

结合危险理论的故障检测免疫模型

王三虎

(吕梁学院 计算机科学与技术系, 山西 吕梁 033000)

摘要: 为了实现快速准确的网络故障检测与诊断, 将危险理论与动态克隆选择算法相结合, 提出了用于网络故障检测的危险理论免疫模型; 并针对网络故障特点, 对危险理论与动态克隆选择算法进行了改进; 首先采用危险理论模型对抗原进行危险信号浓度识别, 并利用成熟检测器检测已知故障类型, 其次用改进的动态克隆选择算法对未知故障进行有效的学习; 通过对多种网络故障类型检测的仿真实验, 证明了新模型不仅具有更好的检测效果和动态适应性, 而且能够提高检测效率与准确率。

关键词: 危险理论; 人工免疫系统; 网络故障检测; 动态克隆选择算法

An Immune Model for Network Fault Detection Based on Danger Theory

Wang Sanhu

(Department of Computer Science and Technology, Lvliang University, Lvliang 033000, China)

Abstract: Aiming at the fast and accurate network fault detection and diagnosis, combining the danger theory with dynamic clonal selection algorithm, a danger theory based immune model used for network fault detection is proposed. And according to the characteristics of network failure to improve danger theory and dynamic clonal selection algorithm. Firstly, the concentration of antigen danger signal is recognized by using the danger theory model, and using the mature detector to detect known fault type. Then an improved clonal selection algorithm is set up to learn the unknown fault types. Experiments were undertaken with various type of network fault diagnosis to demonstrate that the model not only has better detection rate and adaptability but also can improve the efficiency and accuracy.

Key words: danger theory; artificial immune system; network fault detection; dynamic clonal selection algorithm

0 引言

随着网络系统规模和复杂程度不断增加, 对网络维护也变得更加复杂, 及时发现网络故障并对其进行准确定位有助于尽快恢复网络运行。人工免疫系统作为一种智能技术, 它所具有的异常检测、多样性、交互性、连续学习、记忆等特性为智能诊断技术的发展提供了新的理论基础^[1-2]。

在人工免疫系统中, “自己/非己”识别模式(SNS)通过系统对异常变化的成功监测而使免疫系统发挥作用, 而检测成功的关键是系统能够区分自己和非己数据, 并通过免疫应答机制排除“非己”。由于其简单易行, 在异常检测和故障诊断等领域得到广泛应用。但由于计算机网络的设备更换、网络设置改变, 自体和非自体之间的界限会越来越模糊, 导致系统不能正确识别网络状态, 降低了系统的诊断准确率^[3-4]。

随着生物免疫机制的进一步发展, 出现了一种有别于“自己-非己”识别的另一免疫应答模式——“危险理论”^[5]。与传统SNS模式相比, 危险理论认为引起免疫响应的关键因素是机体产生的危险信号(Danger Signal)^[6], 并且“危险”并不等同于“非己”, 免疫应答是由产生于多种不同危险信号之间相互关联的“危险”引起的。

动态克隆选择算法(DynamiCS), 具有动态的学习和识别能力, 可检测到动态变化的网络入侵行为^[7]。如果诊断系统可以以环境中出现的新状况进行适应性学习, 将有效提高故障诊

断的准确性。

为实现快速准确的网络故障诊断, 本文利用危险理论原理和人工免疫系统的学习机制, 根据故障检测与诊断的特点, 对危险理论与动态克隆选择算法进行了改进, 并提出一种用于网络故障检测的免疫模型, 新的模型结合了危险理论良好的危险检测效果以及动态克隆选择算法的动态适应性, 用于网络故障诊断可以有效的提高故障的识别率。并对动态克隆选择算法中检测器的克隆变异策略进行了优化, 使成熟检测器的生成效率有明显的提高, 保证了较高的检测率和学习效率。

1 危险理论模型

鉴于“自己/非己”识别机制存在的问题, 免疫学家 Polly Matzinger 从全新的角度, 提出了一种危险理论。该理论认为, 受损细胞向抗原提呈细胞 APC 发出的危险信号是启动免疫响应的关键。APC 捕获到危险信号后, 使处于静止状态的 APC 转变成激活状态, 然后经过一系列细胞间的相互作用, 激活 T 细胞启动特异性免疫应答, 杀死抗原恢复机体健康。没有危险信号时, 抗原提呈只会使 T 细胞失活。图 1 所示为危险模型的免疫识别过程。

危险理论认为引起免疫应答的关键因素不是机体中被检测出非己抗原, 而是机体中是否存在足够强度的危险信号, 它不是完全否定传统免疫理论, 而是对其进行了补充。如果被检测抗原对机体细胞产生了损害, 这些机体细胞就会发出危险信号, 抗原提呈细胞就会搜集这些危险信号, 并将其提呈给免疫细胞, 进而激活免疫应答清除该抗原。这很好地解释了传统免疫理论不能解释的问题。

2002 年, 随着 UWeAicklin 博士将危险模式理论引入人工免疫, 新的危险理论的模型不断被出现, 并在工程领域得到应

收稿日期: 2014-02-08; 修回日期: 2014-03-27。

基金项目: 吕梁学院自然基金项目(ZRXN201216, ZRXN201308)。

作者简介: 王三虎(1969-), 男, 山西吕梁市人, 硕士, 副教授, 主要从事网络通信教学和科研工作方向的研究。

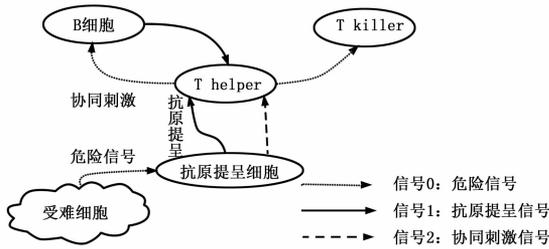


图 1 危险模式应答过程

用。使得新的危险理论开始挑战传统的 SNS 模型。危险理论认为：免疫应答并不是由非自体引起的，而是由机体识别到的危险信号引起的。该新理论确实能解释传统的 SNS 模型所不能解释的一些现象。但是危险信号的计算是一个非常复杂的过程，要消耗大量的系统资源。

危险理论以危险信号为先导只识别系统中发出危险信号的细胞周围区域中的抗原，不需要匹配和处理危险域外的其他异己抗原。这样可以很大程度的降低系统的免疫响应的规模跟次数，并且还可以保证更加精确的识别和清除有害抗原，实际操作性更强。传统免疫系统为使能匹配更多范围的抗原，在生成抗体时需要不断进行变异，而危险理论只识别危险信号并不注重抗原的异己性，这样抗体的产生过程就不需要复杂的成熟变异，当自体集发生变化的时候，所受影响也远小于 SNS 模式。由此可见，危险理论模型具有很强的容错性以及较强的自调节性。

根据危险理论原理设计实现的人工免疫系统不需要事先训练大量样本，也不用区分自体或是非自体，所以更加简单快速，具有检测效率高、实时性强等特点。

2 结合危险理论的人工免疫检测模型

利用危险理论的工作原理以及人工免疫系统的识别及学习机制，综合采用改进的危险理论、动态克隆选择算法的优点对检测过程进行优化，提出用于网络故障检测与诊断的危险理论免疫故障检测模型。模型主要包括危险信号检测模块、免疫响应及抗原学习模块等。

首先获取与网络设备运行状态相关的性能参数值进行数据处理，并组成特征向量，作为抗原集 Ag 。当系统处于正常运行状态时，收集的系状态特征向量所组成的集合即为自体集，用 S 来表示；设备系统中各种类型的故障映射为故障检测器，包括成熟检测器、记忆检测器等。不同故障类型映射为不同的检测器，用 A_i 来表示。图 2 描述了本文提出的检测系统模型。

3 危险信号检测

此模块主要是利用危险理论良好的小样本分类效果对抗原的危险度进行计算。将搜集得到的信号进行处理后，根据周围环境有无危险信号将抗原划分为故障和无故障抗原，将无故障抗原加入自体集作为成熟检测器的训练集，将异常抗原提呈给下个模块进行处理。

3.1 输入信号

1) 危险信号：危险信号在网络故障检测中表示抗原与故障检测器的区域较接近，危险信号计算为：

$$D_{danger} = \sqrt{\sum (Ag - S)^2} \quad (1)$$

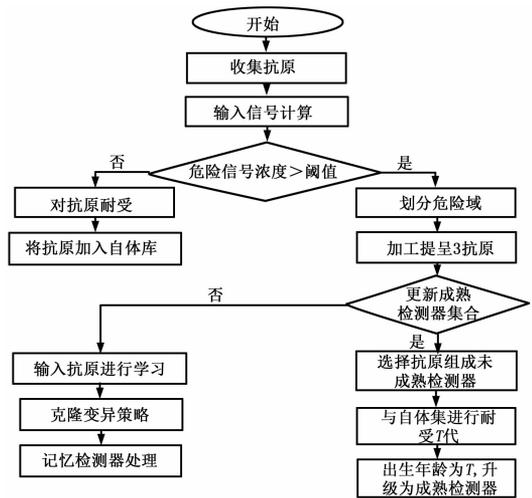


图 2 检测与诊断模型

2) 安全信号：在网络环境中可以用正常的网络特征来表示。

$$D_{safe} = \sqrt{\sum (Ag - A_i)^2} \quad (2)$$

危险信号检测的主要任务为通过输入信号，得到抗原危险信号的浓度 C_{mav} ，并判断 C_{mav} 值属于哪个区间，将判断结果发送给免疫响应模块。

3.2 危险信号浓度计算

抗原危险信号的浓度值可以由以下公式计算。

$$C_{mav} = \frac{D_{danger}}{D_{danger} + D_{safe}} \quad (3)$$

如果危险信号浓度的值超过了一定的阈值，则认为有抗原发生了异常，激活免疫响应；否则，则认为没有异常发生。

3.3 危险区域界定

危险区域的建立是以发现某一危险抗原为基础。如果已经确定了某一抗原存在危险信号的浓度超过危险阈值，则在此其抗原周围一定范围内建立危险区域。

危险区域的建立是通过当前检测抗原与其邻近抗原的相似性值 Aff 获得。选择 k 个相邻抗原，计算每个邻近抗原与当前抗原的相似性值 Aff ，危险域的值由以下公式计算。

$$Z_i = \sum_{j=1}^k Ag_j (Aff(Ag_i, Ag_j) \leq \gamma) \quad (4)$$

其中： Z_i 表示危险区域， Ag_i 表示已经确定为危险的抗原， Ag_j 为抗原 Ag_i 邻近的抗原， γ 为亲和力阈值， k 表示危险区域内的危险抗原个数。

如果输出结果为危险信号浓度超过危险阈值，则激活免疫应答，将抗原提呈到下一阶段的故障类型识别及免疫学习进行免疫响应过程。

4 免疫响应及抗原检测

由危险信号激活免疫响应过程，将抗原提呈并利用成熟检测器种群进一步对抗原进行故障类型的识别。本模型中对成熟检测器的生成过程进行了改进，依据危险信号检测模块的处理结果，从含有危险信号的抗原集合中随机抽取 n 个抗原形成相对应的 n 个未成熟检测器，再与自体进行耐受 T 代，最后将存活下来的未成熟检测器加入到成熟检测器中，这样可以大大提高成熟检测器的生成效率。

4.1 成熟检测器更新

故障知识库中保存能检测已知故障类型的成熟检测器，每种预测故障类型对应一个成熟检测器。利用故障知识库中的成熟检测器对抗原数据进行已知故障的检测与诊断，识别抗原所属的故障类型。计算成熟检测器与抗原之间的亲和力，对抗原进行响应，判断故障类型。

成熟检测器种群 $A(k)$ 的亲和力通过式 (5) 计算得到。

$$\varphi(A(k)) = [\varphi(A_1(k)), \varphi(A_2(k)), \dots, \varphi(A_n(k))] \quad (5)$$

其中： φ 为亲和力计算函数，则成熟检测器种群 $A(k)$ 中的第 i 个检测器 A_i 的亲合力用式 (6) 表示。

$$fit(i) = \varphi(A_i) \quad (6)$$

式中， $fit(i)$ 表示检测器与抗原的亲合力大小。

对第 k 代的成熟检测器群 $A(k)$ 依据亲和力大小，将群体分为记忆单元和一般单元，即：

$$A(k) = [M(k), A_b(k)]$$

其中： $M(k) = \{A_1(k), A_2(k), \dots, A_t(k)\}$ ， $A_b(k) = \{A_{(t+1)}(k), A_{(t+2)}(k), \dots, A_n(k)\}$ ， $t = fix[n \times (S_c + D_{is})]$ 。

$fix(\cdot)$ 为下取整函数， $fix(x)$ 表示不大于 x 的最大整数， S_c 是为了保证记忆单元基本大小而设置的常量，取值为 0.8；另外

$$D_{is} = \sqrt{\frac{1}{(n-1) \times n} \sum_{j=1}^n \sum_{i=1}^n D_{ij}} \quad (7)$$

其中： $D_{ij} = \exp\{\|A(k)_i - A(k)_j\|\}$ ， $i \neq j$ ； $i, j = 1, 2, \dots, n$ ， $\|\cdot\|$ 为任意范数，在十进制编码中多取为欧几里德距离， $A(k)_i$ 表示第 k 代检测器群体中的检测器 A_i 。 D_{is} 度量了检测器间的多样性， $0 \leq D_{is} \leq 1$ ， D_{is} 越大，表明检测器的多样性越好，反之，多样性越差。

如果故障知识库中的检测器未能检测出抗原类型，说明该抗原为未知故障，需要免疫学习，识别该抗原，生成新的检测器并经过耐受成为成熟检测器，将其保存到记忆检测器种群中。由于其模拟了生物免疫机制中的二次应答，所以记忆检测器集合能够快速检测出与之前相同或是相似的抗原。

4.2 克隆变异策略

故障知识库中的成熟检测器未能检测出抗原时，需要对抗原进行学习，即检测器耐受。根据危险区域定义初始未成熟检测器，再对未成熟检测器经过耐受，达到成熟，检测器的耐受过程包括克隆算子、变异算子。通过克隆变异算子，能够有效地扩展检测器的检测范围，并有机会获得更佳检测器。而通过对检测器的评估。

1) 克隆算子：学习过程中的克隆算子借鉴标准克隆选择算法的克隆机制，新检测器和原有检测器将竞相匹配更多的抗原以获得高亲和力。拥有较高亲和力的检测器将获得较长的生存期和更多的克隆体。

在检测器繁殖过程中的第 t 次迭代时，设每个检测器 A_i 克隆的检测器数量为 $N_{ci}(t+1)$ ，它产生的克隆数量为 $N_{ci} = round(\beta * m(t)/i)$ 。其中， β 为繁殖系数。

克隆产生的检测器总数为 $N_c(t+1)$ ，将其定义为：

$$N_c(t+1) = \sum_{i=1}^{m(t)} round \frac{\beta * m(t)}{i} \quad (8)$$

2) 变异算子。实施克隆操作后的检测器要经历变异，因为变异操作可使免疫系统具有可扩展性、动态适应性和自学习等能力，使免疫系统能够快速有效的识别抗原。本文采用高斯

变异，因为高斯变异可以产生位于原始检测器附近的变异，使得原始检测器的种群信息得以保留。

对检测器 $A_i (i = 1, 2, \dots, n)$ 的第 j 维分量 $A_i(j), i = 1, 2, \dots, n$ 进行如下操作：

$$A_i'(j) = A_i(j) + \eta'(j)N(0,1) \quad (9)$$

$$\eta'(j) = \eta_i(j) \cdot \exp(\tau' \cdot N(0,1) + \tau \cdot N_i(0,1)) \quad (10)$$

其中： $A_i(j)$ ， $A_i'(j)$ ， $\eta_i(j)$ ， $\eta'(j)$ 分别表示向量 a_i ， a_i' ， η_i ， η_i' 的第 j 个分量， $N_i(0,1)$ 表示对应每一个 j 所产生的正态随机数， $N(0,1)$ 是满足均值为 0，方差为 1 的正态随机变量， τ 和 τ' 分别取 $(\sqrt{2\sqrt{n}})^{-1}$ 和 $(\sqrt{2n})^{-1}$ 为参数值^[9]。

每个检测器的可克隆的数目取决于其亲和力的大小，而对于亲和力为零的检测器，不能参与克隆变异的过程，它们将被淘汰出整个繁殖过程。

经过对检测器的克隆和变异操作，检测器数量大大增加，位置也发生了一定的变化，检测器的检测范围可能相互重叠，导致冗余检测器的产生，因此，需要消除这些冗余检测器。经过克隆和变异的繁殖过程，生成的成熟检测器保存于成熟检测器库中，以便下次类似抗原出现时能够进行快速检测。

5 算法描述

结合危险理论的免疫算法的流程描述如下：

Step1：收集周围环境中的抗原；

Step2：对于每组抗原数据，计算其相应的危险信号值和安全信号值，将其作为危险模型的输入信号；

Step3：根据输入信号计算危险信号的浓度 C_{csm} ，并划分危险域的范围；

Step4：将危险信号 C_{csm} 与阈值对比，若 $C_{csm} >$ 阈值，则激活免疫响应，将危险域内的抗原进行加工提呈；否则，返回 Step2；

Step5：成熟检测器处理输入抗原，通过计算亲和力判断成熟检测器是否识别抗原，若亲和力小于预定义的阈值则将该检测器的匹配数加 1。当所有的抗原都处理完成后，统计各个成熟检测器的匹配数；

Step6：将生存期达到 T 代的检测器升级为成熟检测器；

Step7：记录每个成熟检测器的年龄以及与抗原的匹配数，将匹配数达到规定阈值的成熟检测器升级为记忆检测器，将存活年龄达到阈值的从成熟检测器集合删除；

6 实验及结果分析

网络的当前运行信息是诊断网络所处状态的关键，可基于 SNMP 协议对网络设备进行轮询，以获取当前网络状态下相关设备的 MIB 变量取值。实验中定义了 7 种网络状态，如表 1 所示，包括正常状态和 6 种故障状态。

表 1 网络故障类型集合

编号	故障类型
0	正常状态
1	接口不通
2	带宽不足
3	通信协议不兼容
4	缓冲区不足
5	接口性能出现问题
6	线路故障

获取与网络设备运行状态相关的性能参数值，进行数据处理，得到的故障特征集合如表 2 所示。

表 2 网络故障特征集合

编号	故障特征
S1	接口工作状态
S2	输入丢弃率
S3	输出丢弃率
S4	输入错误率
S5	输出错误率
S6	接口利用率
S7	未知协议率

危险信号的获取与处理，使用 RIIL v6.0 IT 综合管理平台来检测网络设备的 CPU 利用率，获取网络的拓扑情况，并获取相应设备的网络参数；使用 Allot NetXplorer 来检测网络的流量状况。通过对危险信号的采集，当网络出现故障时，根据警报的来源锁定发生异常情况的设备，并以该设备为中心划分危险域，采集与其相连的设备中的 MIB 信息作为输入抗原。

在故障检测中，为了将结合危险理论的算法与“自己/非己”识别模式进行比较，用检测率和误检率为衡量检测系统指标。分别用两种模型对同 9 组不同的实验样本进行测试，测试结果统计如表 3 所示。

表 3 使用 SNS 模型进行故障检测的结果统计

检测次数	检测率 (%)		误检率 (%)		系统使用时间 (s)	
	SNS	SNS-DANGER	SNS	SNS-DANGER	SNS	SNS-DANGER
1	86	92	14	8	0.326	1.23
2	57	90	43	10	0.348	1.74
3	76	97	24	3	0.323	2.03
4	69	93	31	7	0.590	1.89
5	73	90	27	10	0.464	1.35
6	89	96	11	4	0.632	1.24
7	47	92	53	8	0.398	1.38
8	76	93	24	7	0.476	2.16
9	81	98	19	2	0.359	1.67

对本文提出的模型进行数据分类实验的结果如图 3 所示。

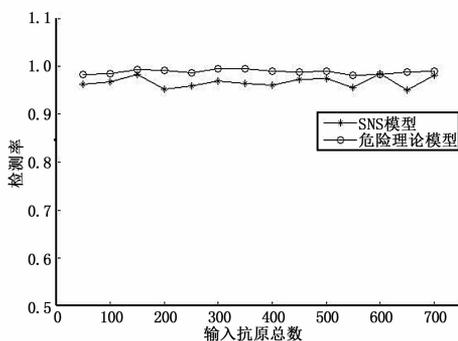


图 3 两种模型的检测率对比图

图 3 是危险理论免疫模型与 SNS 算法分别经历环境状态变化时抗原数目不断增加时的检测率对比关系，可以看出 SNS 算法的检测率有明显的波动。对比而言，危险理论免疫

算法整体检测率都高于 SNS 算法，且稳定性较好。危险理论免疫算法可以维持在一个较高的识别率状态。

7 学习效率

以接口性能问题和线路故障以及带宽不足的复合故障为例，对本文所提出的学习方法与 SNS 算法进行了实验比较。

由于在提出的学习机制中使用改进的克隆与变异的方法，使抗体克隆变异的范围缩小，可以在很大程度上提高系统的检测效率，加快了检测器的进化速度；最佳亲和力的收敛速度相对较快，并且学习精度也较高，提高了检测率。

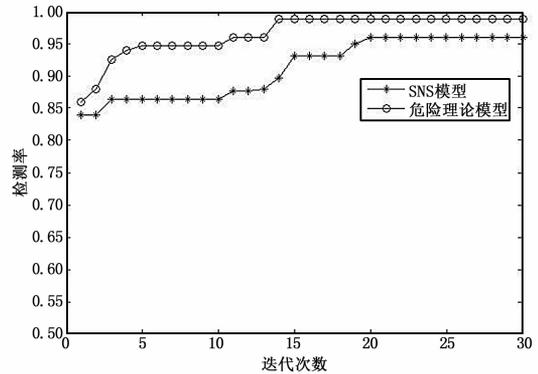


图 4 两种模型的检测率对比图

图 4 所示的对比结果可以很清楚地看出本文提出的新的故障检测模型的检测率要高于传统的 SNS 算法。图中 SNS 算法在运行到第 2 代的时候检测率有个明显的提升，这是由于本实验中成熟检测器的匹配阈值设置为 5，当成熟检测器匹配数达到 5 之后则升级为记忆检测器，从而提升了算法的检测率；基于危险理论的免疫模型在初始的几次迭代运行中，检测率是保持持续上升的，这主要是因为新模型不断的对成熟检测器中亲和度较低的检测器进行变异操作，从而使检测器能更好的匹配更多的抗原。

8 结论

结合动态克隆选择算法和危险理论，设计实现了一种新的免疫检测模型用于网络故障的检测与诊断，并对动态克隆选择算法中成熟检测器的动态更新以及克隆变异策略进行了优化。该模型将收集到的抗原利用危险理论提前进行预分类，从含有危险信号的抗原集合中抽取部分抗原作为动态克隆选择算法的输入，利用危险理论所具有的良好分类效果简化了动态克隆算法的自我更新过程。通过实验验证了对检测器生成策略优化之后成熟检测器的生成效率有明显的提高，保证了较高的故障检测率和学习效率。

参考文献

[1] Reischuk R, Textor J. Stochastic search with locally clustered targets: learning from T cells [A]. ICARIS [C]. 2011, 6825: 146-159.

[2] 宋建厚, 陈良琼, 刘道华. 改进的免疫算法参数自适应调整的优化设计 [J]. 计算机测量与控制, 2013, 21 (5): 1297-1299.

[3] Laurentys C A, Ronacher G, Pallhares R M, et al. Design of an artificial immune system for fault detection: a negative selection approach [J]. Expert Systems with Application, 2010, 37: 5507-5513.

