 如果 $\text{POS}_{\text{IND}(P)}(\text{IND}(Q))=\text{POS}_{\text{ind}(P)-\{a_i\}}(\text{IND}(Q))$，则称 a_i 为 P 中 Q 不必要的，否则 a_i 为 P 中 Q 必要的。

算法描述如下：

```
Begin
    For i=1 To m
    Do
    {
    If (C-{a_i},POS_(C-{a_i}))(Q)=POS_C(Q)) Then
      {C'=C-{a_i}};/*删除 a_i 所在的列且合并 U 中重复的行;
*/}
    Else C'=C
    }
```

将删除后的剩余关键词组成新的向量放入到抗原特征向量中。这样既保证了选择的特征向量对邮件内容的代表意

义，同时又限制了特征向量长度的过度膨胀。

1.5.2 免疫识别

免疫识别是免疫系统的主要功能，识别的本质是区分“自我”和“非我”。免疫识别过程就是抗体对抗原的识别过程，抗体对抗原的识别是通过结合（或匹配）过程实现的，相应的人工免疫系统中的抗原识别通过特征匹配来实现，其核心是定义一个匹配阈值，对匹配的度量可以采用多种方法，如 Hamming 距离、R 连续位匹配方法等。

IISF 在免疫识别阶段，等待新邮件的到来并将其分类为正常邮件或垃圾邮件。其中记忆细胞检测流程如图 3 所示。成熟细胞的情况类似于记忆细胞 。

由免疫细胞的识别流程可以看出，免疫识别过程同时是一个系统学习的过程，学习的结果是免疫细胞的个体亲和度提高，群体规模扩大，并且最优个体以免疫记忆的形式得到保存。

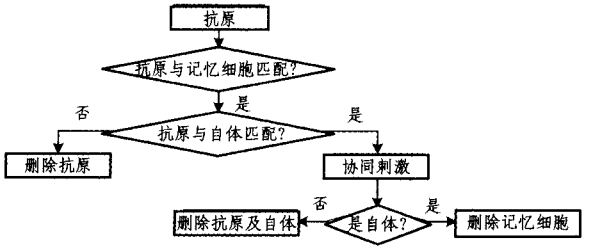

图 3 记忆细胞检测流程图

1.5.3 规则生成

用户的反馈模拟免疫模型的协同刺激。将用户反馈结果作为规则决策表的决策属性；抗原特征作为条件属性，生成决策表，通过特征约简，有效地分析和处理各种不完备信息，从而导出垃圾邮件的系统决策判断，并从中发现隐含的知识，揭示潜在的规律。这里协同刺激的主要功能包括消除识别正常邮件的免疫细胞及抗体库中对应的关键词；增加成功识别垃圾邮件的免疫细胞的浓度；提取新确认的垃圾邮件的特征关键字。

特征约简采用经典的区分矩阵方法，得到属性核值。将核属性加入到区分函数中进行合取运算；输出结果：析取范式的每个合取项就对应决策表的一个约简。$\mathrm{COR}_D(C)$ 中包含的即为决策表的核。

根据决策表生成规则并计算核心特征的权重，其中权重表达了当前邮件特征值对垃圾邮件判断决策的影响大小。在不同的环境下，相同的特征对决策输出会有不同的影响，即权重对环境的敏感性。在不同特征分类中的每个特征对于决策结果的重要度也不尽相同，重要程度定义为：

$$\mathrm{card}(U)k_i=(\mathrm{card}(U)-\mathrm{card}(\mathrm{pos}_{(c/c_i)}(D)))/\mathrm{card}(U)$$

其中 $\mathrm{card}(U)$ 表示集合个数；k_i 越大则条件属性 c_i 即第 i 个指标对评标决策越重要，得到重要度后我们进行权重计算 $w_i=k_i/\sum_{i=1}^{n}k_i$。获得的核以及对应权重形成新的规则，用其更新抗体规则库，新抗原的产生可根据权重大小优先生成。

1.5.4 免疫学习

免疫学习大致可以分为两种：一种发生在初次应答阶段，即免疫系统首次识别一种新的抗原时，其应答时间相对较长；而当机体重复遇到同一抗原时，由于免疫记忆机制的作用，免疫系统对该抗原的应答速度大大提高，这个过程是一个增强式学习过程，对应于再次应答。免疫系统强大的学习能力使之成为一个随环境改变而不断完善的自适应系统，IISF 的免疫学习主要应用在以下方面：

(1)抗体库的动态更新。IISF 在检测垃圾邮件的过程中不断学习并提取新的垃圾邮件特征关键字，使抗体库具有动态性和实时性。

(2)对用户反馈行为的学习。系统可以学习某段时间内用户对疑似邮件的判断习惯，从而提高识别效率。这样系统能更清晰地反映实际抗原环境。

1.5.5 克隆选择与变异

垃圾邮件为了躲避过滤器对某些敏感词组的拦截，常常采用相似的词组或改变文法排序冒充成正常邮件，免疫学中对于这种伪装的行为理解为克隆选择与变异。

克隆选择理论描述了获得性免疫的基本特性，其基本思想是只有那些能识别抗原的细胞才得以增殖。IISF 中，邮件在被免疫细胞分类为垃圾邮件后如果得到协同刺激的肯定，说明该细胞与抗原具有较高的亲和力，它就会繁殖，产生大量的抗体。这种抗体的产生是一个学习过程，也是一个优化过程。

免疫细胞在克隆时要经历变异，IISF 将变异后产生的新细胞作为未成熟免疫细胞，这样利于短期内使其达到成熟，而且使系统具有更好的实时性。

1.5.6 规则训练

免疫细胞不仅要尽可能多地识别抗原，还必须保证不能错误地将正常邮件产生的特征向量误认作抗原（错误肯定）。因此，新生成的免疫细胞必须经过自体耐受方能成熟，若在这个过程中与任意自体字符串匹配，就将死亡。

1.5.7 免疫细胞的动态演化

首先，垃圾邮件的定义范畴是变化的，抗体库必须进行动态更新，删除对于那些已经不属于垃圾邮件范畴的特征向量；其次，免疫系统要用有限的抗体识别无限的抗原，这就需要不断补充新的免疫细胞，随着时间的推移，记忆细胞会无限扩增，因此在 IISF 中引入细胞的动态演化机制。免疫细胞的动态演化可以控制抗体的规模，确保免疫细胞的多样性，淘汰劣势细胞并保留那些最好的免疫细胞[1]。其算法如下所示：

```
Begin
For each ab∈Ib
{
/* 未成熟细胞的动态演化 */
  If (ab.age>a) then
    {从 Ib 中删除该细胞；将细胞加入到 Tb 中；}
  If (|Tb|+|Ib|<δ) then
    {补充新的未成熟细胞；}
  };
For each ab∈Tb
{
If (ab.count>β) Then
/* 成熟细胞的动态演化 */
  {从 Tb 中删除该细胞；将细胞加入到 Mb 中；ab.count=0；}
If (ab.age>λ) Then /* λ 为常数，成熟细胞死亡年龄 */
  {删除该细胞；补充新的成熟细胞；}
}
If (记忆细胞数量>ξ) Then /* 记忆细胞的动态演化，ξ 是系统允许的记忆细胞最大数 */
  {按 LRU 算法淘汰掉最近最少使用的记忆细胞，将其放入 Tb；}
End
```

2 仿真实验

使用 CCERT 中文邮件样本集 2005-Jul 的子集对 IISF 的识别能力进行测试，每次测试中垃圾邮件占 69.21%，正常邮件占 30.79%。

本文采用正确率、精确率和虚报率几个指标来评价系统，其中正确率＝准确判断出的垃圾邮件数/判断出的垃圾邮件总数，精确率＝准确判断出的垃圾邮件数/垃圾邮件总数，虚报率＝误判为垃圾邮件数/正常邮件总数。过滤指标曲线图如图 4 所示。

（下转第 65 页）

4 实例应用

本文使用UCI数据库提供的网络入侵测试数据集进行实验分析。该数据集中包括正常数据和两种异常数据，每种数据包括网络数据包的包头信息、网络连接和传输信息等37个属性，将数据集分为训练集和测试集两部分。

(1)首先对原始数据进行预处理，将原始数据中的噪声去掉并转换为神经网络可处理的特征向量。原始数据属性包括：UKEY，ID，diff_source_hosts，dst_bytes，duration，serror_rate，SYN等37个属性，限于篇幅，此处不一一列出。使用粗糙集方法将原始数据中多余的和不相关的信息去掉，删除与本文采用的入侵检测方法无关的属性。

(2)将神经网络无法处理的符号字段转换成数值字段。将部分属性重新编码，如将ACK，PSH，FIN，SYN，URG属性中的值"NULL"用"0"替换。destination_host和source_host属性表示IP地址，分别为其增加一个新的属性，表示IP地址中前两段的网络号信息，内部网络IP地址用"0"值，外部网络IP地址用"1"值。

(3)对各属性值进行归一化处理，减少由于记录间字段数值差异过大而对网络训练产生的不良影响。将dst_bytes，duration，destination_port，source_port属性的值除以10得到0～6.5535之间的数值。

(4)将经过数据预处理的条件属性用改进后的神经网络系统进行训练，并在测试集中进行入侵检测试验。为检验模型的泛化性能，我们将训练集分为两组，第一组在训练集中选取40条记录作为训练样本数据，每一类数据各20条；第二组在训练集中选取100条记录作为训练样本数据，每一类各50条。测试时，在测试集中各选取50条正常数据、50条问题数据。限于篇幅，训练及测试过程略。

(5)将结果与训练集进行比对。鉴于有效降低入侵检测的误报率及漏报率是入侵检测的公认难题，我们用误报率作为标准进行比较。两种算法在两组不同数据集的检测结果(误报率)如表1所示。

表1 两种系统进行入侵检测的误报率

误报率(%)	传统神经网络系统	改进后的神经网络系统
第一组	23.6	19.1
第二组	36.2	25.5

由此可以看出，使用相同训练样本训练而得到的网络模型进行测试，改进后的系统入侵检测识别能力明显好于传统粗糙神经网络。尤其是经过不良信息过滤隐单元的过滤，在入侵检测前期就去掉了大量的明显的网络入侵信息，简化了神经网络的输入维度，从而使整个入侵检测系统的误报率显著降低。

结束语 互联网中的入侵、黑客行为非常繁杂，有的显而易见，有的则需要专门的入侵检测系统进行专门检测。本文对传统的基于粗糙集的神经网络系统进行了改进，将Web信息过滤技术加入入侵检测系统，先期过滤掉能够明显识别的入侵信息，以减轻神经网络系统的运行负荷，提高了入侵检测系统的适用性。

参考文献

1 唐正军. 网络入侵检测系统的设计与实现. 北京：电子工业出版社，2002

2 Lippmann R P, Cunningham R K. Improving Intrusion Detection Performance Using Keyword Selection and Neural Networks. Computer Networks—the International Journal of Computer and Telecommunications Networking, 2000, 34 (4) : 597～603

3 蒋建春，马恒太，任党恩，等. 网络安全入侵检测：研究综述. 软件学报，2000(11)：1460～1466

4 单征. 基于网络状态的入侵检测模型. 信息工程大学学报，2002(3)：9～14

5 周志华，曹存根. 神经网络及其应用. 北京：清华大学出版社，2004

(上接第62页)

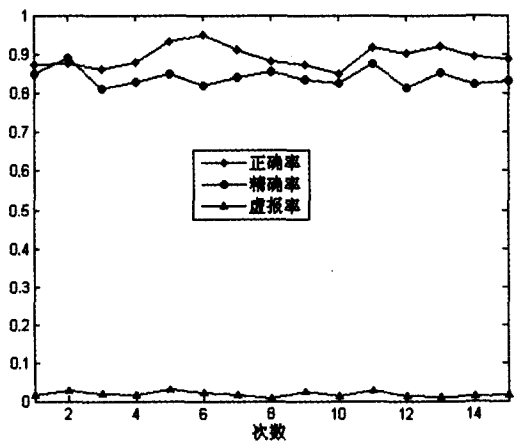

图4 过滤性能指标曲线图

从图4可以看出，正确率和精确率都比较稳定，说明系统的识别能力良好，虚报率较低且值的变化较小，说明系统将正常邮件识别为垃圾邮件的概率较低，即使在这种情况下，由于具有协同刺激机制，邮件也不会被误删，说明系统具有较高的可靠性。此外，影响结果的因素有很多，例如成熟细胞的激活阈值、可信任邮件数目等，部分参数是相互作用的，应保持合适的比例。

小结 本文提出了一种基于免疫机制的智能反垃圾邮件过滤器，主要完成邮件分类、抗体库更新和对用户反馈的协同认证等。该过滤器可以识别垃圾邮件的特征变化并学习、记忆新的垃圾邮件特征，从新的垃圾邮件内提取特征向量，学习用户的行为习惯。系统首先使用可信任邮件列表和黑地址列表对邮件进行初次筛选，其余邮件则由基于免疫的过滤器进一步审查，这种双层机制减少了对一些特征明显的邮件的审核，在提高判别效率的同时进一步增强了系统的可靠性。性能测试表明，IISF对垃圾邮件具有良好的识别能力，并具备一定的自学习和自适应性。

参考文献

1 李涛. 基于免疫的网络监控模型[J]. 计算机学报，2006，29(9)：1515～1522

2 肖人彬，王磊. 人工免疫系统：原理、模型、分析及展望. 计算机学报[J]，2002，25(12)：1281～1293

3 Dasgupta D, Attoh-Okine N. Immunity-based systems: A survey. In: Proc. IEEE International Conference on systems[C], Man and Cybernetics, 1997. 369～374

4 李洋，方滨兴，王申. 基于用户反馈的反垃圾邮件技术[J]. 计算机工程，2007，33(8)：130～132

5 Perone M. An Overview of Spam Blocking Techniques[R]. Barracuda Networks Corp, 2004

6 张修文，吴伟志. 粗糙集理论与方法[M]. 北京：科学出版社，2001. 12～39