第 36 卷 第 1 期 电 子 科 技 大 学 学 报 Vol.36 No.1
2007 年 2 月 Journal of University of Electronic Science and Technology of China Feb. 2007

基于人工免疫原理的反垃圾邮件系统AIASS

张成功，黄迪明，胡德昆

(电子科技大学计算机科学与工程学院 成都 610054)

【摘要】讨论了基于人工免疫原理的反垃圾邮件系统AIASS的生物免疫基础以及各个关键技术环节，对系统环境进行了定量和定性分析，并采用SpamAssassin垃圾邮件库对系统性能进行测试。实验结果表明引入人工免疫原理后，反垃圾邮件系统具备了优于传统方法的自适应能力和稳定性。

关 键 词 垃圾邮件； 人工免疫； 贝叶斯模型； 计算智能
中图分类号 TP301 文献标识码 A

An Anti-Spam System AIASS Based on Artificial Immune Principle

ZHANG Cheng-gong，HUANG Di-ming，HU De-kun

(School of Computer Science and Engineering, Univ.of Electron. Sci. & Tech. of China Chengdu 610054)

Abstract Issued the biologic foundation and various key processes of AIASS, which is an anti-spam system based on artificial immune principle, the system environment is analyzed qualitatively quantitatively. Also a spam library named SpamAssassin to test the performance of system is used. The result of experiment shows that the anti-spam system possesses stronger adaptability and stability than traditional method after introducing artificial immune principle.

Key words spam; artificial immunology; Bayes model; computational intelligence

对计算机网络而言，垃圾邮件作为一种典型的非法信息，其危害之大、影响之广，令人触目惊心。据统计，全球平均每天发送垃圾邮件7.3亿封[1]，按每封造成经济损失1～2美元计算[2]，每年因垃圾邮件而造成的经济损失约为2 700亿美元。因此，如何有效识别和及时拦截垃圾邮件成为当今网络安全技术领域的研究热点。传统的反垃圾邮件(Anti-spam)技术大多采用基于统计学习原理的Bayes方法。在Bayes模型下，从邮件文本提取特征向量 $\boldsymbol{s}=\{s_1,s_2,\cdots,s_n\}$，并假设分量之间保持统计独立。通过训练样本的特征向量各分量出现的频度来计算各分量在各类别邮件中出现的概率，从而计算最大概率类别 ϖ_{NB}，有：

$$\varpi_{\mathrm{NB}}=\arg\max_{\varpi_j\in\varpi}P(\varpi_j)\prod_i P(s_i\mid\varpi_j) \tag{1}$$

式中 $\varpi=\{\varpi_1,\varpi_2\}=\{\mathrm{spam},\mathrm{nonspam}\}$。

Bayes模型充分发掘了已知样本所包含的统计信息，因此其可信度得到了保证。但是对于未知样本，特别是已有样本的变异样本的分类问题，基于Bayes模型的分类系统的执行效果往往不尽人意。例如，“ cash ”是垃圾邮件中出现频率较高的词语，采用传统Bayes模型对包含“ cash”的邮件进行分类，效果比较理想。但是当垃圾邮件发送者为躲避反垃圾邮件系统的拦截而将该词换作“ ca$h ”时，系统的有效性就很难得到保证。因此，如同对计算机病毒所采取的防范措施一样，对于垃圾邮件这种隐蔽性强、变异能力突出的网络非法信息，必须采用具有自适应和自学习能力的分类系统。而生物免疫原理则恰好为此提供了丰富的启示。

1 AIASS的生物免疫机理及功能模拟

免疫系统是生物体抵御外来入侵的有力屏障。在免疫系统中，任何能被识别的外来细胞都称为抗原(Ag)，生物体通过B细胞产生抗体(Ab)响应外来抗原的刺激。抗体在正常状态下是附着在B细胞表面的分子群。一旦有某类抗原入侵，那些对该类抗原有特异性的B细胞立即接受指令，令其抗体分化繁殖。文献[3-6]的克隆选择原理指出，这样的繁殖是带有遗传变异色彩的，在不断分化的过程中，对某种抗原具有特异性的抗体的分子结构逐渐朝着更具亲和

收稿日期：2005 - 03 - 15
基金项目：四川省科技厅科技攻关重点资助项目(03GG-066-021)
作者简介：张成功(1981 –)，男，博士生，主要从事生物智能计算、网络信息技术方面的研究.

力和泛化的方向进化，这一过程称为“ 亲和力成熟”。当抗体的性能经过进化达到一定的亲和力阈值后，抗体即被收入记忆抗体库中成为记忆抗体。因此，以后再次受到相同抗原或者已知抗原的变异体刺激时，系统能利用这些经过进化而具有更高识别效能的记忆抗体更有效地产生免疫应答[7]。此外，B细胞必须在受到T Helper cell的协同刺激(Co-stimulation)的前提下才能被抗原激活，这在一定程度上避免了免疫系统的自体反应。

本文讨论的反垃圾邮件系统(Artificial Immune Anti-Spam System，AIASS)中，抗原即邮件样本的特征向量；抗体即由系统基因库中随机提取的基因片段组成的检测器；基因库即一个垃圾邮件特征字段的集合，由训练样本生成，并且在系统的工作周期中随着新的垃圾邮件被识别以及新记忆抗体的生成而得到持续的扩充。系统中的其他功能实体与生物免疫系统的对照如表1所示。相关的生物免疫学术语见参考文献[8]。

表**1**　**AIASS**与生物免疫系统的功能实体对照

AIASS	特征向量	检测器	邮件分类	人工确认	检测器复制	检测器竞争
生物免疫	抗原	抗体	识别抗原	T-Helper cell	克隆选择	记忆抗体生成

2　AIASS的系统环境及关键技术环节

AIASS的邮件识别流程采用基于Burnet克隆选择原理[3]的免疫算法。由于T helper cell的协同刺激对产生自体反应的B细胞具有抑制作用，因此在AIASS中并未引入阴性选择过程[9]。实质上，T helper cell的协同刺激在算法的后期起到了与阴性选择等同的作用。实验结果也显示，排除阴性选择过程对抗体的错误否定率没有影响，对于错误肯定率的影响也非常有限，而换来的是系统整体效率的提升。

2.1　基因库和抗体结构

AIASS中，基因库以文本文件Gene_Library. txt的形式存在。文件中每行的垃圾邮件关键字段代表一个基因片段。AIASS基因库通过对垃圾邮件训练样本的关键字提取进行初始化，在系统的运行过程中一旦拦截到新的垃圾邮件并得到协同刺激的确认，则提取该垃圾邮件的特征字段并对基因库进行扩充。可以预计，在基因库模块中使用基于时间或者调用频率的动态更新策略，将使基因片段的使用效率更高。但目前尚未引入类似的机制，在今后的研究中期望结合数据库原理对基因片段的动态调节机制开展进一步的探索。系统通过从基因库中随机抽取基因片段生成可变长度的抗体，抗体与抗原的亲和力通过下式计算：

$$\begin{aligned}&\text{affinity}(A_{\text{Ab}}, A_{\text{Ag}}) = \\ &h(A_{\text{Ab}}, A_{\text{Ag}}) / \min(\text{length}(A_{\text{Ab}}), \text{length}(A_{\text{Ag}}))\end{aligned} \tag{2}$$

式中　分子代表抗体与抗原匹配的基因片段数目；分母代表抗体和抗原长度值的最小者。

2.2　抗体生命周期

抗体的生命周期包含训练、应用、种群更新三个部分，如图1所示。

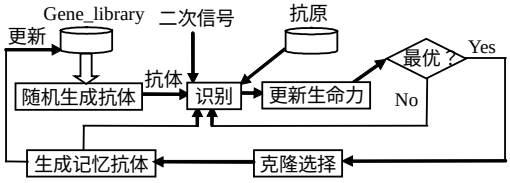

图1　抗体生命周期的各阶段

2.2.1　训练

免疫细胞训练阶段的主要任务是通过提取已明确定义为垃圾邮件的训练样本的特征字段，生成初始基因库，并在此基础上生成初始抗体和初始记忆抗体种群。

2.2.2　识别

在系统的识别阶段，抗体种群中的所有抗体与未知类别的抗原进行亲和力计算，并据此识别该类别的抗原。分别定义抗体识别函数 $R_i(A_{\text{Ag}})$ 和系统识别函数 $R(t, A_{\text{Ag}})$ 如下：

$$R_i(A_{\text{Ag}}) \in \{0,1\} \ ; \ R(t, A_{\text{Ag}}) \in \{0,1\}$$

$$R_i(A_{\text{Ag}}) = 1 \text{ iff } \text{affinity}(A_{\text{Ab}_i}, A_{\text{Ag}}) > \delta \tag{3}$$

$$R(t, A_{\text{Ag}}) = 1 \text{ iff } \exists i, R_i(A_{\text{Ag}}) = 1 \tag{4}$$

式中　$\text{affinity}(A_{\text{Ab}_i}, A_{\text{Ag}})$ 为 t 时刻抗体 A_{Ab_i} 对抗原 A_{Ag} 的亲和力；δ 称为亲和力阈值。

2.2.3　克隆选择过程与变异

在系统的识别阶段，对于每一封邮件(抗原)，抗体种群(包括记忆抗体种群)中的大量抗体对其表现出有差异的亲和力。因此按照克隆选择原理，选择对抗原具有最大亲和力的抗体进入克隆过程。抗体按照与亲和力大小成正比的数目克隆自身；随后又按照与亲和力大小成反比的概率对这些克隆体进行随机变异，即：

$$\text{Clonal\_number}(A_{\text{Ab}_i}, A_{\text{Ag}}) = \left\lfloor K_{\text{c}} \ \text{affinity}(A_{\text{Ab}_i}, A_{\text{Ag}}) \right\rfloor$$

$$\begin{aligned}&\text{Mutate\_possibility}(A_{\text{Ab}_i}, A_{\text{Ag}}) = \\ &\left\lfloor P_{\text{m}}(1 - \text{affinity}(A_{\text{Ab}_i}, A_{\text{Ag}})) \right\rfloor\end{aligned}$$

式中　K_{c}、P_{m} 分别为克隆控制因子和变异概率控制因子。另外，规定AIASS中的抗体变异属于多点

等概率变异。因此，抗体变异的基因位个数服从Binomial分布：

$$P\{\text{Mutate\_number}(A_{\text{Ab}_i}, A_{\text{Ag}}) = v\} = C_L^v p^v (1-p)^{L-v} \quad (7)$$

且p=Mutate-possibility $(A_{\text{Ab}_i}, A_{\text{Ag}})$，$L$=length(A_{Ab_i})。

注意到在AIASS中 $L >> 1$，因此由Demoivre-Laplace定理有：

$$P\left\{\text{Mutate\_number}(A_{\text{Ab}_i}, A_{\text{Ag}}) \le v\right\} \approx \int_{-\infty}^{v} \frac{1}{\sqrt{2\pi Lp(1-p)}} \exp[-\frac{(x-Lp)^2}{2p(1-p)}]\mathrm{d}x \quad (8)$$

因此，抗体的平均变异基因位个数满足：

$$E[\text{Mutate\_number}(A_{\text{Ab}_i}, A_{\text{Ag}})] = Lp \quad (9)$$

由式(2)、(6)和(9)可知：

$$0 \le E[\text{Mutate\_number}(A_{\text{Ab}_i}, A_{\text{Ag}})] \le P_m L \quad (10)$$

由此可以得出结论，即变异概率控制因子 P_{m} 可视为抗体基因位中变异位所占的最大比例。较小的 P_{m} 值避免了当抗体种群的平均识别性能较弱时出现的种群结构的剧烈波动，因此，该参数对于系统的稳定性起着至关重要的作用。

变异后的抗体种群和原始抗体种群中的最优者被选入记忆抗体种群。记忆抗体的生存周期远长于一般抗体。

2.2.4 T helper cell协同刺激

由于免疫细胞是从基因库中随机生成，并且缺少阴性选择过程，因此不能保证抗体在识别阶段不将无害抗原(非垃圾邮件)识别为有害抗原(垃圾邮件)。对这一潜在的缺陷，通过引入类似于生物免疫系统中T Helper cell协同刺激的原理进行弥补。当AIASS识别出垃圾邮件时，将其放入“ 待确认区域 ”，等待系统管理员的最后确认(二次信号)。一旦二次信号表明该邮件为正常邮件时，删除所有识别该邮件的抗体；反之则增强抗体的生命力。

2.2.5 免疫细胞动态更新

在现实环境下，为了保持系统的高效和动态性，抗体种群必须引入动态更新机制。这一机制的实施基于抗体“ 生命力 ”的概念。在抗体种群的每一代(定义为系统在识别阶段每次遭遇抗原的时间段)，若抗体A_{Ab_i} 对抗原A_{Ag} 的识别得到协同刺激(二次信号)的确认，则其生命力 l_i 按下式变化：

$$\Delta l_i = k \cdot \text{affinity}(A_{\text{Ab}_i}, A_{\text{Ag}}) - h \cdot \text{rn}(A_{\text{Ag}}) - C \quad (11)$$

$$\text{rn}(A_{\text{Ag}}) = R_i(A_{\text{Ag}}) / \left|A_{\text{Ab}}\right| \quad (12)$$

式中 k、h和C 为常数，C 为自然衰亡速率；若抗体识别无害抗原，则意味着发生自体反应，应消除该抗体以降低错误肯定率，即令 $l_i = 0$。同时需一并删除基因库中该抗体对应的基因片段。一般抗体和记忆抗体的初始生命力分别为 I_{N}、I_{M}，且满足 $I_{\text{M}} >> I_{\text{N}}$。在样本更新的最后阶段，系统随机生成新的抗体，其目的是：(1) 弥补因自体识别和生命力耗尽导致的抗体数量减少；(2) 增强抗体种群的识别范围，提高系统对未知抗原的识别能力。

3 样本测试

采用SpamAssassin垃圾邮件样本的一个子集对AIASS的识别能力进行测试。训练样本由42封垃圾邮件组成；测试样本的最大数量为2 166封，其中垃圾邮件数量为705(32.5%)，正常邮件数量为1 461(67.5%)。在系统运行期间，所有参数均为常量，如表2所示。

表2 AIASS参数设置

δ	K_c	P_{m}	I_{N}	I_{M}	k	h	C
0.52	7.44	0.65	20	100	3.2	1.9	1.0

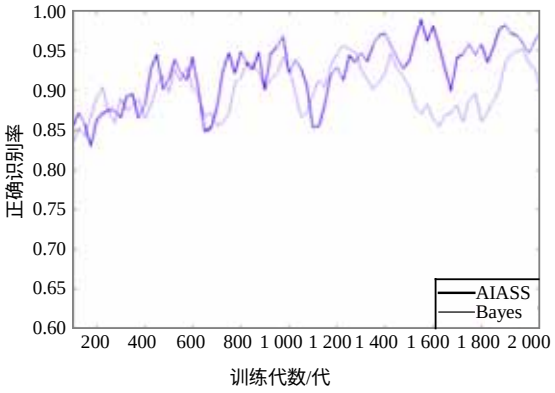

图2 AIASS与Bayes的识别效果比较

4 结 束 语

本文概括了基于人工免疫原理的反垃圾邮件系统AIASS的关键技术，深入讨论了免疫系统的各个关键技术环节，并进行了定量分析。实验结果表明，在反垃圾邮件系统中采用人工免疫原理，使系统具备了更强的自适应能力和稳定性，这对于开发针对变异能力极强的垃圾邮件进行过滤的网络安全系统具有较强的启发意义。在下一步的研究中，计划在抗体种群中引入相互激励和抑制机制，结合免疫网络的相关理论，提高AIASS对未知样本的快速反应能力和抗欺骗能力。

参 考 文 献

[1] ODA T, WHITE T. Increasing the accuracy of a spam-detecting artificial immune system[J]. Evolutionary

Computation, 2003, 1 (1): 390-396.
[2] ATKINS S. Size and cost of the problem[C]// In Proceedings of the Fifty-sixth Internet Engineering Task Force(IETF) Meeting, SpamCon Foundation, San Francisco, 2003.
[3] BURNET F M. The clonal selection theory of acquired immunity[M]. London: Cambridge University Press, 1959.
[4] 竹小明, 许家玿. 动态克隆选择和免疫网络结合的算法[J]. 实验科学与技术, 2006, 4(5):35-37.
[5] 梁宏志, 许家玿. 免疫在入侵检测中的应用基础抗原编码[J]. 实验科学与技术, 2006, 4(6):34-36.
[6] 赖　立, 许家玿. 利用聚类法建立免疫模型自我库[J]. 实验科学与技术, 2006, 4(4): 8-10.
[7] LEANDRO Nunes de Castro, FERNANDO J Von Zuben. Learning and optimization using clonal selection principle[J]. IEEE Transactions on Evolutionary Computation, Special Issue on Artificial Immune Systems, 2001, 6 (3): 239-251.
[8] 陈　仁. 免疫学基础[M]. 北京: 人民卫生出版社, 1982.
[9] FORREST S, PERELSON A S, Allen L, et al. Self-nonself discrimination in a computer[C]// In Proceedings of the IEEE Symposium on Research in Security and Privacy, IEEE Computer Society Press, Oakland, 1994.

编　辑　熊思亮

(上接第95页)

网络传输中存在许多不确定因素，为了在客户端得到实时稳定的监控图像，避免网络抖动对视频播放的影响，本文设计的监控系统在网络接收模块和DirectShow模块的源过滤器之间采用双缓冲技术，即建立两个队列：(1) 空闲缓冲区队列，用于接受数据；(2) 尚未处理的数据缓冲区队列，等待源过滤器读取。双缓冲技术将网络抖动带来的影响减小到最低。

6) 应用程序接口

DirectShow模块是应用程序的直接下层模块，监控系统客户端软件具备功能强大、操作简单、界面漂亮等特点，DirectShow提供了相应的接口，能很好地满足了以上要求：

(1) GraphBulder 接口：建立过滤器图表；

(2) MediaEventEx 接口：获得播放过程中发生的事件，如播放完毕等。主要方法有 SetNotifyWindow，指定处理事件的窗口；GetEvent，获得事件；

(3) IVideoWindow 接口：控制视频窗口属性。主要方法有 put_Owner，指定视频窗口的父窗口；put_FullScreenMode, 指定全屏播放模式；put_MessageDrain，指定一个窗口，用于接收视频窗口发出的鼠标消息等；

(4) IMediaControl 接口：控制过滤器图表的运行。主要方法有 Run，开始运行；Pause，暂停运行；Stop，停止运行。

3　结 束 语

基于COM组件的DirectShow技术使多媒体应用开发模块化，并且DirectShow提供的一系列功能强大的基类极大地简化了开发过程，同时画面的质量和实时性也得到了保证，取得了满意的视觉效果，为MPEG-4的类似应用提供了一个完整的方案。本文设计的监控系统在主动丢包测试时，当丢包率达到10%，画面会出现少量的马赛克，播放依然保持流畅，完全满足远程视频监控的要求。

参　考　文　献

[1] 钟玉琢.基于对象的多媒体数据压缩编码国际标准MPEG-4及其校验模型[M]. 北京: 科学出版社, 2000.
[2] COMER D E. 用TCP/IP进行网际互联[M]. 林　瑶, 蒋慧 译. 北京: 电子工业出版社, 2003.
[3] 路其明. DirectShow开发指南[M]. 北京: 清华大学出版社, 2003.
[4] Audio- Video Transport Working Group, RFC1889-RTP: A transport protocol for real-time application[S]. 1996.
[5] Microsoft Corporation. DirectShow online document ation [DB/OL]. http://www. msdn.microsoft.com/directx/., 2005-03-25.
[6] 吴　萍, 傅　彦. MPEG-4编码和流式数据传输[J]. 实验科学与技术, 2003, 1(3): 79-92.

编　辑　黄　莘