

具有疫苗算子的可变模糊匹配阴性选择算法

王 辉¹, 于立君¹, 毕晓君², 张利军¹

(1. 哈尔滨工程大学 自动化学院, 150001 哈尔滨, whjyylj@163.com;

2. 哈尔滨工程大学 信息与通信工程学院, 150001 哈尔滨)

摘 要: 为了建立可变模糊匹配阴性选择免疫算法的特异性免疫应答, 基于模糊思想并采用疫苗理论, 在监测异常过程中提取疫苗, 研究疫苗算子和正选择算子对算法性能的影响. 仿真结果表明: 疫苗算子增强了抗体库的记忆功能, 使算法二次应答时间明显缩短, 而正选择算子是完成这种功能的具体实施者. 可变模糊匹配阴性选择免疫算法经改进设计后, 具有较强的自学习功能, 对建立特异性免疫应答具有自适应性.

关键词: 模糊匹配; 免疫应答; 疫苗算子; 正选择算子; 自适应性

中图分类号: TU375.4

文献标志码: A

文章编号: 0367-6234(2011)06-0141-04

Adjustable fuzzy matching negative selection algorithm with vaccine operator

WANG Hui¹, YU Li-jun¹, BI Xiao-jun², ZHANG Li-jun¹

(1. Automation College, Harbin Engineering University, 150001 Harbin, China, whjyylj@163.com; 2. Information and Communication Engineering College, Harbin Engineering University, 150001 Harbin, China)

Abstract: To set up inherent immune response of the adjustable fuzzy matching negative selection immune algorithm, based on fuzzy idea, the vaccine theory is used to pick up the vaccine during abnormal monitoring, and the algorithm performance influenced by vaccine operator and positive selection operator is studied. The simulation results show that the antibody storeroom memory function is increased by the vaccine operator and the second response period is shorten obviously, while the positive selection operator is the executant. The improved adjustable fuzzy matching negative selection algorithm has better self-study function, it has self-adaptation for the inherent immune response.

Key words: fuzzy matching; immune response; vaccine operator; positive selection operator; self-adaptation

机体免疫功能是对抗原刺激的应答, 称为免疫应答, 这种应答分为先天性免疫应答和自适应免疫应答. 接种疫苗可以建立自适应免疫应答, 也称为特异性免疫反应^[1-2]. 特异性免疫是机体通过记忆或借助外部力量调节自身以增强免疫能力的过程. 免疫系统通过疫苗的作用, 提高个体的环境适应能力. 焦李成提出基于疫苗的免疫进化算法, 把工程问题的某些先验知识抽象成疫苗, 通过接种疫苗和免疫选择来提高候选解的适应度并防止解的退化^[3], 大大改善候选解的质量, 在该领

域产生了深远的影响. 文献[4]提出了在最优路径中父代最优个体疫苗, 王磊在遗传算法中引入包括免疫疫苗的免疫算子, 构造了一类免疫遗传算法, 并用于求解 TSP 问题^[5]. 文献[6]提出了一个网络入侵检测模型, 该模型集成了阴性选择、克隆选择与接种疫苗算子, 具有分布式、自组织和轻量级的特性. 本文将疫苗理论引入到阴性选择算法中, 建立可变模糊匹配阴性选择免疫算法^[7]的特异性免疫反应功能. 由于阴性选择算法中需多次遍历检测器集和自体集并计算抗原与检测器匹配度, 因此匹配算法的效率及检测器集检测自体集的效率是影响阴性选择算法效率的关键因素, 将疫苗算子加入到检测器集中, 并采用基因位 r 间隔迭次取用思想, 可以大大提高检测效率, 缩短二次应答时间. 同时利用疫苗提取与接种过程动

收稿日期: 2009-12-23.

基金项目: 国家自然科学基金资助项目(60704004); 黑龙江省博士后基金资助项目(LBH-Z09216); 中央高校基本科研业务费专项基金资助项目(HEUCF00401).

作者简介: 王 辉(1976—), 博士, 副教授.

态刷新抗体库,增强抗体库的记忆功能. 仿真表明疫苗算子的加入使算法具有更快的运行速度和更低的检测失败率,抗体库的更新具有自适应性.

1 可变模糊匹配阴性选择免疫算法

可变模糊匹配阴性选择免疫算法是基于普通阴性选择算法提出的,该算法匹配阈值可变,通过检测及调整匹配阈值来减少传统阴性选择算法中“黑洞”的数量. 同时,由于抗体与抗原匹配时具有不确定性和模糊性,算法利用模糊思想,定义模糊相似度,实现了抗体与抗原的带有控制参数的模糊匹配,从而提高系统的检索率和性能.

2 疫苗算子

2.1 疫苗算子

在算法中加入疫苗算子可以增强算法的记忆功能,同时可以提高算法的响应速度. 疫苗算子可以由算法自身在检测过程中产生,这个过程与生物体免疫过程中所产生的记忆细胞类似,因此,疫苗能够增强算法的记忆功能. 同时,由于疫苗只是个体在某些基因位上的特征,其本身并不是个体,因此疫苗需要某种载体来存放. 而疫苗是基于抗原产生的,因此疫苗算子的获取在连续 r 位匹配规则下,应该是连续 r 个基因模式,而疫苗所在的抗原既具有特征基因位,又具有 r 个基因模式,将抗原作为该疫苗的载体进入疫苗抗体库,可形成疫苗算子. 利用可重用的检测元作为疫苗抗体,既保证了疫苗的质量,又增加了抗体的多样性.

2.2 正选择算子

在生物免疫中,免疫细胞所经历的逆于负选择(阴性选择)的另一个成熟抗体的过程就是正选择过程,与负选择同样重要^[8]. 在以往的相关文献中,往往都忽视了正选择这一重要过程. 抗体库在生成时,是一个负选择的过程,算法所关心的是没有任何匹配发生时待选检测器才能进入抗体库. 而在用疫苗进行检测过程中,与疫苗发生匹配的是检测到的二次入侵的抗原,算法所关心的就是匹配是不是发生了,因此是一个正选择的过程.

2.3 提取疫苗

本算法中对疫苗的提取采用自适应的方法. 针对 self 集首先利用可变模糊匹配阴性选择免疫算法的有效检测器集生成算法^[9]生成有效检测器集 D ,改变 self 集的某一个片段,模拟抗原的入侵,并用检测器集 D 检测. 若发生了正选择匹配,则将基因连同 self 集中的载体提取为疫苗,加入到抗体库(检测器集)中. 此时抗体库包含 2 部

分:有效检测器集 D 和疫苗. 利用抗体库继续检测 self 集的下一个元素:先由疫苗进行正选择匹配,若有匹配发生,则产生报警(检测到抗原);若没有检测到,再由有效检测器集 D 进行检测,将检测到的匹配提取为疫苗,更新抗体库. 随着 self 集的改变,抗体库中的疫苗会自适应地产生,并参与下一轮的检测,抗体库得到动态的刷新.

疫苗提取算法步骤如下:1)由可变模糊匹配阴性选择免疫算法生成有效检测器集 D ;2)生成抗体库;3)读待检文件,进行疫苗正选择匹配. 若有匹配发生,产生报警;否则,利用检测器集 D 检测;4) D 检测若有匹配发生,提取为疫苗,转到第 2 步;5)否则,若检测尚未完成,转到第 3 步;6)若检测完成,生成检测报告.

3 具有疫苗算子的可变模糊匹配阴性选择算法仿真

疫苗算子和正选择算子的主要作用就是对抗原形成免疫记忆,从而建立二次免疫应答,并大大缩短二次应答时间,因此,这 2 种算子应加在可变模糊匹配阴性选择免疫算法的监测系统异常部分. 在监测阶段加入疫苗算子和正选择算子,可使算法在监测阶段的二次应答时间大大缩短. 由于提取疫苗时采用自适应方法,使抗体库在每检测到一个抗原时能得到动态刷新,从而使可变模糊匹配阴性选择免疫算法的检测效率明显提高,并能对再次入侵的抗原作快速应答.

3.1 检测率分析

设 P_{Mi} 为 2 个随机串在 r_i 连续位匹配规则下的匹配概率,则

$$P_{Mi} = (l - r_i + 1) \cdot 2^{-r_i} - (l - r_i) \cdot 2^{-r_i-1} = 2^{-r_i} \left(\frac{l - r_i}{2} + 1 \right). \quad (1)$$

对于一个随机串,在匹配阈值为 r_i 时,不与任何 self 串匹配的概率为^[10]

$$f_i = (1 - P_{Mi})^{N_s}. \quad (2)$$

其中: N_s 为 self 集的大小,则该串成为匹配阈值为 r_i 的成熟检测器概率为

$$P_i = (1 - f_1) \cdot (1 - f_2) \cdots (1 - f_{i-1}) f_i. \quad (3)$$

对于 N_{R0} 个未成熟检测器,根据上述概率关系可以生成成熟检测器的个数为

$$N_{Ri} = N_{R0} \cdot P_i. \quad (4)$$

则漏报率 P_f 满足^[11]:

$$P_f = (1 - P_{M1})^{N_{R1}} + \cdots + (1 - P_{MC})^{N_{RC}}. \quad (5)$$

进而可得检测率为

$$P_s = 1 - P_f = 1 - (1 - P_{M1})^{N_{R1}} - \cdots -$$

$$(1 - P_{MC})^{N_{RC}} \tag{6}$$

由于疫苗的加入,使得抗体库中成熟检测器的个数 N_{Ri} 增大,根据式(6)可知抗原检测率明显提高.特别是对具有疫苗基因位的特定抗原,检测失败率大大降低,对于二次入侵的抗原,疫苗检测率为 100%,且应答时间明显缩短.

3.2 基因位 r 间隔迭次取用思想

算法仿真时,生成长度为 256 字节的二进制文件作为 self 集 N_s . 取初始匹配阈值 $r = 7$,最大匹配阈值 $r_c = 11$,通过可变模糊匹配阴性选择免疫算法生成有效检测器集 D. 采用模糊匹配方法实现提取疫苗操作及监测 self 集的改变. 由于 self 集的长度为 256 字节,共有 2 048 个二进制位,并且没有形成独立的字符串元素的形式,因此,首先要将 self 集划分成若干个字符串的形式. 本文在仿真时,采用基因位 r 间隔进行迭次取用,即第 1 个位串的起始位置为 1,第 2 个起始位置为 8,依此类推. 这样既可以减少计算的冗余度,又不会过多地丢失基因位,保证了算法的时间复杂度不会过大,降低检测失败率,提高算法的检测效率.

3.3 仿真分析

仿真时,由于匹配阈值在 7 ~ 11 之间可调,所以 D 集中就包含匹配阈值 7 ~ 11 之间的检测器. 采用 $r = 7$ 进行疫苗匹配检测,根据检测器生成过程可知,匹配阈值大于 7 的检测器,会与 self 集中的个体发生匹配,因此在没改变 self 集时却检测到了某个匹配. 为了保证了在 self 集没有改变时,不会被检测器检测到,取初始匹配阈值 $r = 6$,最大匹配阈值 $r_c = 8$,利用可变模糊匹配阴性选择免疫算法重新生成检测器集 D. 在加入疫苗算子的可变模糊匹配阴性选择免疫算法中取迭次间隔为 6,匹配阈值为 9,分别改变 self 集的某一个片段,检测算法的二次应答时间. 结果发现算法的二次应答时间并没有减少,这说明疫苗算子并没有发挥其真正作用,因此,需进一步改进算法流程.

3.4 取 self 集整体为研究对象的仿真实验

由于二次应答时间除与 self 集的大小、机器的速度及抗体数、疫苗数等有关外,主要取决于算法本身的流程. 由于在监测阶段,算法的设计思想是针对 self 集中的每一个位串进行检测,先用所有疫苗进行正选择检测,若没有匹配发生,再用检测器集进行检测. 初次应答时,抗体库中的疫苗数较少,检测相对较快,但二次应答时已经积累了经验,疫苗数量多,这样每一个位串在进行疫苗检测时耗费的时间相对增多,而下一个位串要在前一个位串检测器集 D 集检测后才能进行检测,检测

器集检测也要耗费一定的二次应答时间,因此,当检测到某一个二次入侵的抗原位串时,其前面的所有位串的检测器集检测耗费的时间与初次应答的基本相同,二次应答所耗费的时间并没有减少.

取 self 集整体为研究对象,调整算法流程,检测操作时仍然采用 r 间隔迭次取用. 首先利用疫苗进行正选择匹配检测,扫描 self 集整体,提高抗体库二次应答的速度,然后再用检测器集检测,提取疫苗并更新抗体库. 仿真实验时改变原文件的某个片断,对疫苗数、二次应答时间及检测失败率等进行检测,每一种实验重复 100 次,取平均值作为实验最终结果,见表 1 及图 1、图 2 所示. 从实验数据可以看出,算法进一步改进后,二次应答时间明显减少,检测失败率显著降低,检测效率大大提高,增强了算法的免疫自适应性. 表 1 中,最小抗体数与实际抗体数并不是成比例变化,这主要是由于初始抗体生成时的随机性以及自我集的相似性,导致实际抗体的冗余性造成的. 在各种类型的实验中,单个字变化的检测是最难的,所需的抗体数较多. 但由于 r 间隔迭次取用的方法可以有效地避免因 self 文件的数据被等长划分而造成的基因缺失,使得算法对单个字的应答可以扩展到字长度的 1/4 次数(8 次),因而对单个字的检测率与其他算法相比大大增加.

表 1 算法进一步改进后的实验结果

初始抗体数	实际抗体数	最小抗体数	抗原数	疫苗数
200	15	10	5	2
400	22	11	4	3
600	38	12	7	5
800	29	9	5	3
1 000	46	18	10	4

3.5 与小生境克隆选择算法的实验对比

仿真时取 $m = 2, l = 32, N_{R0}$ 为初始抗体数, N_R 为实际抗体数(含疫苗), P_f 为检测失败率,并在检测时间相同的情况下与文献[11]的小生境克隆选择算法进行了实验对比分析,实验结果如表 2 所示. 可以看出,即使当 self 文件较大时,加入疫苗算子的可变模糊匹配阴性选择免疫算法也仍然要明显好于小生境克隆选择算法. 这是因为克隆选择算法在克隆过程使抗体间具有极高的相似度,限制了抗体的空间分布,使抗体的多样性受到抑制. 而加入疫苗算子的可变模糊匹配阴性选择免疫算法通过提取疫苗来获得新的抗体,疫苗是来源于 self 信息之外的所遇到过的无限多样的 Nonself 字符,这样既保证了抗体的多样性,也保证了检测率的提高. 表 2 中数据是在检测时间几

乎相同的情况下得到的,从而也进一步验证了加入疫苗算子的可变模糊匹配阴性选择免疫算法的执行效率要明显高于小生境克隆选择算法。

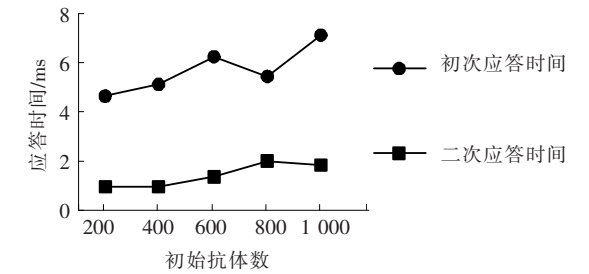


图 1 应答时间对比

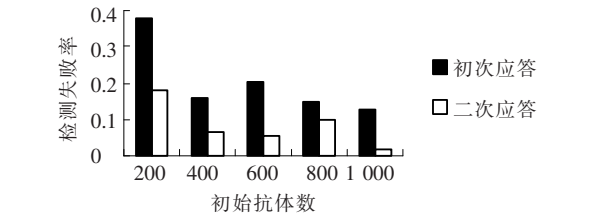


图 2 检测失败率对比

表 2 检测时间相同的实验结果

算法	N_s	N_{R0}	N_R	抗原数目	P_f
加入疫苗算子	16	2	134	499	0.002
的可变模糊	32	2	99	496	0.008
匹配阴性	64	3	123	498	0.004
选择算法	128	11	98	497	0.006
	16	6	124	496	0.008
小生境克隆	32	6	113	463	0.074
选择算法	64	8	135	471	0.058
	128	11	109	458	0.084

3.6 疫苗算子对算法性能影响的分析

加入疫苗算子使算法获得了自学习功能,只需要较少的初始种群,经过不断地学习来逐步充实完善其抗体库.同时,疫苗的提取过程保证了抗体的多样性,并通过疫苗的“记忆”功能来增加算法的检测经验.事实上,疫苗能否发挥作用取决于算法能否积累到“学习经验”,也就是能否检测到 Nonself.检测效果的好坏也和学习次数的多少以及学习质量有很大的关系.此外,疫苗的存在使算法的时间复杂度增加.在进行疫苗的正选择检测过程中,时间复杂度取决于抗体库中疫苗的数量,并与问题的规模成正比.由于疫苗算子的加入使算法中增加了线性阶的时间复杂度,二次应答时间的缩短是以时间复杂度的增加为代价,但时间复杂度增加有限,而应答时间对检测算法来说是很重要的性能指标之一,疫苗算子显著的提高了算法的快速性,增强了算法的免疫自适应能力,明显地改善了算法的检测性能。

4 结 论

1)根据生物免疫系统中疫苗免疫的理论,提

出了一种可变模糊匹配阴性选择免疫算法的改进设计方案,在算法的异常监测阶段加入了疫苗算子和正选择算子。

2)分析了疫苗算子的提取过程,通过疫苗的提取,使正选择算子记住了环境改变的方式,算法的记忆功能得到增强,算法的快速性得到提高,算法对再次入侵的抗原二次应答时间明显减少。

3)采用基因位 r 间隔迭代取用的方法,保证了算法中位串基因信息的完整性,降低了检测失败率,检测效率大大提高.疫苗算子的加入使抗体库能够随着 self 集的改变而动态刷新,改善了算法的性能,可提高算法的免疫自适应性。

参考文献:

[1] 丁永生,任立红. 人工免疫系统理论与应用[J]. 模式识别与人工智能,2000,13(1):52-59.

[2] 肖人彬,王磊. 人工免疫系统原理、模型、分析及展望[J]. 计算机学报,2002(12):1283-1293.

[3] JIAO Licheng, WANG Lei. A novel genetic algorithm based on immunity[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems and Humans, 2000, 30(5):552-561.

[4] 杨辉,康立山,陈毓屏. 一种基于构建基因库求解 TSP 问题的遗传算法[J]. 计算机学报,2003,26(12):1753-1758.

[5] WANG Lei, PAN Jin, JIAO Licheng. The immune algorithm[J]. ActaElectronica Sinica; in Chinese, 2000, 28(7):74-78.

[6] FANG Xianjin. An artificial immune model with vaccine operator for network intrusion detectio[C]//Proc of 2008 IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application. Wuhan: IEEE CS Press, 2008: 488-491.

[7] 王辉. 可变模糊匹配阴性选择免疫算法研究[D]. 哈尔滨:哈尔滨工程大学,2007.

[8] BENTLEY P J, GORDON T, KIM J. New trends in evolutionary computation[C]//The Congress on Evolutionary Computation(CEC-2001). Seoul, Korea: [s. n.], 2001: 162-169.

[9] 王辉,王科俊. 基于免疫的最小有效检测器集生成算法[J]. 计算机工程,2008,34(9):219-221.

[10] ZHANG Yajing, HOU Chaozhen, WANG Fang. A clone selection algorithm with niching strategy inspiring by biological immune principles for change detection[C]//IEEE International Symposium on Intelligent Control. Texas, US: The Westin Galleria Houston, 2003:1000-1005.

[11] 张海英,管洪娜,潘永湘. 一种改进的阴性选择算法[J]. 西安理工大学学报,2005,21(3):306-309.

(编辑 魏希柱)