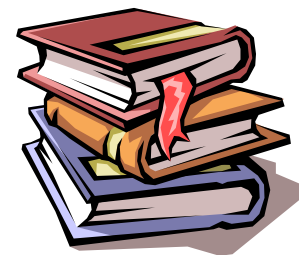


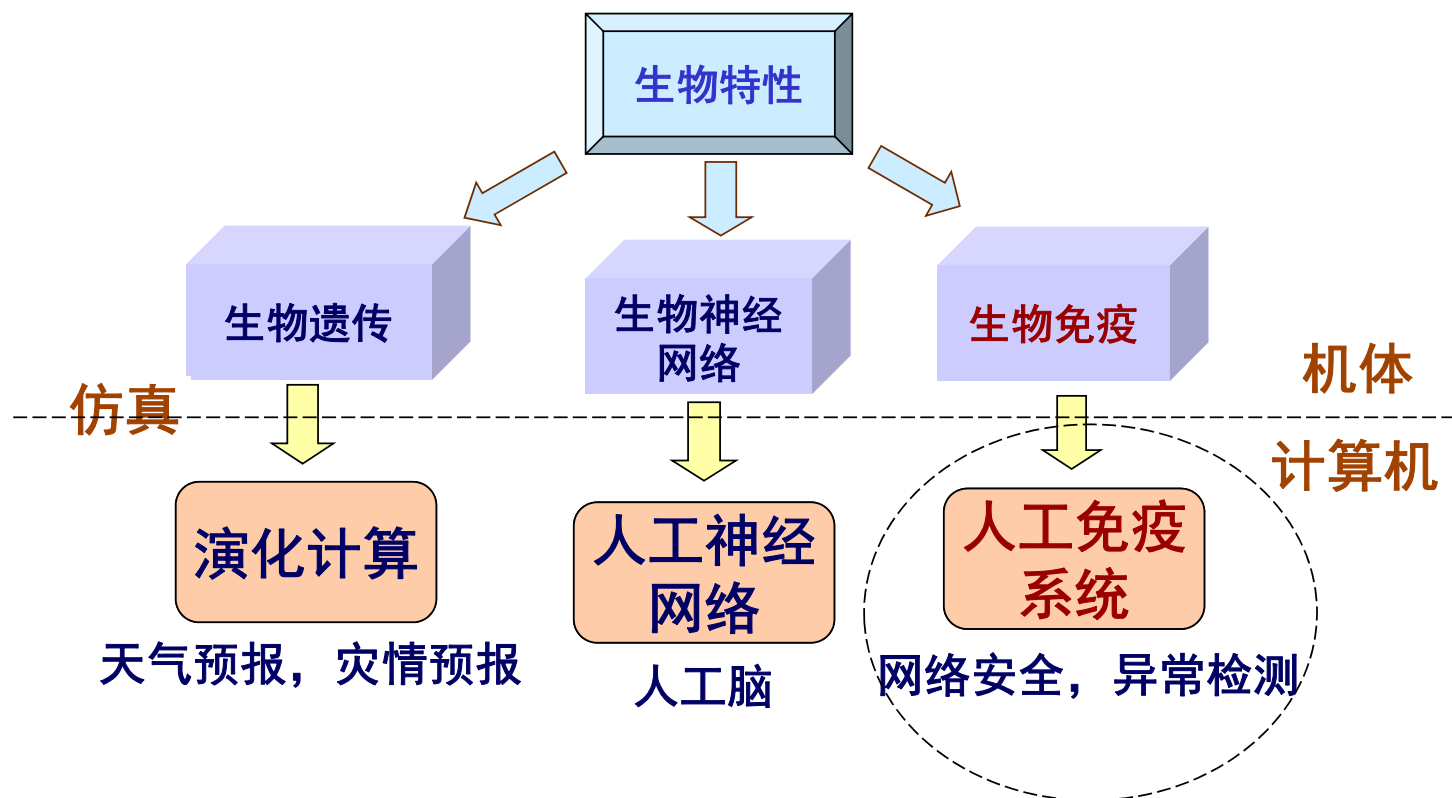
6.11 免疫算法

- 人工智能中的仿生技术
- 什么是免疫
- 人工免疫学
- 相关的免疫算法



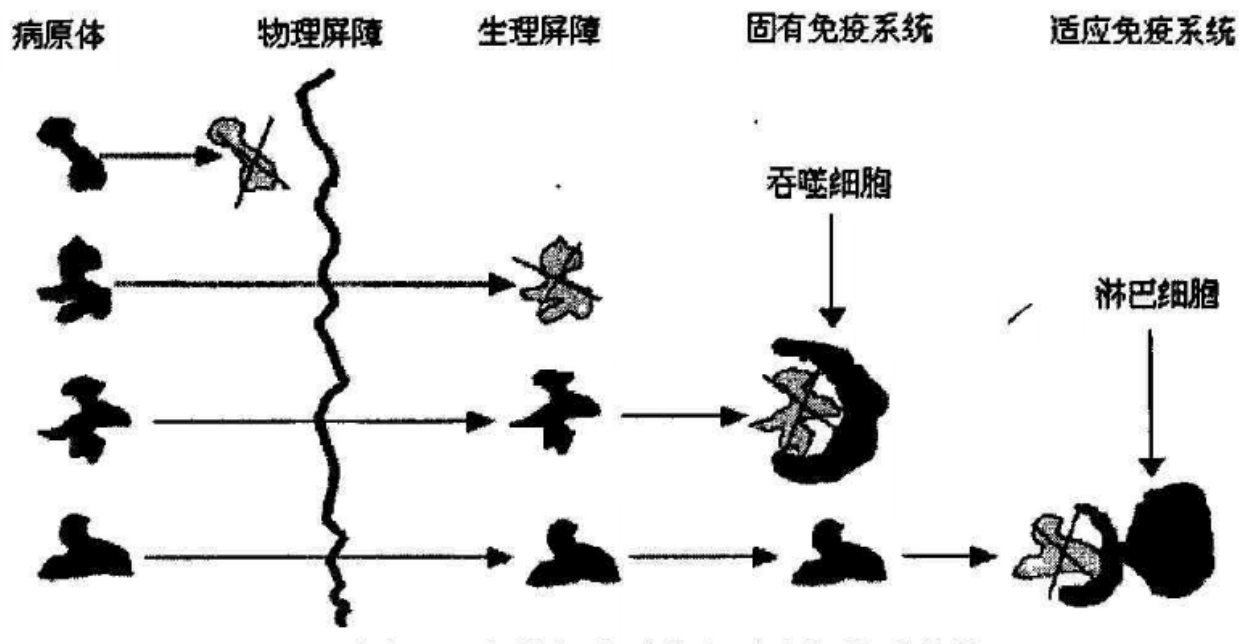
人工智能中的仿生技术

用人工的方法和技术，模仿、延伸和扩展生物智能, 实现机器智能。---生物仿真技术(自然计算)



生物免疫系统概述

免疫系统是生物体一个高度进化、复杂的功能系统，它能自适应地识别和排除侵入机体的抗原性异物，并且具有学习、记忆和自适应调节能力，维护机体内环境的稳定。



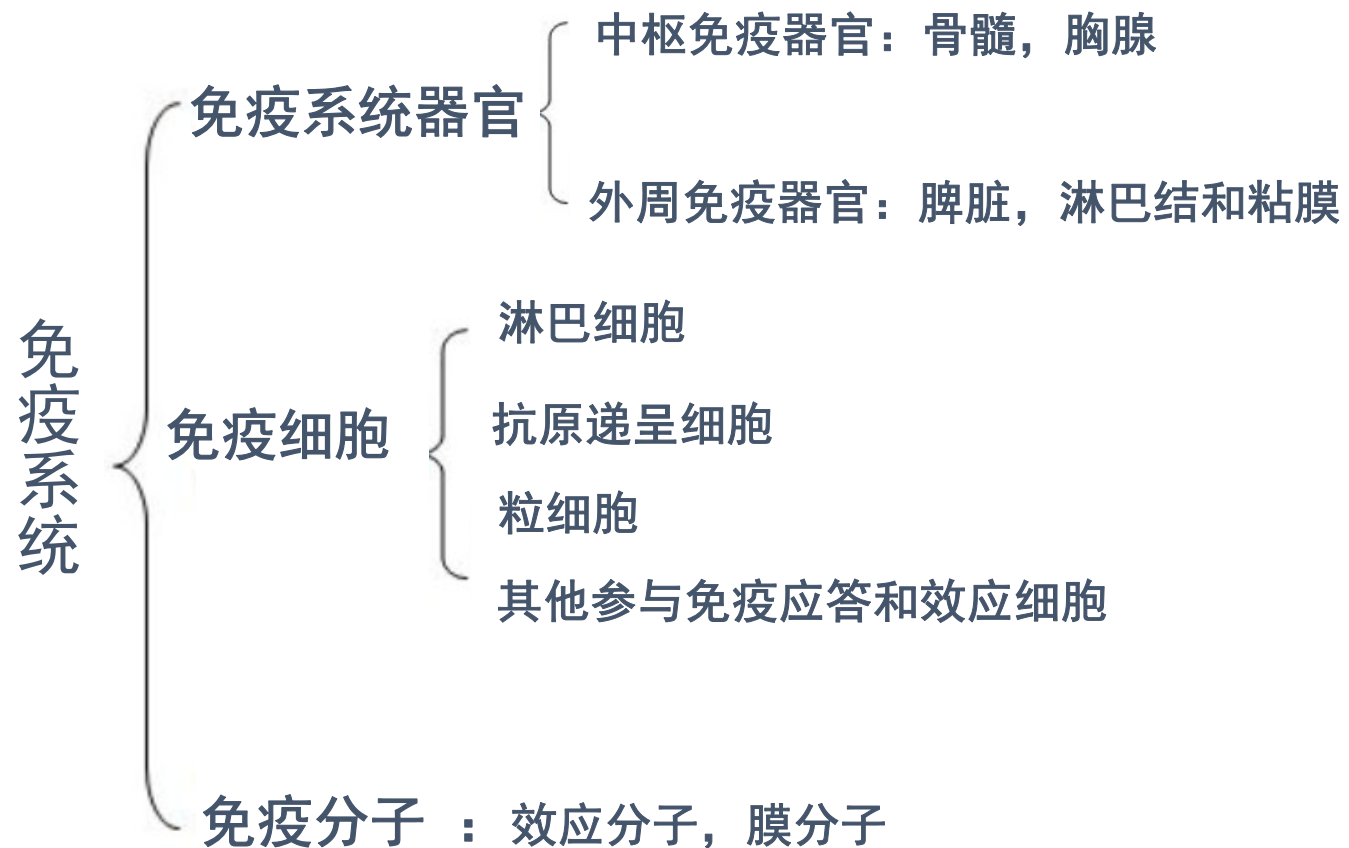
生物免疫系统概述

固有免疫系统和自适应免疫系统比较

类型	特征	细胞	分子
固有免疫	快速应答	噬菌细胞	细胞因子 补体
	低特异性	杀伤细胞	
	无记忆	巨噬细胞	
	开始很慢		
自适应免疫	高度特异 记忆	T 细胞和 B 细胞	抗体

生物免疫系统概述

免疫系统的组成



生物免疫系统概述

免疫系统的功能

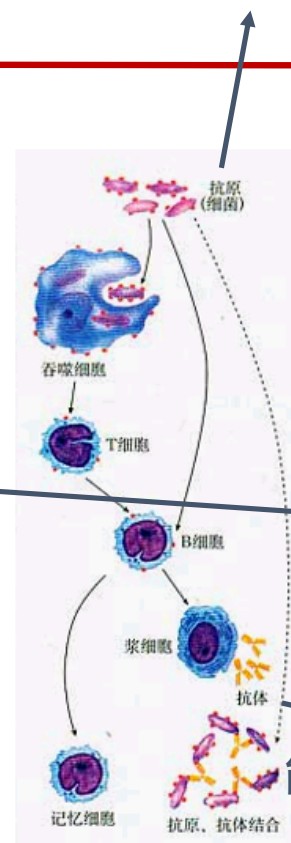
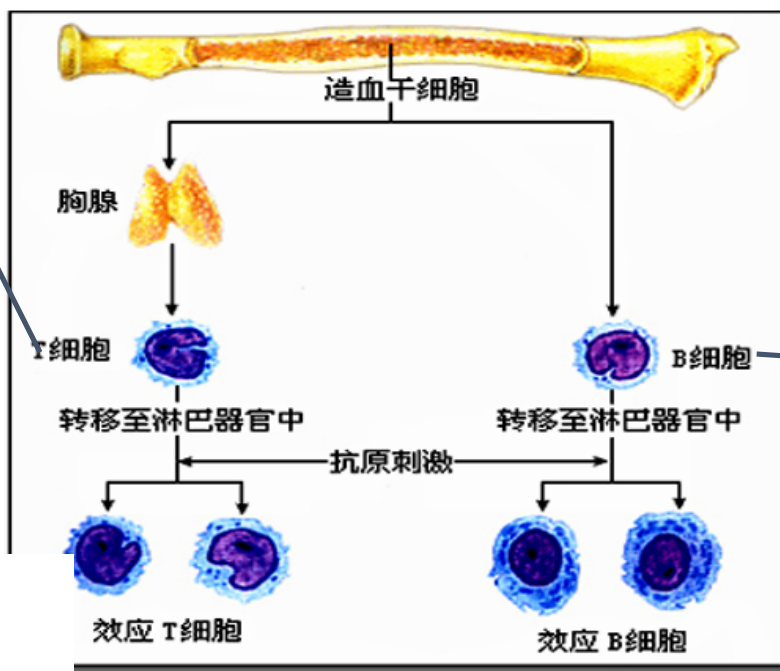
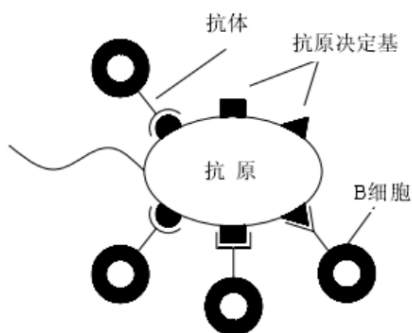
- 免疫防御：机体抵抗其他生物的侵入，抑制其在体内繁殖，扩散，并且排除外来微生物及其产物，保护机体生存的功能。
- 免疫自稳：清除机体内变性、损伤、衰老的细胞，防止形成自身免疫的功能。
- 免疫监视：识别、杀伤和清除体内的突变细胞，防止肿瘤发展的功能。

相关概念

是一种病原体被巨噬细胞等吞噬后产生的带有特异信号的蛋白质片段，能刺激免疫系统产生抗体或形成致敏淋巴细胞。

清除病毒或癌细胞通过刺激或抑制B细胞产生抗体来调节免疫系统平衡。

免疫调节：保证机体免疫平衡。



免疫记忆：免疫系统将侵入抗原反应部分抗体作为记忆细胞保留下来。

产生抗体，提呈抗原以及分泌浆细胞和记忆细胞参与调节。

能够识别并结合其他特异蛋白质。

抗原识别：通过抗原分子表面的表位和抗体分子的对位的化学基进行相互匹配来完成。

生物免疫系统概述

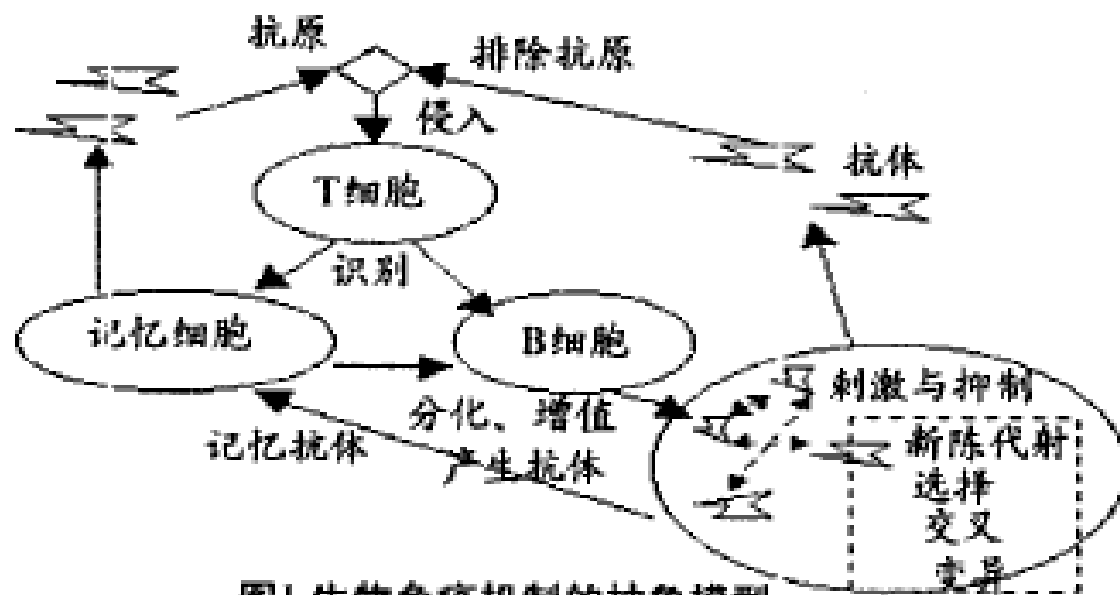


图1 生物免疫机制的抽象模型

当抗体数量>某个阈值时，产生抗体的细胞分化为：

- 抑制细胞- 阻止这种抗体的进一步增加
- 记忆细胞- 将此抗体对应的解记为局部最优解

免疫系统的主要特性

耐受性 分布性 学习性 自治性 多层性

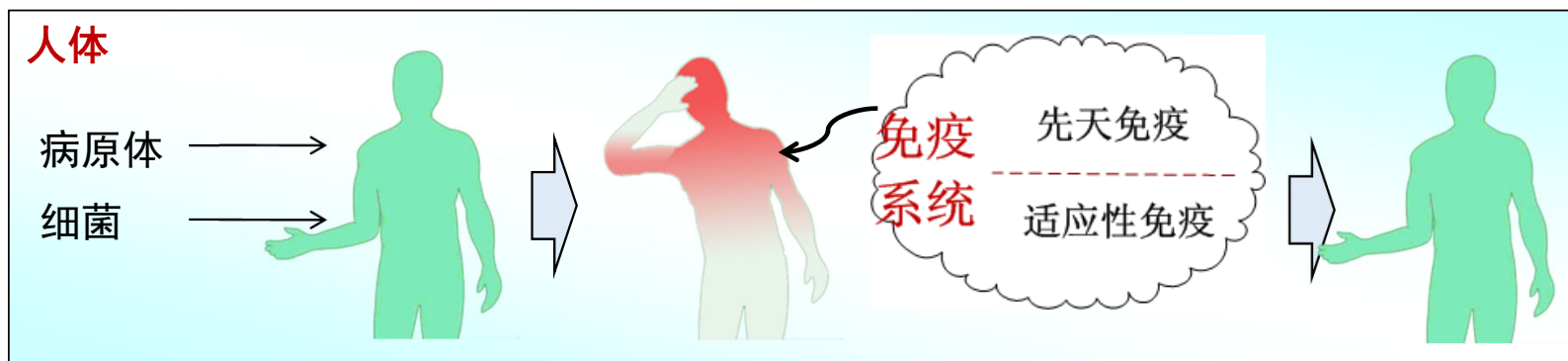
适应性 自体调节性 多样性 动态覆盖性

人工免疫学

计算机系统与生物系统的相似性

- 生物免疫系统：面临大量病毒病菌；计算机系统：面临大量木马、病毒和攻击等。
- 生物免疫系统：目的是保护生物本身、消灭病毒；计算机系统：保护合法的和正常的用户行为，消灭非法的和异常的行为。
-

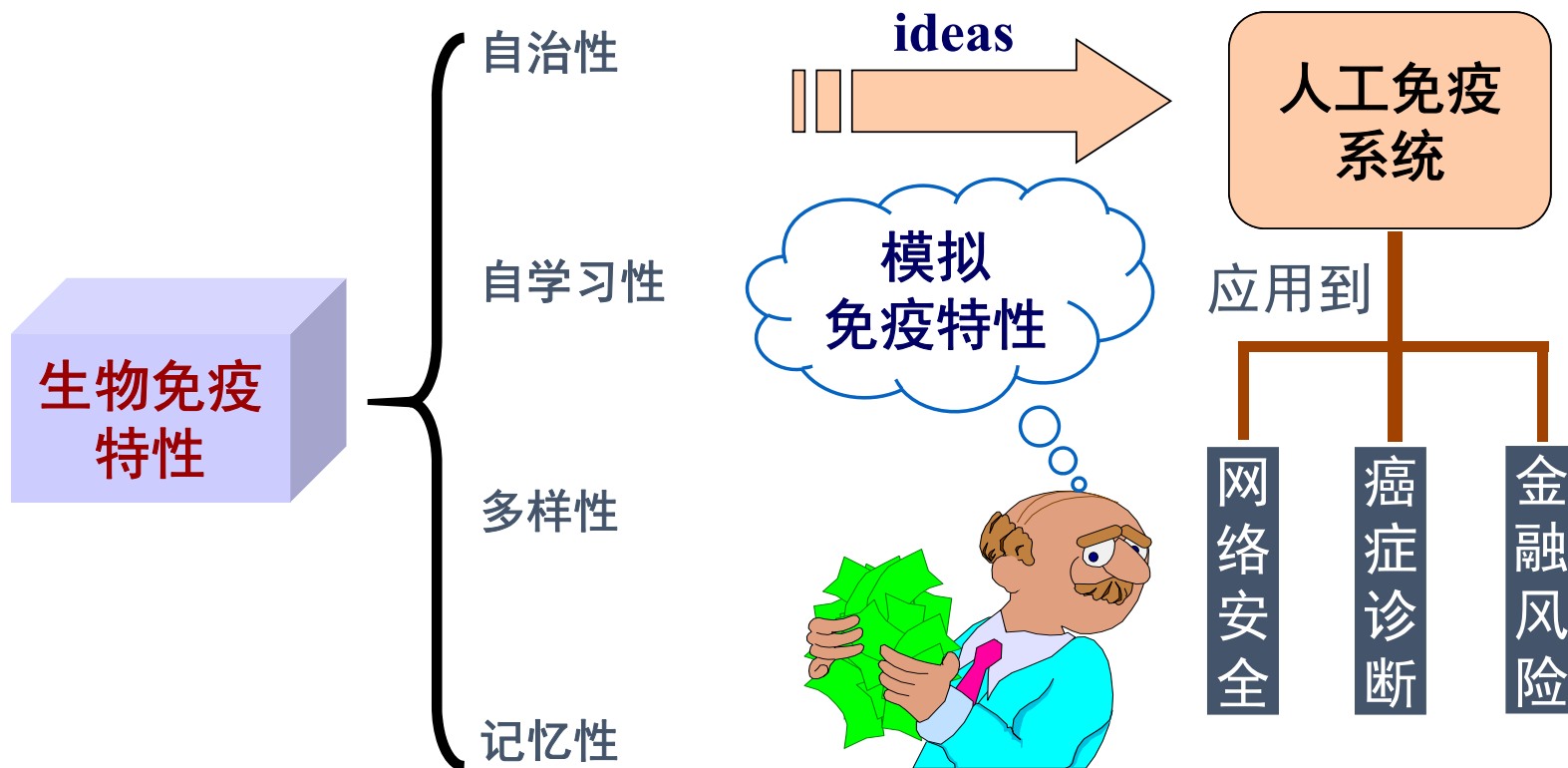
人工免疫学--- ideas



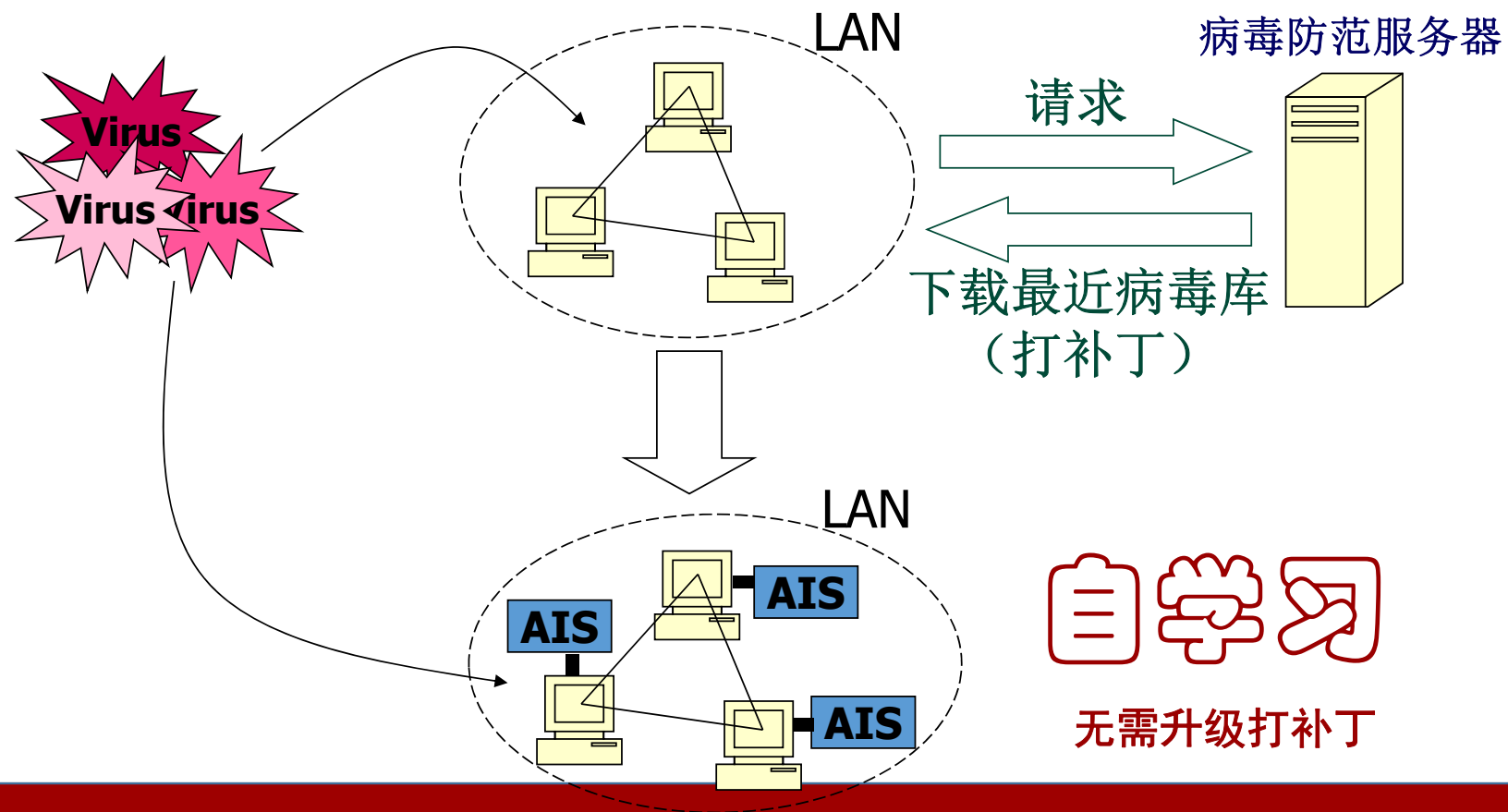
借鉴生物免疫系统的思想建立的计算机安全系统，即

将人工免疫系统引入到入侵检测中

人工免疫学--- ideas



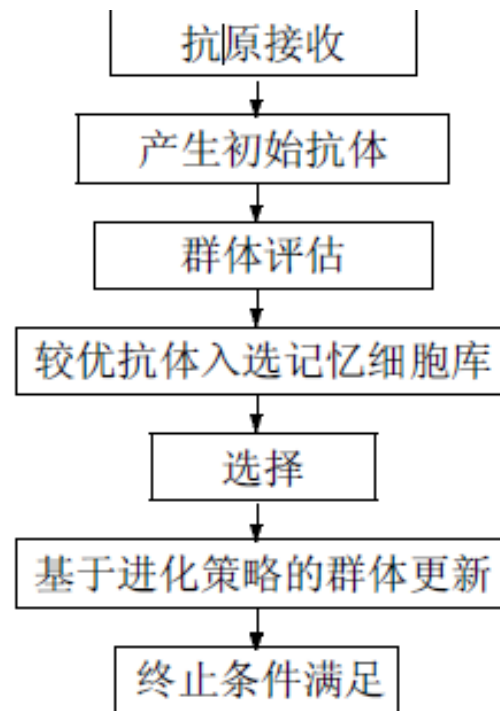
人工免疫学--- ideas



免疫系统与免疫算法的比较

免疫系统	免疫算法
抗原	要解决的问题
抗体	最佳解向量
抗原识别	问题识别
从记忆细胞产生抗体	联想过去的成功
淋巴细胞分化	优良解(记忆)的保持
细胞抑制	剩余候选解的消除
抗体增加(细胞克隆)	利用遗传算子产生新抗体

免疫算法步骤



将生物免疫模型
作为一种优化算法

图 2 免疫算法流程图

阴性选择算法

人工阴性选择方法是“自己/非己”区别的计算机模拟之一，首先设计应用于变化检测。

1. 任务类似
2. 所处环境类似
3. 所采用的检测方法类似

阴性选择算法

生物免疫系统与入侵检测系统的概念对比关系

生物免疫系统	入侵检测系统
缩胺酸/抗原决定基	被检测的行为模式串
受体	检测模式串
淋巴细胞 (T—细胞、B—细胞)	检测器
抗原	计算机中的监控进程
绑定	检测模式串和异己模式串的匹配
耐受性 (否定选择)	否定选择
淋巴细胞克隆	检测器复制
抗原检测	入侵检测系统的检测
抗原消除	检测器响应

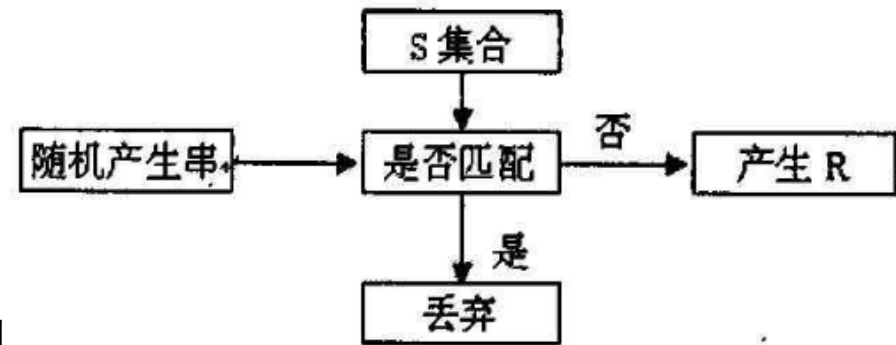
阴性选择算法

传统阴性选择算法实际上可分为三个阶段：定义“自己”、产生检测器、监测异常的发生。一般将定义“自己”与产生检测器阶段看作一个阶段，因此大致分为两个阶段：产生阶段和检测阶段。在产生阶段，给定一个称为“自己”集合 S 的已知模式集合，测试 T 细胞受体识别和结合“自己”模式的能力。如果 T 细胞受体识别来自“自己”集合的字符串，则忽略它；否则，作为一个竞争细胞进入检测器集合。检测器可用于监测系统的异常变化。

阴性选择算法

阴性选择算法：

收集自我集S
随机产生大量的字符串作为未成熟检测器 D_j ($j=1,2, \dots, N$)
while (成熟检测器集大小 < 规定的大小)
{ 计算未成熟检测器和自我集中每个个体的亲和力 aff_{ji} ($i=1,2, \dots, M; j=1,2, \dots, N$)
 if (亲和力 aff_{ji} > 阈值)
 删除该未成熟检测器 D_j
 else
 将 D_j 放入成熟检测器中
}



否定选择算法

匹配规则

Hamming匹配规则 设某字符集 m 上有长度为 l 的 X 和 Y 二进制字符串， m^l 表示 l 个字符的位串空间。Hamming规则通过计算两个字符串对应位置的相似度 $H(X, Y)$ ，并将它们与预先设定的判断阈值 σ 相比较来完成。

$$H(X, Y) = \sum_{i=1}^l \overline{(X \oplus Y)} \quad X, Y \in m^l$$

$H(X, Y)$ 值越大，表示 X, Y 越相似。当 $H(X, Y) \geq \sigma$ 认为 X, Y 匹配。Hamming匹配规则可以简单描述为：字符串 X 和 Y 匹配，当且仅当它们在 σ 位以上的位置上有相同的位数。

匹配规则

例如：取 $l=8$ ， $\sigma=4$ ，X和Y取值见：

A例：
X: 1 0 0 1 1 0 1 1
Y: 1 1 0 0 1 1 1 0

A例中X和Y匹配

B例：
X: 1 0 0 1 1 0 1 1
Y: 0 1 0 0 0 1 1 0

B例中X和Y不匹配

B例中只有当 $\sigma \leq 2$ 时，X和Y才表现为匹配。

阴性选择算法

rcb (r-contiguous-bits) 匹配规则，也称为r连续位匹配规则，规则指出：如果两个二进制字符串X和Y至少有连续r位上的符号相同，则称X和Y相匹配。即字符串X和Y匹配，当且仅当它们在r位以上连续相同的位置有相同的字符，记做：

X: 1 0 1 0 0 0 1 1

Y: 1 1 1 0 0 0 0 1

字符串X和字符串Y有连续4位的字符完全一样，如果 $r=4$ ，此时X和Y发生匹配，如果 $r>4$ ，则X和Y不匹配。

阴性选择算法

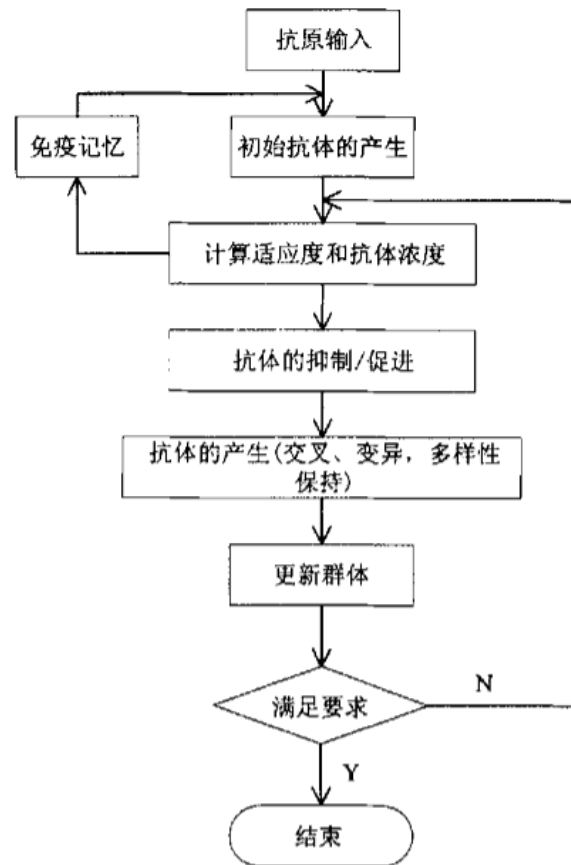
存在的问题：

1. 检测器集合覆盖空间重叠问题
2. 检测器自适应差及效率问题
3. 黑洞问题
4. 检测器收敛性问题

免疫遗传算法

在遗传算法中引入了免疫记忆库和浓度控制的免疫机制，克服了遗传算法易于陷入局部收敛的缺陷；在一般免疫算法中引进了标准遗传算法的交叉和变异过程来处理抗体，加快了抗体适应度的提高，从而加速算法的收敛。

免疫遗传算法流程图

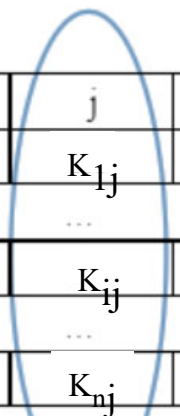


免疫遗传算法

多样性

当我们识别抗原以后，我们会随机产生一个初始抗体群，对于抗体群的大小，和抗体长度都是我们自定义的参数值。

现在假设一个免疫系统的初始抗体群如下图所示：抗体群大小为 N ，遗传因子共 M 位



	1	2	...	j	...	$M-2$	$M-1$	M
抗体 1				K_{1j}				
				...				
抗体 i				K_{ij}				
				...				
抗体 N				K_{nj}				

图 1 基因的信息熵

免疫遗传算法

为了判断抗体是否具有多样性，根据信息论的定义，我们用信息熵来表示抗体的多样性，抗体由m位等位基因构成，先计算各个等位基因的信息熵。

N个抗体第j个基因的信息熵为：

$$H_j(N) = \sum_{i=1}^S (-P_{ij} \log P_{ij}) \quad (1)$$

P_{ij} 表示在j位置的等位基因取值i的概率（等位基因取值范围是s个符号离散值）。

N个抗体对应共N个在j位置的等位基因，那么这N个值中假设取值为i符号的有 n_{ij} 个，那么

$$P_{ij} = \frac{n_{ij}}{N}$$

再计算该抗体群的信息熵，共M位等位基因组成一个完整的抗体

$$H(N) = \sum_{j=1}^M \frac{H_j(N)}{M} \quad (2)$$

	1	2	...	j	...	M-2	M-1	M
抗体 1				K_{1j}				
				...				
抗体 i				K_{ij}				
				...				
抗体 N				K_{Nj}				

免疫遗传算法

亲和性

抗原和抗体之间、抗体和抗体之间的匹配程度可以用亲和性来描述。

为保证抗体种群的多样性，在生成新一代抗体的过程中必须适度控制相同抗体的数量，避免近亲繁殖，这就需要计算抗体u和抗体v之间的相似度 $g(u, v)$ ，计算公式如下：

$$g(u, v) = \frac{1}{1 + H(u, v)}$$

$H(u, v)$ 为抗体u与抗体v所共有的信息量，当 $H(u, v)$ 为0时说明抗体u与抗体v完全一致，其相似度为1；当抗体u与抗体v不一致时， $H(u, v)$ 很大。

$H(u, v)$ -由u, v两个抗体组成的新抗体群计算对应的熵。（相当于计算新抗体群多样性）

免疫遗传算法

亲和度有多种表示形式，如Euclidean距离、Manhattan距离以及Hamming距离。

如果把抗体的坐标分别表示为 $(ab_1, ab_2, \dots, ab_i, \dots, ab_L)$ 和 $(ag_1, ag_2, \dots, ag_i, \dots, ag_L)$

$$\text{Euclidean距离 } D_E = \sqrt{\sum_{i=1}^L (ab_i - ag_i)^2}$$

$$\text{Manhattan距离 } D_M = \sqrt{\sum_{i=1}^L |ab_i - ag_i|}$$

若把抗体看作字符序列，则Hamming距离可用下式表示：

$$D_M = \sum_{i=1}^L \delta$$

$$\text{其中 } \delta = \begin{cases} 1 & \text{if } ab_i = ag_i \\ 0 & \text{otherwise} \end{cases}$$

免疫遗传算法

亲和性

抗原和抗体之间亲和性用于表明抗体对抗原的识别程度。

抗体j与抗原间的亲和度 m_j ：当 $m_j=1$ 时，表示抗体j与抗原完全亲和，抗原被消灭，抗体j即为所求问题的最优解。其求解公式如下：

$$m_j = \frac{1}{1 + OP_j}$$

其中， OP_j 为所求问题的目标函数。

免疫遗传算法

抗体生成的促进和抑制

当一种抗体和抗原相遇时，若发生亲和现象，则认为靠近最优解；反之，远离最优解。当寻优过程中如果抗体在染色体群中浓度过高，则会导致寻优过程停滞不前，造成早熟收敛。所以本算法则用控制浓度的方法来抑制相同或相似抗体的数量。所谓抗体的浓度 C ，是指群体中相似抗体所占的比重，即

$$C_i = \frac{\text{与抗体 } i \text{ 相似度大于 } \lambda \text{ 的抗体数和}}{N}$$

其中 λ 为相似度常数，一般取为 $0.9 \leq \lambda \leq 1$

免疫遗传算法

计算抗体浓度，并找出浓度较大的个体，记为个体1, 2, ..., t，则定义该t个个体的浓度概率为：

$$p_d = \frac{1}{N} \left(1 - \frac{t}{N}\right)$$

其中 $1 < t < N$,

其它 $(N-t)$ 个个体的浓度概率为 $\frac{1}{N} \left(1 + \frac{t^2}{N^2 - N * t}\right)$

个体的亲和力概率 P_f 用轮盘赌选择法计算。

个体的选择概率 P 由适亲和力概率 P_f 和浓度概率 P_d 两部分组成，即：

$$p = \alpha p_f + (1 - \alpha) p_d$$

式中， α —亲和系数， $\alpha > 0$ ； $P_f, P_d < 1$

几种免疫优化算法

在遗传算法的进化大框架上，引入免疫的抗体自我调节，免疫响应过程，免疫疫苗接种，免疫记忆等概念，发展了下面几种免疫优化算法：

- 基于免疫自我调节机制的免疫优化算法
- 基于免疫响应的免疫优化算法
- 克隆选择算法
- 基于疫苗接种的免疫优化算法
- 基于免疫抗体记忆的免疫优化算法

基于免疫自我调节机制的免疫优化算法

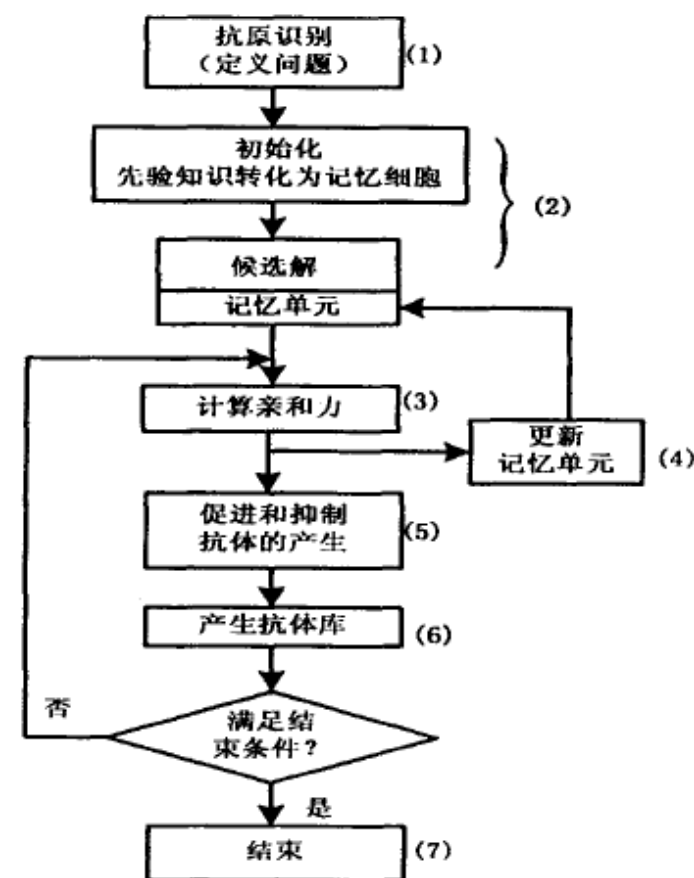
在遗传算法中，如果根据适应度函数选择出的双亲基因非常接近，那么所产生的后代相对双亲也必然比较接近，所期待的改善就比较小，这样基因模式的单一性不仅减慢进化历程，而且可能导致进化停滞，过早收敛于局部最优点。

在免疫调节中，那些与抗原亲和力大并且浓度较低的抗体会受到促进，而与抗原亲和力小或浓度较高的抗体会受到抑制，以此保证抗体的多样性。将这一概念应用到标准的遗传算法中，保证群体多样性，改善未成熟收敛，提高遗传算法的性能。个体浓度越小，选择概率越大，而个体浓度越大，则选择概率越小，这样，在保留高适应度个体的同时，进一步确保个体多样性，能避免早熟现象。

基于免疫响应的免疫优化算法

这种免疫优化算法从体细胞理论和免疫网络理论中得到启发，实现了类似于免疫系统的自我调节和生成不同抗体的功能，具有以下特点：

- 它在记忆单元基础上运行，对于曾经出现过的抗原，免疫算法产生相应抗体的速度比以前更快，提高了算法的总体搜索能力，确保了快速收敛于全局最优解；
- 它有计算亲和力的程序，体现了抗体的多样性；
- 它通过促进或抑制抗体的产生，体现了免疫反应的自我调节功能。



克隆选择——原理

- ◆ 1958年Burnet等提出了著名的克隆选择学说
- ◆ 抗体是天然产物，以受体的形式存在于细胞表面，抗原可与之选择性地反应
- ◆ 抗原与相应抗体受体的反应可导致细胞克隆增殖，其中某些细胞克隆分化为抗体生成细胞，通过抗体基因频繁的变异和编辑，在经过免疫后得到改善,即亲和度成熟。另一些形成免疫记忆细胞以参加之后的二次免疫反应。

克隆选择——原理

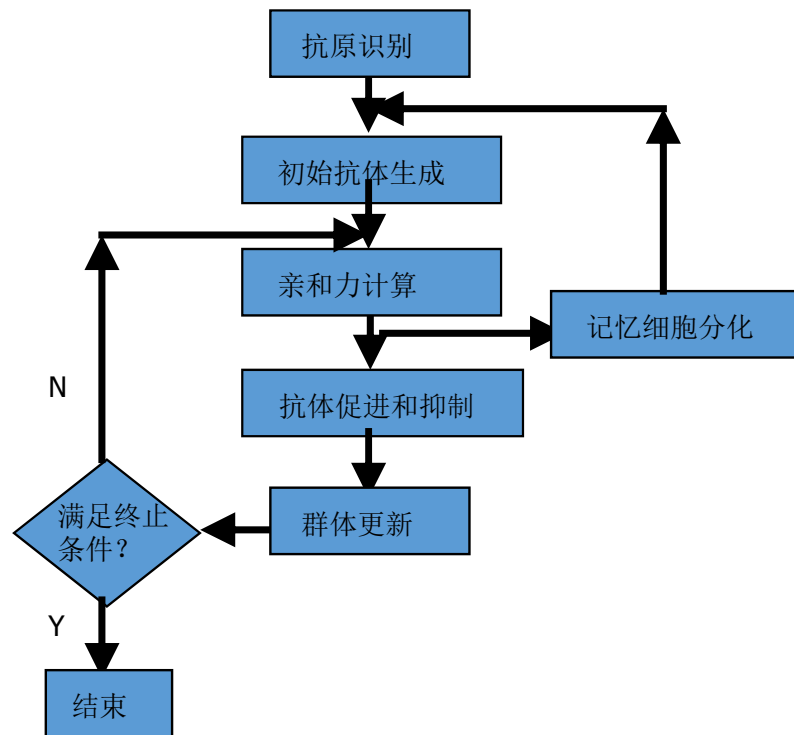
克隆选择学说的核心论点是:

- ①带有各种受体的免疫活性细胞克隆早已存在,抗原的作用只是选择并激活相应的克隆;
- ②细胞受体和该细胞后代所分泌的产物（抗体）具有相同的特异性。

克隆选择——算法

- (1) 产生一个初始群体，随机产生种特征组合作为初始抗体群体。
每个抗体对应一种特征组合
- (2) 计算每个抗体的亲和度，达到一定阈值或者迭代到一定次数，
选择最佳个体
- (3) 被选出的每个体克隆出若干个体，并使其发生变异，从而形成
下一代群体
- (4) 计算亲和度
- (5) 选择新群体中的一些个体替换初始群体中的一些父代个体。
- (6) 计算亲和度
- (7) $k=k+1$ ，返回 (3)

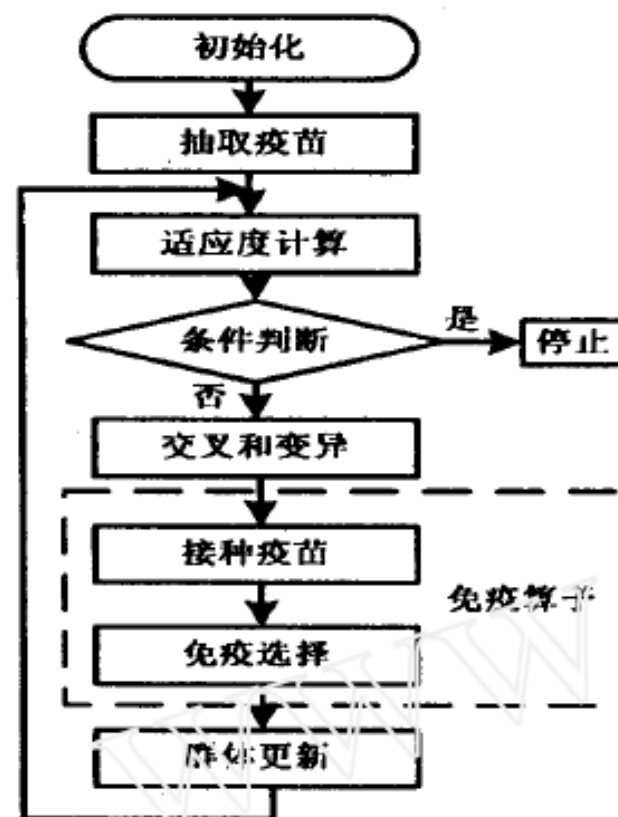
克隆选择——算法



基于疫苗接种的免疫优化算法

疫苗接种是免疫记忆临床应用的一个重要方面。将这一免疫概念用于遗传算法，有效地利用求解问题的一些基本的、显而易见的特征信息或知识，提取“疫苗”，使问题求解更具有针对性。

在保留原算法的优良特性的前提下，引入了一个新的算子——免疫算子，有选择、有目的地利用待求问题中的一些特征信息或知识，提取“疫苗”来抑制其优化过程中出现的退化现象，在进化选择过程中，通过“接种疫苗”和“免疫选择”来知道搜索过程。

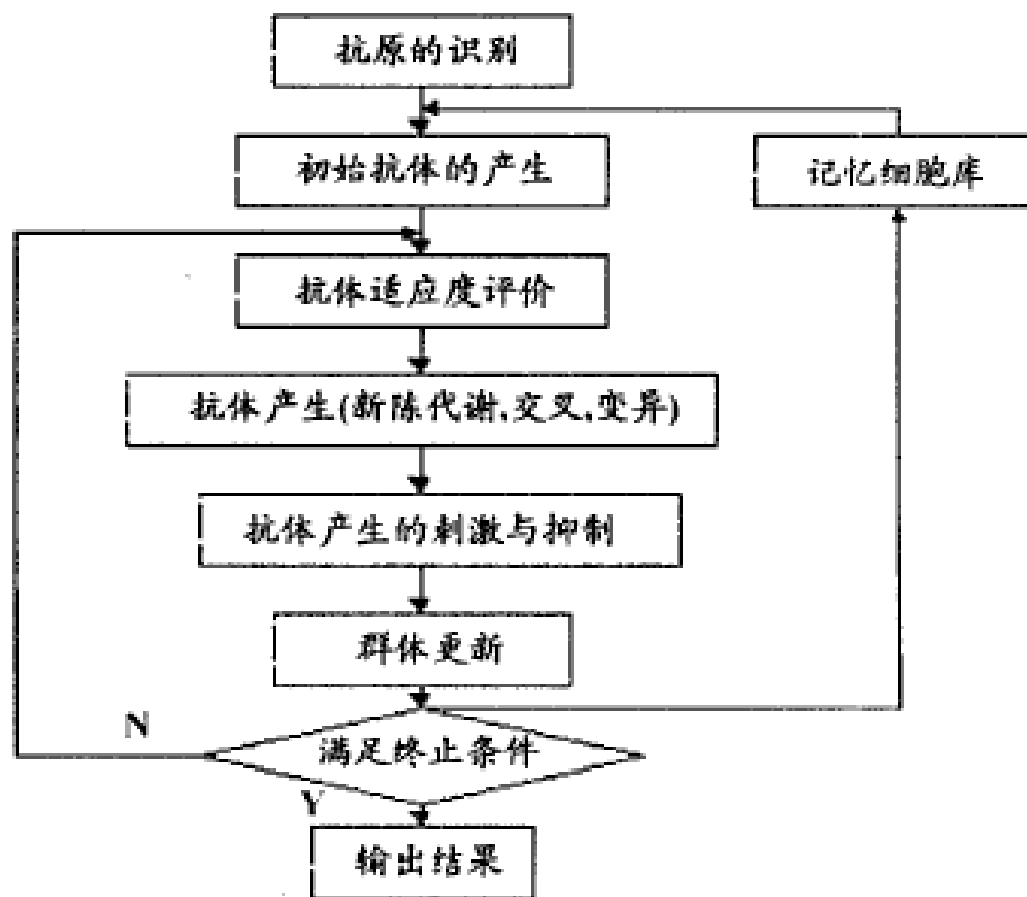


基于免疫抗体记忆的免疫优化算法

在适应性免疫调节过程中，对特殊抗体的记忆起着重要的作用。这种算法提出的免疫优化算法是对传统GA的改进，与GA相比，增加了两个新算子：一个是记忆细胞，一个是抑制细胞，这两种算子都是对重要抗体的记忆，记忆细胞是对具有高亲和力抗体的记忆，用于全局搜索过程中促进优良个体的搜索，而抑制细胞则是对高浓度抗体的记忆，在局部搜索过程中抑制相似解的搜索。

基于免疫算法的TSP问题的求解

所要用到的免疫算法框架



算法的设计与实现

在算法的实现中，将TSP问题的目标函数对应于抗原，而问题的解对应于抗体。

(1) 抗体编码及适应度函数

抗体采用了以遍历城市的次序排列进行编码，每一抗体码串形如： V_1, V_2, \dots, V_n ，其中， V_i 表示遍历城市的序号。程序中抗体定义为一维数组A(N)，N表示TSP问题中的城市数目，数组中的各个元素 $A(i)(i \in \{1, 2, \dots, N\})$ 取值为1至N间的整数，分别表示城市的序号，根据问题的约束条件，每一数组内的各元素值各不相同。

适应度函数取路径长度 T_d 的倒数，即 $Fitness(i) = 1 / T_d(i)$

$$T_d = \sum_{i=1}^{n-1} d(v_i, v_{i+1}) + d(v_1, v_n)$$

表示第i个抗体遍历城市的路径长度。

算法的设计与实现

(2) 抗体记忆细胞群

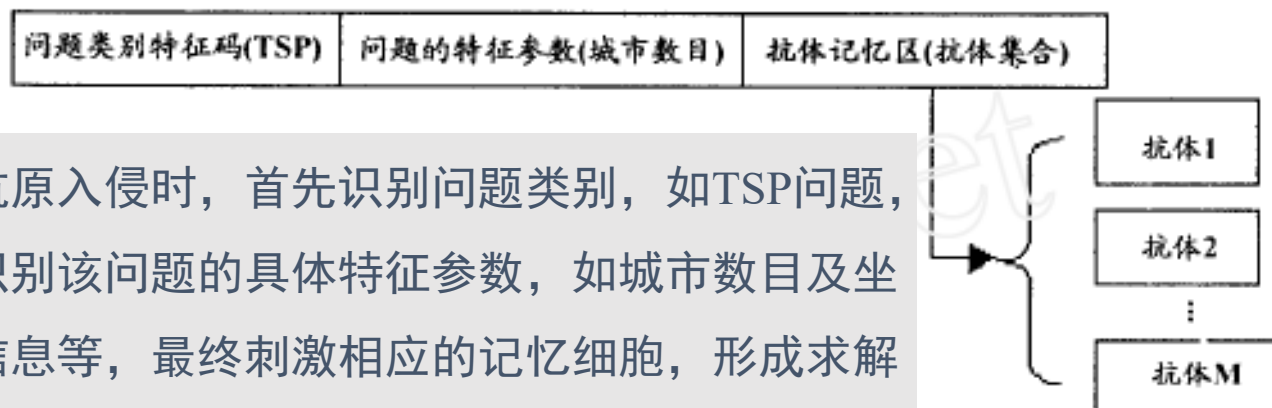
抗体记忆机制的引入是免疫算法的一个重要特点。系统在完成一个问题的求解后，能保留一定规模的求解过程中的较优抗体，这样，在系统接收同类问题的求解时，以所保留的记忆细胞作为初始群体，从而提高了问题求解的收敛速度，体现了免疫算法自适应环境的能力。

在这里设定记忆细胞规模为50，在求解过程中的每一代抗体群体的更新时，将适应度最好的M个（这里设为5个）抗体入选记忆细胞库中，每次入选记忆抗体时，若记忆细胞库已满额，则替换掉适应度最差的个体，同时比较库中有否与入选抗体相同的记忆细胞，以保证记忆细胞群的多样性。

算法的设计与实现

(2) 抗体记忆细胞群

记忆细胞库的结构



当抗原入侵时，首先识别问题类别，如TSP问题，再识别该问题的具体特征参数，如城市数目及坐标信息等，最终刺激相应的记忆细胞，形成求解问题的初始抗体群。

算法的设计与实现

(3) 初始群体产生

初始群体的产生有两种途径：

- 1) 针对待求解问题的特征，若能与记忆细胞群的结构信息相匹配，则由库中的记忆细胞组成初始抗体群，不足部分的抗体随机产生；
- 2) 否则随机产生初始群体。在这里设定群体规模为100。

算法的设计与实现

(4) 新抗体生成

根据免疫系统中抗体产生细胞的来源特点，每一代新抗体的产生主要基于：

1) 基于遗传操作的新抗体的产生：交叉采用部分匹配交叉策略，变异采用随机多次对换方式。这里设定的交叉概率为0.85，变异概率设定为0.03.由此产生新的抗体群体（群体规模为 N ,为初始群体规模）。

2) 随机产生的 P 个新抗体，类似于B细胞的新陈代谢功能，它对于进化过程中的抗体多样性的保持，提高全局搜索能力，防止未成熟收敛方面有较好的作用。

基于上述方式生成新的 $N + P$ 抗体，我们采用下述的基于浓度的策略，来修正各抗体的适应度值，并产生下一代群体（群体规模为 N ）。

算法的设计与实现

(5) 基于浓度的群体更新

在进化过程中，为了保持抗体的多样性，避免陷入局部最优，体现抗体之间的相互促进和抑制作用，在这里采用基于浓度的群体更新方法。即基于下述方式调整各抗体的适应度，在选取适应度较高的N个抗体组成新的抗体群。

$$\begin{aligned} Fitness(i) = & Fitness(i) \\ & + \alpha \cdot (1 - C) \cdot (1 + Sgn(\gamma - C)) \cdot \frac{Fitness(i)}{MaxFitness} \cdot Fitness(i) \\ & + \beta \cdot \frac{Fitness(i)}{MaxFitness} \cdot Fitness(i) \end{aligned}$$

其中， α, β 为 0—1 间可调参数， γ 为较优抗体的浓度阈值， $Fitness(i)$ 为第*i*个抗体的适应度， $MaxFitness$ 为抗体的最大适应值， $Sgn(x)$ 为符号函数， C 为较优抗体的浓度设为：

$$C = \frac{\sum_i (\eta \cdot MaxFitness \leq Fitness(i) \leq MaxFitness)}{\text{抗体总数 } N}$$

其中， η 为 0—1 的可调参数，这里取 0.8

算法的设计与实现

(5) 基于浓度的群体更新

在群体的更新过程中，一般总是希望适应度高的抗体被保留下来，但如果此类抗体过于集中，即浓度过高，则很难保证抗体的多样性，很容易使算法陷入局部极优，而丢失那些适应度较差但却可能保持这较好的进化趋势的抗体。因此，采用上述基于浓度机制的多样性保持策略，使得在新一代抗体群中，各适应度层次的抗体维持着一定的浓度。

结论

上面的几种免疫优化算法都是针对传统遗传算法存在的问题，如局部搜索能力差，存在未成熟收敛等现象，从而导致算法的收敛性能力差，需要很长时间才能找到最优解，结合不同的免疫信息处理机制发展而来的，因此可以遗传算法与免疫优化算法是融合发展的。

遗传算法与免疫算法的比较

- 二者的不同之处

	遗传算法	免疫算法
搜索目的	全局最优解	多峰值函数的多个极值
评价标准	解的适应度	解的适应度高且本身浓度较低
操作	选择和交叉为主	选择和变异为主
记忆库	无	有