

基于残余抗原学说的动态记忆风险识别模型

陶媛¹ 胡珉² 王萍¹

(上海大学计算中心 上海 200072)¹ (上海大学悉尼工商学院 上海 201800)²

摘要 以小概率事件风险识别为研究对象,提出一个基于残余抗原学说的动态记忆风险识别模型 DMRIM。DMRIM 针对小概率事件风险的无规则等特点,将风险的强度和频度直观地、动态地映射为残余抗原的浓度,以残余抗原刺激免疫记忆、指导抗体进化、控制识别器的生命周期,突破了传统的记忆细胞生命周期,实现了识别器分布自制,提高了小概率事件的辨识能力。仿真实验表明,DMRIM 充分体现免疫记忆的动态性,有效地识别小概率事件,其可行性在实际应用中得到了验证。

关键词 残余抗原,免疫记忆,动态风险识别,小概率事件

中图分类号 TP18 **文献标识码** A

Dynamic Memory Risk Identification Model Based on Residual Antigen

TAO Yuan¹ HU Min² WANG Ping¹

(Computation Center, Shanghai University, Shanghai 200072, China)¹

(Sydney Institute of Language & Commerce, Shanghai University, Shanghai 201800, China)²

Abstract The paper focused on the risk identification of small probability events. A Residual Antigen based Dynamic Memory Risk Identification Model DMRIM was proposed. As small probability events have no rules to follow, the intensity and frequency of risk is intuitively and dynamically mapped to the concentration of residual antigen by DMRIM. By using residual antigen, DMRIM stimulates immune memory, guides antibody evolution and controls the life-cycle of the detector. It breaks through traditional memory cell life-cycle and realizes distribution and self-made of the detector, thereby improving recognition of small probability events. The simulation experiments prove that DMRIM achieves dynamic memory and recognize small probability events effectively. Its feasibility is verified in practical application.

Keywords Residual antigen, Imunne memory, Dynamic risk identification, Small probability

免疫记忆是免疫应答的主要方式之一,在人工免疫系统中具有重要地位,并广泛应用于网络入侵检测、风险识别、异常检测等方面^[1,2]。在人工免疫学界,记忆细胞学说^[3,4]是被广泛采纳的观点,在处理优化问题上更是取得成功应用^[5,6]。一般以记忆细胞的形式存储局部最优解,再以抗体抗原的亲合力等为衡量标准不断地比较这些局部最优解,最终获得全局最优解,非常适合优化问题特点。

作为动态风险中特例的小概率事件,其在具有动态风险随时间变化的特性的同时,还具有无规则性等特点,表现为时间上发生的突变或异常,从而使小概率事件风险识别更加困难。但是,小概率事件广泛地存在于工程、气候、传染病暴发等领域。如何识别小概率事件的突变和异常,是目前研究的热点。例如,文献[7]从动力学结构角度,采用 Fisher 信息检测和识别系统动力学结构所发生的突变。但是,从动力学的角度,突变是系统由一个稳定状态到另一种稳定的状态^[8]。因此,从动力学角度出发很可能不能识别不稳定状态的风险。

生物免疫系统在面临的环境完全是不确定的、时变的,外界入侵物(抗原)的形式几乎是无限的情况下,对已知和未知

的抗原都能出色防御,这种特性使模拟生物免疫系统运行机理的人工免疫系统在识别小概率风险上具有天然优势。但是,人工免疫系统在这方面研究应用还非常少。对于传统的基于记忆细胞学说的免疫记忆方法,若识别器成功地识别了某个模式,则将该识别器作为记忆细胞加以区别对待,以备再次识别相同的模式^[9,10]。免疫记忆通过记忆细胞来维持,这种机制导致记忆是静态的,不善于捕捉小概率事件的特征变化,因此不完全适应小概率事件风险识别。

2004 年 Wilson 首次提出人工免疫系统可以采纳残余抗原学说简化记忆模型,突破了传统采用记忆细胞学说建立记忆模型的方法,不再区分记忆细胞与普通细胞。Wilson 使用了一种基于细胞有丝分裂过程的生命周期控制算法,其原理是:细胞中存在一种称为着丝点(Tolemere)的器官,细胞每次分裂时其长度会缩短,当其长度缩短为 0 时,细胞会因为无法分裂而死亡^[11]。这样,记忆会被大大延长,但是最终仍然会消失,如果应用在小概率事件风险识别中,不能完全适应小概率事件的无规则等特点。本文提出一种基于残余抗原学说的动态记忆风险识别模型 DMRIM (Dynamic Memory Risk

到稿日期:2013-08-18 返修日期:2013-11-13 本文受上海市科委“科技创新行动计划”重点项目(13511504803),上海高校青年教师培养资助计划(ZZSD12060),上海大学创新基金(SDCX2012017)资助。

陶媛(1981-),女,博士,讲师,主要研究领域为智能信息处理和模式识别,E-mail:taoyuan103@shu.edu.cn;胡珉(1970-),女,博士,副教授,硕士生导师,主要研究方向为数据挖掘和建筑信息化;王萍(1975-),女,博士,副教授,主要研究方向为数据挖掘。

Identification Model Based on Residual Antigen)。与 Wilson 不同的是,DMRIM 将小概率事件的风险强度和频度以残余抗原的形式直观地、动态地映射为残余抗原的浓度,通过残余抗原刺激免疫记忆,这样系统记忆的抗原形式是最有价值的,资源的利用率也会提高。

1 模型描述

残余抗原学说的中心思想是:少量抗原被保留在低水平下持续地刺激免疫系统。低浓度抗原的刺激不至于引发免疫应答反应,但是可以使免疫系统一直处于对可能到来的感染的预备状态。如果再次遇到高浓度的经验感染(或相似的结构),免疫系统则处于待命状态。

依据残余抗原理论,低浓度的残余抗原(浓度在安全范围内的残余抗原)持续地刺激免疫系统,但这种刺激不至于引发免疫应答反应,仅仅可以使免疫系统一直处于对可能到来的感染的预备状态。这种残余抗原的刺激水平表示为残余抗原浓度,为了模拟残余抗原刺激的频度和强度的变化,将风险的强度和频度直观地、动态地映射为残余抗原的浓度,适应小概率风险识别的特性。基于残余抗原学说的动态记忆风险识别模型 DMRIM 的核心结构如图 1 所示。

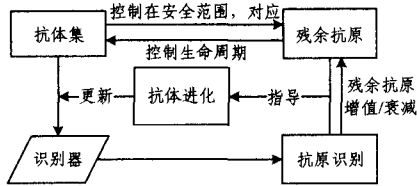


图 1 DMRIM 的核心结构

如图 1 所示,残余抗原刺激免疫记忆,表现为抗体(识别器)的生命周期;通过残余抗原增值和衰减机制将风险的强度和频度直观地、动态地映射为残余抗原刺激水平(通过浓度体现),并指导抗体进化。

抗体 Ab 和残余抗原 Ra 是有对应关系的,定义残余抗原群 RA 与抗体群 AB 的对应关系如式(1)所示:

$$Match = \{ \langle Ab, Ra \rangle \mid Ab \in AB, Ra \in RA \wedge f_m(Ab, Ra) = 1 \} \quad (1)$$

其中, $f_m(Ab, Ra)$ 是 Ab 与 Ra 之间的距离,可以为 Euclidean 距离、Manhattan 距离、Hamming 距离等,如果小于一定的阈值,则成立, Ab 与 Ra 互为抗体和相应的残余抗原。

任何一个抗体 $Ab(Ab \in AB)$ 所涉及的内容可以用一个三元组 $Ab = (p, r, aff)$ 描述,其中 p 为抗体的残余抗原; r 为抗体繁殖的活化度; aff 为抗原识别标准,这里采用抗体与抗原的亲合力。

2 残余抗原刺激

为了模拟残余抗原刺激的频度和强度的变化,将风险的强度和频度直观地、动态地映射为残余抗原的浓度,定义残余抗原增值与衰减机制分别如式(2)与式(3)所示。

$$p(t) = p_0 + k * p(t-1) \quad (2)$$

其中, $p(t)$ 为 t 时刻残余抗原的浓度; $p(0) = p_0$ 表示一直处于增值活动中的起始浓度,为残余抗原初始浓度,是常数; k 为增值系数, $0 < k < 1$ 。

$$p'(t') = p'(t'-1) - k' * p_0' \quad (3)$$

其中, p_0' 表示抗原一直没有增值活动的起始浓度; k' 为衰

系数, $0 < k' < 1$; t' 表示抗原一直没有增值活动的起始浓度的最初时刻到当前所用的时间。

定理 1 设每次被提呈给抗体群的抗原为风险抗原,识别器都匹配抗原成功,表明此类威胁在持续增加,其相应残余抗原的浓度是递增的,且有最大上限,是有限值。

证明:设每次被提呈给抗体群的抗原为风险抗原,识别器都匹配抗原成功,相应残余抗原处于增值操作,则有:

$$p(0) = p_0, p(1) = p_0 + k * p_0$$

$$p(2) = p_0 + k * p(1) = p_0 + k * (p_0 + k * p_0)$$

...

$$p(t) = p_0 + k * p(t-1)$$

迭代得:

$$p(t) = p_0 + k * p_0 + \dots + k^{t-1} * p_0$$

$$= p_0(1 + k + \dots + k^{t-1})$$

$$= p_0(1 - k^t) / (1 - k)$$

$\therefore 0 < k < 1$

$$p(t) - p(t-1) = p_0(k^t - k^{t-1}) / (1 - k) > 0$$

$p(t)$ 为增函数

$\therefore \lim_{t \rightarrow \infty} p(t) = p_0(1 - k^t) / (1 - k) = p_0 / (1 - k)$ 为有限值。

由此得证,每次被提呈给抗体群的抗原为风险抗原,识别器都匹配抗原成功,其相应残余抗原的浓度是递增的,且有最大上限,即是有限值,且最大值与初始浓度有关。定理 1 确保了 DMRIM 在使抗原残留下来的同时,也可将残余抗原浓度控制在安全范围内,避免残余抗原的再次伤害。

定理 2 设每次被提呈给抗体群的抗原为风险抗原,识别器都匹配抗原失败,其残余抗原的浓度是递减的,且在有限时间内递减为 0。

证明:设每次被提呈给抗体群的抗原为风险抗原,识别器都匹配抗原失败,相应残余抗原处于衰减操作,则有:

$$p'(0) = p_0', p'(1) = p_0' - k' * p_0'$$

$$p'(2) = p'(1) - k' * p_0' = p_0' - 2 * k' * p_0'$$

...

$$p'(t') = p'(t'-1) - k' * p_0'$$

迭代得:

$$p'(t') = p_0' - k' * p_0' - \dots - k' * p_0' = p_0' - k' * t' * p_0'$$

$\therefore 0 < k' < 1$

$$p'(t') - p'(t'-1) = -k' * p_0' < 0$$

$\therefore p'(t')$ 为减函数

当 $p'(t') = p_0' - k' * t' * p_0' = 0$ 时

$t' = 1/k'$ 为有限值

由此得证,当每次被提呈给抗体群的抗原为风险抗原,识别器都匹配抗原失败,其残余抗原的浓度是递减的,且在有限时间内递减为 0。

定理 1 和定理 2 证明了残余抗原增值与衰减机制能确保大部分抗原物质被免疫系统清除引起免疫应答的同时,少量抗原物质得以繁殖,并且残余的少量抗原保留在低水平下,使其浓度控制在安全范围内,也就是浓度有最大安全上限。一旦浓度过高达到安全浓度最大上限,即为风险,必须采取措施将残余抗原浓度控制在安全范围内。

3 残余抗原控制的识别器

通过残余抗原刺激控制识别器的生命周期,这种激活状态表现为识别器的一个生命周期。这样,识别器的生命周期

是分布式的,残余抗原控制的识别器生命周期如图 2 所示。

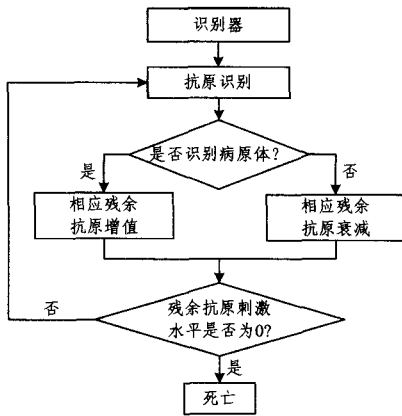


图 2 残余抗原控制的识别器生命周期

具体流程如下:

Step 1 如果识别器在一次识别过程中识别出病原体,则其相应的残余抗原浓度增值,否则其相应的残余抗原浓度衰减。

Step 2 如果识别器的残余抗原的浓度为 0,识别器死亡,否则进入下一轮识别。

由上可见,识别器(抗体)与残余抗原两者相互影响。从识别器角度出发,具体表现为:识别器识别出病原体→相应残余抗原增值→相应残余抗原浓度增加→识别器寿命增加;识别器未识别出病原体→相应残余抗原衰减→相应残余抗原浓度减少→识别器寿命缩短。从残余抗原角度出发,具体表现为:残余抗原浓度增加→相应识别器受到刺激变大→相应识别器的寿命增加;残余抗原浓度衰减→相应识别器受到刺激变小→相应识别器的寿命缩短。这一过程不断重复,残余抗原控制的识别器生命周期突破传统的记忆细胞生命周期,具有较高的自动性和适应性,实现分布自制。

4 抗体进化

抗体进化包含繁殖、变异和优选 3 步。

4.1 繁殖

当抗体 $Ab(Ab \in AB)$ 的 $Ab.r > r_{threshold}$ (繁殖阈值) 时,抗体可以进入繁殖, r 的标准可以根据实际情况而定。鉴于 DMRIM 针对小概率事件风险的识别,这里 r 由抗体的残余抗原浓度和抗体与抗原的亲合力两方面决定,如式(4)所示:

$$Ab.r = Ab.aff * e^{\delta Ab.p} \quad (4)$$

其中, δ 起到调节作用。抗原的攻击频度越大或是强度越大,抗体对该类抗原的活化度越大。

这里,抗体繁殖有以下两种方式。

4.1.1 克隆

在免疫算子中,克隆最常被用于抗体的繁殖。为了克隆数目易于控制,可以采用式(5),根据亲和力由高到低排序后的抗体在序列中的位置计算其克隆数目 N_c^i 。

$$N_c^i = \left\lceil \frac{P_i [nPop * b - (nPop - 1)] + (nPop - 1)}{nPop} \right\rceil \quad (5)$$

其中, $P_i = aff_i / \sum_{k=1}^{nPop} aff_k$, $nPop$ 是抗体群大小, b 是定义的最好抗体的最大克隆数。依据式(5), N_c^i 与 P_i 成正比(即与其亲和力成正比),且 $1 \leq N_c^i \leq b$ [12]。

4.1.2 细胞分裂

细胞分裂是活细胞繁殖其种类的过程,是一个细胞分裂

为两个细胞的过程。抗体的繁殖也可以模拟细胞分裂。抗体获得分裂机会时,则分裂为两个子代细胞。

4.2 变异

抗体变异后,形成新的识别能力,使得免疫系统具有更高的适应性。变异算是免疫算子中的一个重要成员,有很多种方法,可以根据实际情况选择。这里定义变异概率如式(6)所示:

$$g = \beta \exp(-1/aff(Ab, Ag)) \quad (6)$$

其中, β 起到调节变异概率的作用, g 随着亲和力的增大而减小。

4.3 优选

一般在优选方面,根据模型或算法的出发点和应用对象,选择优选标准,这里依据变异后亲和力和不降低原则。如果无子代生成,则继续保留父代;否则基于子代与父代“相似程度”继承父代的残余抗原浓度,并初始新抗体的活化度,如式(7)所示。

$$f_{in} = \{Ab_{n1}, \dots, Ab_{nk} \mid Ab_{ni}.p = f_{ni} * Ab_{fa}.p \wedge \sum_{i=1}^{k-1} Ab_{ni}.p = Ab_{fa}.p \wedge Ab_{ni}.r = 0\} \quad (7)$$

其中, k 为父代繁殖子代个数, f_{ni} 表示依据 f_{in} 标准第 i 个子代继承父代残余抗原的系数。例如, $f_{ni} = aff_{ni} / \sum_{i=1}^{k-1} aff_{ni}$,亲和力和 aff_{ni} 可以是 Euclidean 距离、Manhattan 距离、Hamming 距离等。

5 仿真实验与结果分析

初始抗体群 AB 和相应的残余抗原群 RA ,依据抗体 Ab 的残余抗原浓度 $Ab.p = p_0$,及初始抗体 Ab 的活化度 $Ab.r = 0$ 。 $Ab.aff$ 用欧几里德距离表示。抗体的繁殖采用文中提到的分裂方法。

5.1 DMRIM 记忆的动态性

选取 Breast Cancer data set of UCI [13] 作为实验测试集,该数据集的 10 个属性中包含 9 个固定属性和一个类别属性。依据类别属性,该数据集分两类,包含良性的 458 条,恶性的 241 条。选取 Breast Cancer 数据集的部分数据作为 5 组实验的实验样本,即抗原集,测试 DMRIM 记忆的动态性,如图 3 所示。

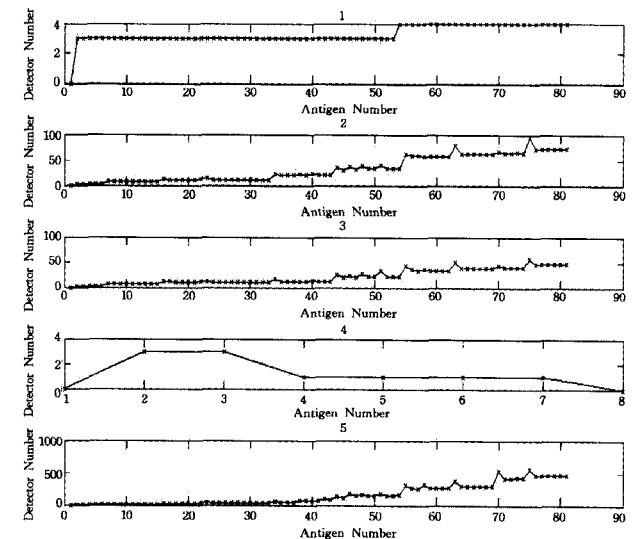


图 3 DMRIM 的动态性实验结果

图 3 反映了识别器数目随抗原输入而变化的情况,其中:横坐标表示抗原序列,序号代表了输入抗原的个数;纵坐标表

示识别器数目。从图 3 可以看出,每个子图的识别器数目的总体变化趋势相似,这是因为 5 组实验的实验样本相同,抗原对识别器的刺激情况相同;识别器数目的具体变化情况不同,这是因为每组实验的参数 $r_{threshold}$ 、 p_{max} 和 k'/k 等不同,导致残余抗原浓度刺激水平不同。实验证明了 DMRIM 的识别器群体数目随输入抗原对其刺激水平而动态变化。

5.2 DMRIM 识别性能的测试

5.2.1 无缺失属性的数据样本实验

选取 Breast Cancer 中无缺失属性的 683 条数据作为实验样本,采用 10 次交叉验证,并将 DMRIM 与已知的比较著名的 CSA^[13] (Clonal Selection Algorithm) 和 DCA^[14] (Dendritic Cell Algorithm) 进行对比分析,实验结果如表 1 所列。

表 1 无缺失属性的数据样本的识别精度

实验	DMRIM			CSA			DCA		
	TR	TPR	TNR	TR	TPR	TNR	TR	TPR	TNR
1	96.72%	93.55%	100%	80.2%	97.8%	65.7%	91.03%	90.5%	93.1%
2	92.97%	91.3%	100%	96.9%	99.02%	92.8%	92.1%	95.8%	87.5%
3	96.92%	95.12%	100%	90.02%	81.88%	97.4%	96.23%	95.65%	100%
4	96.94%	95.8%	91.5%	91.46%	88.13%	98.02%	88.89%	78.79%	97.44%
5	94.2%	94.12%	94.23%	90.24%	89%	91.32%	96.23%	97.83%	85.71%
6	92.94%	94.2%	87.5%	94.11%	91.03%	95.47%	86.79%	84.78%	100%
7	98.82%	100%	95.65%	83.33%	87.8%	87.5%	95.08%	96.67%	93.55%
8	97.56%	100%	91.3%	90.28%	90.91%	89.74%	86.79%	84.78%	100%
9	97.4%	96.97%	100%	83.33%	90.91%	76.92%	88.52%	93.55%	83.33%
10	97.62%	97.14%	100%	87.75%	72.73%	100%	88.52%	88.89%	88.14%
平均值	96.21%	95.82%	96.01%	88.76%	88.92%	89.49%	91.02%	90.74%	92.88%

从 10 次交叉得到的总的识别精度来看,DMRIM 和 DCA 的识别精度都达到了 90% 以上,其中:DMRIM 识别精度为 96.21%,DMRIM 的正向与反向的识别精度分别为 95.82% 和 96.01%,明显高于 CSA 和 DCA 的识别精度。可见 DMRIM 的识别效果良好。

5.2.2 包含缺失属性的数据样本实验

在工程领域中,对数据进行采集时,得到的数据样本很可能包含缺失属性的数据,因此,研究对包含缺失属性的数据样本的分类性能是非常具有实用意义的。从表 2 可以看出:当正向数据与反向数据比例不均时,占大比例的数据样本主导总的识别精度,为了避免这种情况,采用查全率(Recall)和查准率(Precision)的组合作为评价算法分类性能的标准。正向数据的查全率和查准率用式(11)和式(12)计算:

$$Recall(P) = TP / (TP + FN) \quad (11)$$

$$Precision(P) = TP / (TP + FP) \quad (12)$$

反向数据的查全率和查准率分别用式(13)和式(14)计算:

$$Recall(N) = TN / (TN + FP) \quad (13)$$

$$Precision(N) = TN / (TN + FN) \quad (14)$$

其中,TPR(True Positive Rate)是指正向数据识别精度,即正向数据确实识别为正向的比例,TNR(True Negative Rate)是指反向数据识别精度,即反向数据确实识别为反向的比例。TPR 与 TNR 用式(8)和式(9)计算,识别精度 TR 用式(10)计算。

$$TPR = TP / (TP + FP) \quad (8)$$

$$TNR = TN / (TN + FN) \quad (9)$$

其中,TP 为正向数据的正确识别数;FP 为正向数据的错误的识别数;TN 为反向数据的正确识别数;FN 为反向数据的错误的识别数。

$$TR = (TP + TN) / (TP + FP + TN + FN) \quad (10)$$

其中,TP+FP+TN+FN 是实验样本的总数。

其中,TP 为正向数据的正确识别数;FP 为正向数据的错误的识别数;TN 为反向数据的正确识别数;FN 为反向数据的错误的识别数。

正向数据识别准确率 P_value 和反向数据识别准确率 N_value 用式(15)和式(16)计算:

$$P_value = \frac{(1 + \beta^2) Recall(P) * Precision(P)}{\beta^2 * Recall(P) + Precision(P)} \quad (15)$$

$$N_value = \frac{(1 + \beta^2) Recall(N) * Precision(N)}{\beta^2 * Recall(N) + Precision(N)} \quad (16)$$

其中, $\beta=1$ 。识别精度只作为查全率和查准率评价标准的一部分。这样,只有在查全率和查准率都高的情况下,识别准确率才可能高。也就是说,在两类数据比例相差较大的情况下式(15)和式(16)可更全面地评价识别结果。

选取 Breast Cancer 数据集中无缺失属性的 300 条数据与这 16 条缺失属性的数据作为实验样本,用 5 次交叉验证。DMRIM、CSA 和 DCA 均采用平均值方法处理缺失属性的数据样本。将 DMRIM 与 CSA 和 DCA 进行对比分析,实验结果如表 2 所列。其中,TMR 是指 16 条缺失属性的数据的识别精度。

表 2 包含缺失属性的数据样本的实验

实验	DMRIM			CSA			DCA		
	P_value	N_value	TMR	P_value	N_value	TMR	P_value	N_value	TMR
1	95.27%	93.02%	66.67%	78.2%	100%	37.7%	86.67%	90.48%	66.67%
2	98.47%	97.63%	100%	89.55%	90.91%	0%	97.78%	87.75%	37.7%
3	100%	100%	100%	83.33%	83.33%	66.67%	91.76%	66.67%	100%
4	90.2%	90.91%	66.67%	84.21%	89.66%	37.7%	97.83%	85.71%	37.7%
5	98.9%	93.33%	100%	56%	86.42%	100%	66.67%	91.74%	66.67%
平均值	96.57%	94.98%	86.67%	78.26%	90.05%	48.41%	88.14%	84.47%	61.75%

从表 2 可以看出:DMRIM 对于包含缺失属性的数据样

本,在正向数据识别准确率和反向数据识别准确率上都取得

了较好的效果,明显优于 CSA 与 DCA;对缺失属性的数据的识别精度也高于 CSA 与 DCA。

5.3 DMRIM 对小概率事件的识别测试

DMRIM 可以通过对残余抗原浓度最大上限 p_{max} 和残余抗原衰减/系数关系 k'/k 的设置,调节识别器的残余抗原浓度增加与衰减的速度;通过对分裂阈值 $r_{threshold}$ 的设置,控制识别器的分裂速度,以延长小概率事件识别器的生命周期,达到有效地识别小概率事件的目的。例如:

Step1 如果识别器 Ab 与抗原的亲合力大于 $aff1_{threshold}$,则 Ab 的残余抗原衰减系数取 k_1' ,残余抗原浓度最大上限取 $p1_{max}$,分裂阈值取 $r1_{threshold}$,否则,转 Step 2。

Step2 如果识别器 Ab 与抗原的亲合力大于 $aff2_{threshold}$,则 Ab 的残余抗原衰减系数取 k_2' ,残余抗原浓度取最大上限 $p2_{max}$,分裂阈值取 $r2_{threshold}$,否则,识别失败进入残余抗原衰减操作。

其中, $aff1_{threshold} > aff2_{threshold}$, $k_1' < k_2'$, $p1_{max} > p2_{max}$, $r1_{threshold} < r2_{threshold}$ 。

选取 Iris data set of UCI 作为实验测试集,该数据集的 5 个属性中包含 4 个固定属性和 1 个类别属性,共 150 条。Iris 的类别属性值分为 Iris-setosa, Iris-versicolor 和 Iris-virginica 3 类,依据本次实验目的,对 Iris 数据集进行抽样得到数据样本,如表 3 所列。

表 3 数据样本的分布情况

类别	Iris-setosa	Iris-versicolor	Iris-virginica
样本数	40	50	10
比例	40%	50%	10%

DMRIM 对 Iris 数据集的实验结果如表 4 所列。

表 4 实验结果

评价标准 \ 类别	Iris-setosa	Iris-versicolor	Iris-virginica
P_value	96.3%	97.96%	95.24%

从表 4 可以看出,DMRIM 对 Iris-setosa、Iris-versicolor 和 Iris-virginica 的识别准确率都很高。实验证明,DMRIM 能有效地识别小概率事件(Iris-virginica)。

5.4 DMRIM 在盾构施工进出洞风险监测中的应用

盾构进出洞过程存在着大量的不确定和不可预见的风险因素。我们选取了长江隧道工程出洞阶段中采集到的土体压力监测数据并采用 DMRIM 进行风险识别的应用实验。长江隧道分为上行线和下行线。在施工过程中,上行线已经出现了明显的风险特征,事后专家经分析给出了可能出现的风险时间点 2008 年 5 月 18 日 6:30 AM(数据 246 条)。我们抽取 2008 年 5 月 12 日 9:46 到 2008 年 5 月 28 日 10:00 区间内上行线施工所采集的 684 条土压力数据进行实验。

土压力的数据采集是通过多个测点来进行测量的。对数据做了如下预处理:将土压力数据按照时间顺序排列,每一时间点的数取该点测量时的最大值、最小值和平均值,并进行归一化处理。这样,经过预处理后输入数据为 3 维数据(即每条数据包含该时间点的最大、最小和平均值)。DMRIM 的实验结果如表 5 所列。

表 5 实验结果

	TR	P_value	N_value
DMRIM	92.5%	92.54%	88.43%

从表 5 可以看出,DMRIM 取得了良好的识别准确率。不同于标准工程问题,实际工程应用问题的风险状态不会以高频率转换,因此个别风险数据的出现不足以说明风险发生。DMRIM 对土压力数据的风险识别如图 4 所示,在图 4 中,为了方便理解识别结果,工程状态用紫红色点表示,其中,0 表示安全,-1 表示风险,状态切换在虚线连接点处。识别结果用黑色点表示,其中,0 表示安全,1 表示风险,状态切换在实线连接的点处,横坐标的序号代表一个时间点。

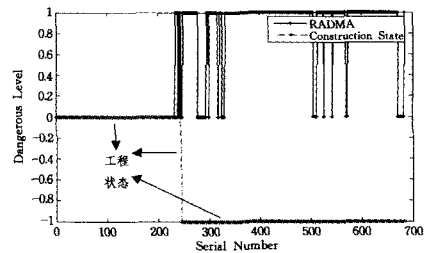


图 4 DMRIM 对土压力的风险识别

如图 4 所示,DMRIM 在图 4 中第 234 个点(对应 5 月 17 日 12 点)首先给出了风险状态,之后密集而持续地给出风险状态。这说明 DMRIM 给出的风险点是 5 月 17 日 12 点。上述实验证明了 DMRIM 在实际应用中的可行性。

结束语 目前小概率事件风险识别方法相对缺乏,主要是从管理角度出发,或是基于随机理论,依赖历史资料数据,容易陷入规则化或概率化,无法充分考虑到小概率事件的非规则性等特点。本文针对小概率事件的风险特点,提出一种基于残余抗原学说的动态记忆风险识别模型 DMRIM。DMRIM 突破传统记忆细胞记忆模型的模式,通过建立残余抗原增值/衰减机制赋予残余抗原生命,将风险的强度和频度直观地、动态地映射为残余抗原的浓度,由残余抗原刺激控制识别器的生命周期,指导识别器的进化。实验证明了模型记忆的动态性和识别性能,以及实际应用中的可行性。针对 DMRIM 在实际应用中的表现,需要开展进一步的研究。

参考文献

- [1] Dasgupta D, Yu Sen-hua, Nino F. Recent Advances in Artificial Immune Systems: Models and Applications [J]. Applied Soft Computing, 2011, 11(2): 1574-1587
- [2] Boukerche A, Juca K R L, Sobral J B, et al. An artificial immune based intrusion detection model for computer and telecommunication systems [J]. Parallel Computing, 2004, 30(5/6): 629-646
- [3] 李涛,刘赛,苏贝. 人工免疫记忆模型研究 [J]. 计算机技术与发展, 2006, 16(7): 29-31
- [4] Jerne N K. The immune system [J]. Scientific American, 1973, 229(1): 51-60
- [5] Tan K C, Goh C K, Mamun A A. An evolutionary artificial immune system for multi-objective optimization [J]. European Journal of Operational Research, 2008, 187(2): 371-392
- [6] 尚荣华,焦李成,公茂果,等. 免疫克隆算法求解动态多目标优化问题 [J]. 软件学报, 2007, 11(18): 2700-2711
- [7] 蔡舒平,戴理,倪溟. 基于 Fisher 信息理论的突变检测新方法 [J]. 物理学报, 2013, 62(18): 1-9
- [8] 颜鹏程,侯威,胡经国. 基于 Logistic 模型的均值突变时间序列临界预警研究 [J]. 物理学报, 2012, 61(18): 1-8

[9] Castro P A D, Zuben F J V. Multi-objective Feature Selection Using a Bayesian Artificial Immune System[J]. International Journal of Intelligent Computing and Cybernetics, 2010, 3(2): 235-256

[10] Ciccazzo A, Conca P, Nicosia G, et al. An Advanced Clonal Selection Algorithm with Ad Hoc Network-Based Hypermutation Operators for Synthesis of Topology and Sizing of Analog Electrical Circuits[C]// Artificial Immune Systems-7th International Conference, ICARIS 2008, Proceedings, 2008. Tiergartenstrasse 17, Heidelberg, D-69121, Germany; Springer Verlag, 2008; 60-70

[11] Wilson W O, Garrett S M. Modelling Immune Memory for Pre-

diction and Computation[C]// 3rd International Conference in Artificial Immune Systems, ICARIS2004, 2004. Catania, Sicily, Italy; Springer, 2004; 386-399

[12] 陶媛, 吴耿锋, 胡琨. 一种基于进化与免疫的动态多目标人工免疫系统模型[J]. 计算机科学, 2010, 37(1): 217-22

[13] De Castro L N, Von Zuben F J. Clonal selection algorithm with engineering applications[C]// GECCO 2002-Workshop Proceedings, 2000. Las Vegas; Proc. of GECCO'00, 2000; 36-37

[14] Aickelin U, Greensmith J. Sensing danger: Innate immunology for intrusion detection[J]. Information Security Technical Report, 2007, 12(4): 218-227

(上接第 184 页)

图以 $|V|$ 为横坐标, 时间(秒)为纵坐标。

实验结果如图 9、图 10 所示。

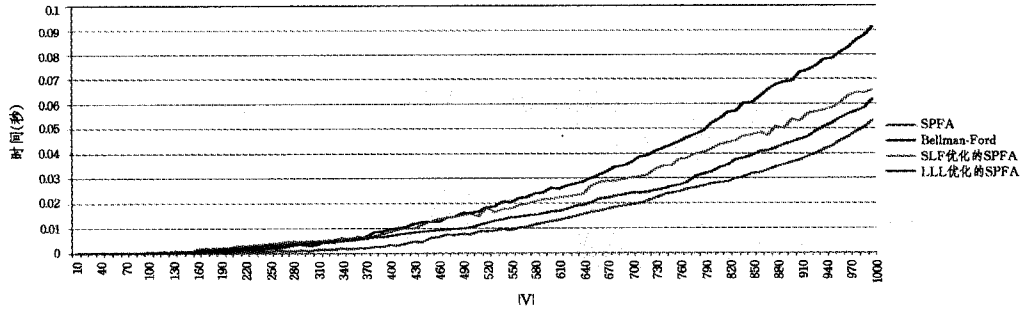


图 9 无源点可达负圈

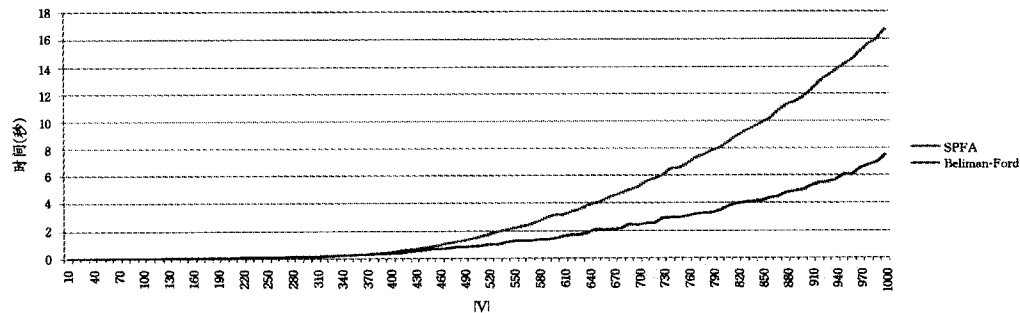


图 10 有源点可达负圈

结束语 从实验结果中我们可以明显看出: 无论是稀疏图, 还是稠密图, 当图中不包含负圈的时候, SPFA 算法的实际运行时间都要好于经过简单优化的 Bellman-Ford 算法; 当图中存在负圈的时候, 则 SPFA 算法的实际运行时间要差于经过简单优化的 Bellman-Ford 算法。经过 SLF 或 LLL 优化的 SPFA 算法并不能始终对 SPFA 算法进行优化, 甚至最坏情况复杂度为指数级, 远远高于普通的 SPFA 算法, 因此在一般情况下并不会使用这些优化。在实际情况中, 我们一般只能事先知道我们所处理的问题是否会有负边出现的情况, 而无法知道若出现负边, 是否会有可能出现负圈。如果我们事先知道了所处理的情况不会出现负边, 虽然 SPFA 算法这时会优于 Bellman-Ford 算法, 但可以采用性能更优的 Dijkstra 算法; 若我们事先只知道有可能会出现负边, 但不能保证是否一定不会出现负圈, 那么这时 SPFA 算法并不会比 Bellman-Ford 算法更有优势。

参考文献

[1] 段凡丁. 关于最短路径的 SPFA 快速算法[J]. 西南交通大学学报, 1994, 29(2): 207-212

[2] Cherkassky B V, Goldberg A V, Radzik T. Shortest paths algorithms: Theory and experimental evaluation [J]. Mathematical Programming, 1996, 73(2): 129-174

[3] 姜碧野. SPFA 的优化与应用[C]// 国家集训队 2009 年论文集. 2009

[4] Cormen T H, Leiserson C E, Rivest R L, et al. Introduction to Algorithms (Third Edition) [M]. MIT Press, Cambridge, MA, 2009; 651-655

[5] Chen Z L, Powell W B. A note on bertsekas' small-label-first strategy [J]. Networks, 1997, 29: 111-116