

人工免疫算法综述

施建刚, 陈 罡, 高 喆

(上海大学 信息化工作办公室, 上海 200444)

摘 要:人工免疫算法的研究,已成为人工智能研究领域的一个重要内容。它突出地体现了现代科学发展的多层次、多学科和多领域的相互渗透、相互交叉和相互促进的特点,对信息科学和计算机科学的发展具有重要意义,同时也为工程实践人员提供了许多富有成效的技术和方法。因此,将人工免疫系统的原理应用在计算机领域有着重要的理论意义和实际应用价值。

关键词:人工免疫算法;遗传算法

中图分类号:TP312

文献标识码:A

文章编号:1672-7800(2008)11-0068-02

0 引言

从信息处理的观点看,免疫系统是与遗传系统、神经系统并列的人体三大信息系统之一。人类从免疫系统中不断获得新的启示并创造出越来越多智能方法,人工免疫算法就是其中的一种方法。在人工免疫算法中,被求解的问题视为抗原,抗体则对应于问题的解,改进的人工免疫算法与GA相似,人工免疫算法也是从随机生成的初始解群出发,采用复制、交叉、变异等算子进行操作,产生比父代优越的子代,这样循环执行,逐渐逼近最优解。不同的是人工免疫算法的复制算子模拟了免疫系统基于浓度的抗体繁殖策略,出色地保持了解群(对应于免疫系统中的抗体)的多样性,从而克服了GA解群多样性保持能力不足的缺点。

1 人工免疫算法

人工免疫算法将抗原和抗体分别对应于优化问题的目标函数和可行解。把抗体和抗原的亲合力视为可行解与目标函数的匹配程度:用抗体之间的亲合力保证可行解的多样性,通过计算抗体期望生存率来促进较优抗体的遗传和变异,用记忆细胞单元保存择优后的可行解来抑制相似可行解的继续产生并加速搜索到全局最优解,同时,当相似问题再次出现时,能较快产生适应该问题的较优解甚至最优解。人工免疫算法的基本步骤如下:

(1)问题识别。根据给定的目标函数和约束条件作为算法的抗原。

(2)产生抗体群。初始抗体群通常是在解空间用随机的方法产生的,抗体群采用二进制编码来表示。

(3)计算抗体适应值。即计算抗原和抗体的亲和度。

(4)生成免疫记忆细胞。将适应值较大的抗体作为记忆细胞加以保留。

(5)抗体的选择(促进和抑制)。计算当前抗体群中适应值相近的抗体浓度,浓度高的则减小该个体的选择概率—抑制;反之,则增加该个体的选择概率—促进,以此保持群体中个体的多样性。

(6)抗体的演变。进行交叉和变异操作,产生新抗体群。

(7)抗体群更新。用记忆细胞中适应值高的个体代替抗体群中适应值低的个体,形成下一代抗体群。

(8)终止。一旦算法满足终止条件则结束算法。否则转到(3)重复执行。人工免疫算法基本流程如图1所示。

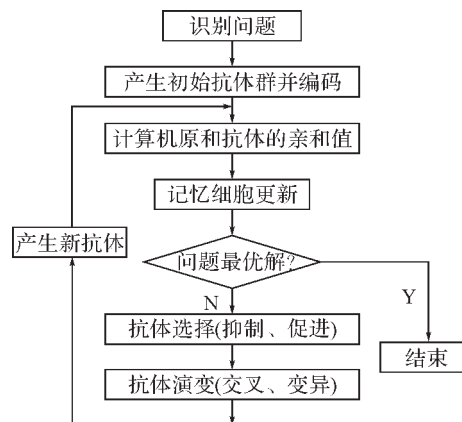


图1 人工免疫算法基本流程

作者简介:施建刚(1976~),男,上海人,上海润信息系统有限公司副总经理,研究方向为数字化校园、一卡通消费;陈罡(1980~),男,江苏无锡人,上海润信息系统有限公司软件部经理,研究方向为校园应用系统开发及整合;高喆(1980~),男,浙江鄞县人,硕士,上海大学信息化工作办公室助理工程师,研究方向为数字化校园。

2 人工免疫算法的改进

2.1 基于信息熵的人工免疫算法

标准人工免疫算法由于其采用了信息熵的概念来描述抗体的多样性,故也称作基于信息熵的人工免疫算法。其特点如下:①采用信息熵作为评价抗体群的相似度,较容易地与生物免疫系统中的抗体相对应,更能够客观地反应其含义;②采用基于浓度的选择机制,既鼓励适应度高的抗体(解),又可抑制浓度高的抗体,从而保证了算法的收敛性及抗体群体的多样性。

根据抗体群平均信息熵的概念计算抗体亲合度和浓度的方法可知,抗体群的所有抗体在同一基因座上的等位基因各不相同,抗体群的平均信息熵最大,抗体的亲和(相似)度最小;反之,则亲和度较大。然而,在优化计算中,这种利用信息熵的概念计算抗体(即解)亲和度和抗体浓度的方法存在一定的问题,例如对于由抗体 $u=\{0111111111\}$ 和抗体 $v=\{1000000000\}$ 构成的抗体群,其平均信息熵 $H(2)=0.693147$,表明抗体 u 和抗体 v 是两个很不相似的抗体。但是我们知道在优化计算中它们是两个很相近的解(通过解码后)。还比如:抗体 $u=\{0100000000\}$ 和抗体 $v=\{0011111111\}$ 等。由此,可以看出基于信息熵的人工免疫算法具有一定的缺陷。为了防止这种缺陷的发生,提出了一种改进的基于欧氏距离求解抗体多样性的人工免疫算法。

2.2 基于欧氏距离的人工免疫算法

一种改进的基于欧氏距离的人工免疫算法,采用欧氏距离代替信息熵,来反映抗体的多样性和计算抗体与抗体的亲和度。

基于欧氏距离的人工免疫算法的算法流程如图2所示。
两种算法性能比较结果如表1所示。

表1 两种人工免疫算法比较

性能	算法运行	每次搜	所耗	收敛于全局
算法	次数	索代数	时间	最优次数
基于信息熵的人工免疫算法	40	400	20.5	35
基于欧氏距离的人工免疫算法	40	400	3.2	36

其中,群体规模为 N 为30。算法运行一次是指算法从最初的随机产生群体到最后得出问题的解为止,称作算法运行了一

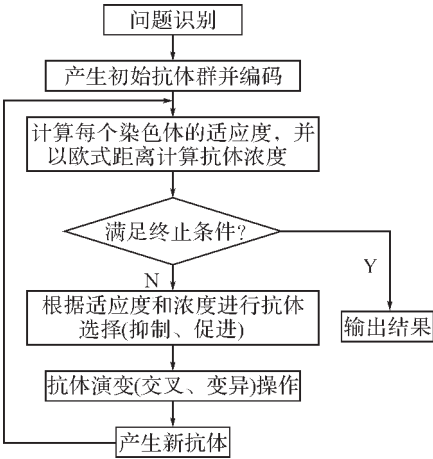


图2 基于欧氏距离的人工免疫算法流程

次,在此规定算法重复执行的次数为40次,而搜索代数是指在一次算法运行过程中产生新群体的代数为400代。所耗时间是算法运行40次所花费的总时间。同时在基于信息熵的人工免疫算法中,解抗体采用34位的二进制编码来表示,即每个自变量由17位基因组成,而在基于欧氏距离的人工免疫算法中,解抗体采用10位十进制数编码来表示,即每个自变量由5位基因组成,这一方面能保证解的精度,同时也能减少算法运行时间。

由表1可以看出,在基于信息熵的人工免疫算法中,由于要计算每代抗体群中每个个体的信息熵,使得算法的速度大大减缓,比改进了的基于欧氏距离的人工免疫算法在该算例实验中慢了几乎6.8倍。同时也可看出,当求解问题中的自变量增加时,信息熵的计算量将呈爆炸式增长的,这将极大地阻碍了基于信息熵的人工免疫算法的搜索效率,从而显现出基于欧氏距离的人工免疫算法的优点来。

参考文献:

[1] 罗小平.人工免疫遗传学习算法及工程应用研究[D].杭州:浙江大学博士论文,2002.
[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).

(责任编辑:杜能钢)