

改进免疫遗传算法及其应用研究

张建萍¹, 刘希玉², 谭业武¹

(1. 滨州学院, 山东 滨州 256603; 2. 山东师范大学, 山东 济南 250014)

摘要:在传统遗传算法基础上提出一种改进遗传算法——免疫遗传算法(IGA),该算法将生物系统免疫思想引入遗传算法中,通过计算抗体之间的亲和度来促进和抑制抗体,既保留了群体中较优抗体又保证了抗体多样性,避免搜索进化的过早收敛,得到全局最优解。文中采用此改进算法来控制无源光中光分支器和光网络单元中位置分配选择,通过浓度因子控制解空间的搜索方向,帮助逃离局部极值。利用 Visual C++ 6.0 对改进的免疫遗传算法和传统遗传算法进行比较,证明 IGA 的有效性和优越性。

关键词:无源光网络;遗传算法;免疫算子

中图分类号:TP301.6

文献标识码:A

文章编号:1673-629X(2008)10-0166-04

Study on Improved Immune Genetic Algorithm and Its Application

ZHANG Jian-ping¹, LIU Xi-yu², TAN Ye-wu¹

(1. Binzhou University, Binzhou 256603, China;

2. Shandong Normal University, Jinan 250014, China)

Abstract: In this paper, based on a genetic algorithm, an improved effective immune genetic algorithm is presented. In order to overcome premature convergence and find out optimal solution, the immune mechanism of creature is used in IGA and antibodies will be promoted or restrained according to the computation result of affinity between antibodies, which reserves the excellent antibodies as well as guarantees the variability of antibody. In addition, the IGA's local searching ability is improved by combining it with gradient method. The computation results show that both global and local searching abilities of the IGA are improved and premature convergence is overcome. The effectiveness and the superiority of IGA is proved on Visual C++ 6.0 by optimization experiments using another optimization algorithm comparisons to.

Key words: powerless optical network; genetic algorithm; immune operator

0 引言

伴随着遗传算法应用的深入开展,遗传算法有着其他优化算法不可比拟的优点,因此,遗传算法在优化计算^[1]中得到了广泛的应用,将遗传算法用于解决各种实际优化问题后,人们发现遗传算法^[2]也会由于各种原因,产生所谓“早熟”及收敛慢的问题,从而影响算法向全局最优解方向的搜索。文中通过深入分析遗传算法的特性,以及各操作算子对模式的影响,寻找遗传

算法早熟收敛的原因,并对其进行改进,以改善其收敛特性和收敛速度^[3]。文中抽取了生物体免疫系统的免疫机制,将其引入到遗传算法中,并与遗传算法紧密结合。通过保持个体的多样性,来避免早熟收敛和提高收敛速度。显然,这一研究课题不仅对发展遗传算法具有极大的新意,而且有一定的搜索性。另一方面,由于生物的免疫机制引入大大改善遗传算法的性能,因此遗传算法在优化计算中可以得到更广泛的应用。以无源光网络中光分支器和光网络单元中位置分配选择路径最小化^[4]为例,运用 Visual C++ 6.0 对改进的免疫遗传算法和传统遗传算法进行比较^[5],证明了 IGA 的有效性和优越性。

收稿日期:2008-01-26

基金项目:国家自然科学基金资助项目(6037405);山东省自然科学基金重大项目(Z2004G02);“泰山学者”建设工程专项经费资助项目;山东省滨州学院“青年人才创新工程”科研基金项目(BZXYQ NLG200711)

作者简介:张建萍(1979-),女,山东滨州人,助教,主要研究方向为数据挖掘、遗传算法;刘希玉,博士生导师,教授,主要研究方向为信息管理和信息管理系统(MIS)的研究及开发;谭业武,教授,主要研究方向为数据挖掘、计算机教育技术。

1 方案控制算法

1.1 问题描述及模型建立

在无源光网络(PON)网络中有 N 个光网络单元(ONU),位置为 U_1, U_2, \dots, U_n 和 M 个光分支器

(OBD), 位置为 D_1, D_2, \dots, D_m 。要求将各个 ONU 连到 OBD, 再将 OBD 连到光线路终端(OLT)(位置为 t_0) 的全部通信费用(路径和)最小。ONU 到 OBD 以及 OBD 到 OLT 的所有连接都是点状和星状结构。约束条件 1: 每个 ONU 恰好连一个 OBD; 约束条件 2: 1 个 OBD 最多只能有 K 个 ONU 和它相连。那么, 这个问题就转化为求节省通信费用(最小路径和)的组合优化问题。此问题的求解目标函数为:

$$C(t, n) = \min \sum_{i=1}^N \sum_{j=1}^M (d_{ij} + D_j) * X_{ij} \quad (1)$$

式中: d_{ij} 表示 ONU _{i} 到 OBD _{j} 的距离, D_j 表示 OBD _{j} 到 OLT 的距离, X_{ij} 表示 ONU _{i} 是否通过 OBD _{j} 连接到 OLT。

1.2 编码表示

针对文中的问题采用浮点数编码。

设 PON 网络中有 1 个 OLT, M 个 OBD(每个 OBD 最多允许与 K 个 ONU 连接)和 N 个 ONU, 其中 $M = 5, N = 10$ 。群体 $G(t)$ 中第 n 个个体的 ONU 1~10 分别连接到 OBD 的 2、3、5、1、5、5、2、2、3、4, 则该个体的基因序列 $p = 2351552234$, 即任一 $p_i (i = 1, 2, \dots, S)$ 就形成一个抗体。由于任意形成的抗体必须保证 OBD 不会过载, 即每个 OBD 最多允许与 K 个 ONU 连接, 所以文中建立 OBD 容量控制表 OBDCControl 记录群体中每个个体各 OBD 已有的连接数目, 如在此处 OBDCControl(t, n) = 13213。通过 OBD 容量控制表缩小抗体空间, 提高搜索效率。此种方法降低问题的时间复杂度。

1.3 遗传算子的选择

1.3.1 选择算子

文中问题求解采用轮盘赌选择机制, 设群体大小为 N , 个体 i 的适应度值为 f_i , 则个体 i 被选择的概率为 $p_f = \frac{f_i}{\sum_{i=0}^{N-1} f_i}$, 显然个体的适应度越大被选中的概率越高。

1.3.2 交叉算子

若对父个体 $p_1 = (23515 | 52234)$ 和 $p_2 = (51421 | 53352)$ 进行简单的单点交叉, 若随机断点选择在 5 号基因位, 则交叉后的结果为:

$$p_{12} = (23515 | 53352) \quad p_{21} = (51421 | 52234)$$

那么在 $p_{12} = (23515 | 53352)$ 中, OBDCControl(t, n) = 12304, 因此群体第五个个体中的 OBD 已有数目为 $4 > 3$, OBD 过载, 违反约束条件 2, 交叉后的抗体失去其有效性。因此, 用一组称为重排操作的新操作来处理这类问题^[6]。

针对此类问题的特殊性, 采用基于路径表示的顺序交叉(OX)操作。OX 操作能保留排列并融合不同排列的有序结构单元。两个父个体交叉时, 通过选择父个体 1 的一部分, 保留父个体 2 中城市码的相对顺序生成子个体。例如, 对下面两个父个体的表示, 随机选择两个交叉点“|”

$$p_1 = (2351 | 5522 | 34)$$

$$p_2 = (5142 | 1533 | 52)$$

首先, 两个交叉点之间的中间段保存不变, 得到:

$$p_{12} = (XXXX | 5522 | XX)$$

$$p_{21} = (XXXX | 1533 | XX)$$

然后, 记取父个体 2 从第二个交叉点开始基因码的排列顺序, 当到达表尾时, 返回表头继续记录基因码, 直至到达第二个交叉点结束, 这样便获得了父个体 2 从第二个交叉点开始的基因码排列顺序为 5-2-5-1-4-2-1-5-3-3。对于父个体 1 而言, 已有基因码有 5, 5, 2, 2, 将它们从父个体 2 的基因码排列顺序中去掉, 得到排列顺序 1-4-1-5-3-3, 再将这个排列顺序复制给父个体 1, 复制的起点也是从第二个交叉点开始, 以此决定子个体 1 对应位置的未知码 x , 这样子个体 1 生成: $p_{12} = (1533 | 5522 | 14)$, 子个体 2 为: $p_{21} = (5522 | 1533 | 42)$ 。

1.3.3 变异算子

变异采用多对基因位换位。预先确定一个正整数 U_e , 对抗体 $C = (C_1, C_2, \dots, C_n)$ 随机取一个正整数 $r (1 \leq r \leq U_e)$, 再在抗体 C 中随机取 r 对基因位做换位操作。如: 抗体做多对基因位换位的操作时, 若选择正整数 U_e 和 r 都为 2, 随机选取的位置为 4 和 7, 则交换以后的抗体为(1 6 8 3 5 7 4 2)。

1.4 终止准则

给定遗传算法的遗传代数 G_g , 当运算达到该代数时, 运算终止。如在迭代完成后依旧不能达到利用率的要求, 可以通过人工调整尝试加大 G_g 再次运算, 直至得到满意的值。

2 基于改进的免疫遗传算法

理论上证明, 采用这种保留每代最优解的算法将收敛到最优解, 但是, 当 ONU 与 OBD 的个数增加时, 收敛到最优解的时间可能为无穷, 所以, 改进上述的遗传算法来提高收敛的速度非常有必要。遗传算法中, 选择算子是在原来的种群中选择, 交叉算子是来源于原来种群的模式变化, 变异算子是用来冲破原来的模式在全局内变化。所以, 交叉和变异算子都是希望能够冲破局部极小值, 在全局内搜索的策略, 也有一定效

果,但是,这种变异和交叉是盲目的、无方向的。即使是自适应遗传算法,也只是考虑了交叉和变异的概率随适应度分布的改变而改变情形,没有考虑遗传算法的问题依赖性^[7]。所以,要真正明确解的分布,应当考虑具体个体的编码情况。

免疫算法^[8]将上述的思想进一步推进而引入了一个重要的概念:浓度,以此来描述个体与个体的相似程度。依据相似度并综合考虑适应度来决定保留的个体。这样,一方面保留了适应度高的解(即利用率高),又使得解能够被分散开来不至于限于局部极小而过早收敛。根据文中实际问题的需要,相似程度可用信息熵来度量。于是,文中提出如下改进算法。

算法的改进主要在基于求解 PON 建设总费用最少问题的相似度的重新定义。

2.1 抗体之间的相似度定义

相似度 A_{ij} ,是指两个抗体 i 和 j 之间相似的程度。定义如下,且有 $A_{ij} \in (0,1]$:

$$A_{ij} = \frac{1}{1 + H(2)} \quad (2)$$

A_{ij} 越大表示 i 和 j 两个抗体越亲和。(2) 式中的 $H(2)$ 为基于抗体 i 和 j 采用信息熵公式求其值。实际上,相似度是抗体间及抗体和抗原间相似程度的一种度量。只要满足两个条件即可,一是 $H(2) \geq 0$,且 $H(2) = 0$ 时,两个抗体的基因完全相同;二是两个抗体的基因差异越大,则 $H(2)$ 的值越大。为使计算简单、易于工程设计的实现,采用计算欧式距离的方法,即将两抗体看作两个向量,计算这两个向量之差的 Euclid 范数。设抗体 $i = (i_1, i_2, \dots, i_m)$ 和 $j = (j_1, j_2, \dots, j_m)$ 则有:

$$H(2) = \sqrt{\sum_{i=1}^n (v_i - w_i)^2} \quad (3)$$

假设 N 为抗体总数即群体规模。则抗体 i 浓度的计算可以采用下式且有 $C_i \in (0,1]$:

$$C_v = \frac{1}{M} \sum_{w=1}^M a_{yvw} \quad (4)$$

某种抗体的浓度越大,则说明这种抗体和其他抗体的相似性越大。

考虑到为了在下述的免疫选择中,使 C_i 随进化代数 K 的增加而最终趋于 1,以避免算法后期的振荡,有利于算法最终收敛,对式(4)作如下修正:

$$C_v = \left(\frac{1}{M} \sum_{w=1}^M a_{yvw} \right)^{(1 - \frac{K}{K_{max}}) \cdot \beta} \quad (5)$$

式(5)中的 β 为系统参数,取 $\beta = 0.5$ 。

2.2 免疫选择

采用综合考虑个体的选择概率、亲和度常数及个体的浓度概率来决定选择个体。抗体的选择概率公式

为: $P = \alpha P_f + (1 - \alpha) P_d$,其中, P_f 由轮盘赌选择法计算, α 为亲和度常数, P_d 为个体的浓度概率,然后根据计算出的 P 选择个体进入下一代。该选择体现了免疫系统对抗体抑制和促进的自我调节作用,适应度高、浓度低的抗体得到促进,生存几率较大;相反,适应度低、浓度高的抗体得到抑制,生存几率较小。它可有力维护群体多样性,防止算法早熟。

3 仿真实验

3.1 SGA 与 IGA 结果比较一

上述算法思想用 VC++ 6.0 来实现。在程序的实现时,选定交叉概率为 0.6,变异概率为 0.05,OBDD 的最大连接数为 $K = 4$,ONU 数目为 $N = 20$,OBDD 数目为 $M = 10$,群体规模 $G = 100$ 。通过反复改变遗传代数来决定终止的条件。分别用 SGA 和 IGA 进行计算,求解结果分别如图 1、图 2 所示。

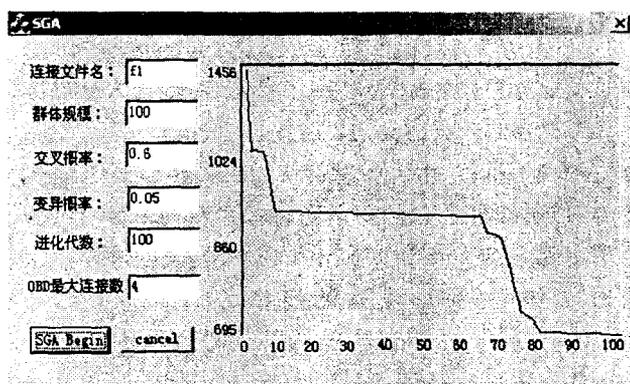


图 1 SGA 求解结果

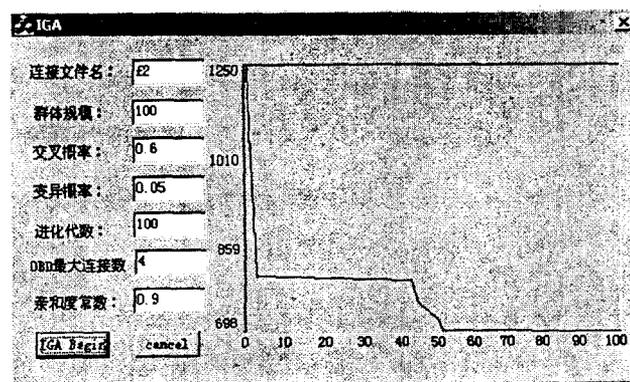


图 2 IGA 求解结果

从以上两图可以看出免疫遗传算法在第 50 代左右搜索到了最优解,而遗传算法在第 80 代左右才搜索到最优解,说明与 GA 比,IGA 有着较快的收敛速度,显示了该算法在求解 NP 问题的潜力。

3.2 SGA 与 IGA 结果比较二

对每种算法各进行 50 次实验,最后对多组数据求平均值进行分析,得到文中所提出的改进遗传算法与

传统遗传算法的比较数据如表 1 所示。

表 1 SGA 与 IGA 结果比较

	最优适值	平均适值	收敛率/%	CPU/s
SGA	0.062	0.027	100	2.04
IGA	0.089	0.031	100	1.91

4 结束语

以求解无源光网络中 OBD 与 OUN 位置分配问题为例,对人工免疫遗传算法进行了深入的研究,提出改进的免疫遗传算法。该算法在产生多样性个体能力、相似性矢量距选择概率的一般性、最优个体记忆功能等方面具有突出的优点。

实验证明在优化问题中应用改进的免疫遗传算法确实取得了良好的效果。

参考文献:

- [1] Buckles B P, Petry F E. Genetic Algorithms[M]. Los Alamitos, California: IEEE Computer Society Press, 1992.

(上接第 122 页)

虑底层网络的影响。与其它模拟器相比,通过调节底层网络参数,可以发现 P2P 算法对底层网络的敏感程度,为进一步分析、修改算法提供依据。

目前,该模拟器只能在一个工作站上进行模拟实验,未来的工作主要放在并行模拟的实现。首先在多核处理器上进行模拟,这其中涉及到并发程序设计。希望通过 Java 的线程机制简化这方面的工作。之后,将尝试将模拟器部署在多个工作站上,考虑使用 Java 中的 RMI 机制进行实现。

参考文献:

- [1] Rowstron A, Druschel P. Pastry: Scalable, Decentralized Object Location, and Routing for Large - Scale Peer - to - Peer systems[J]. Lecture Notes in Computer Science, 2001, 218: 329 - 350.
- [2] Yang W, Abu - Ghazaleh N. GPS: A General Peer - to - Peer Simulator and its Use for Modeling BitTorrent[C]//In Pro-

(上接第 165 页)

- [4] 洪 玫,刘 佳.基于服务数据对象的异构系统数据集成方案研究[J].计算机应用,2007(6):21 - 23.
- [5] 冀汶莉,龚尚福.煤矿安全远程综合监控系统异构数据集成技术的研究[J].工矿自动化,2007(4):6 - 9.
- [6] 王艳华,薛胜军,蒲秋梅.公积金监管系统中的多级数据集成研究[J].武汉理工大学学报:交通科学与工程版,2007(3):544 - 547.

- [2] 李敏强,寇纪淞.遗传算法的基本理论与应用[M].北京:科学出版社,2003.
- [3] 钱光超,贾瑞玉,张 然,等.基于一种改进的遗传算法的空间聚类分析[J].计算机技术与发展,2007,17(12):71 - 75.
- [4] 周伟良,何 鲲,曹先彬,等.基于一种免疫遗传算法的 BP 网络设计[J].安徽大学学报:自然科学版,1999,23(1):63 - 66.
- [5] 四维科技,赵 辉.Visual C++ 系统开发实例精粹[M].北京:人民邮电出版社,2005.
- [6] Oliver L M, Smith D J, Holland J R C. A Study of Permutation Crossover Operators on the Traveling Salesman Problem [C]//In: Proceedings of the Second International Conference on Genetic Algorithms. Lawrence: Erlbaum Associates, 1987: 24 - 130.
- [7] 莫宏伟.人工免疫系统原理与应用[M].哈尔滨:哈尔滨工业大学出版社,2002:154 - 155.
- [8] 宋亚男,邓飞其,叶家玮.基于改进免疫遗传算法的不规则图形排样[J].计算机工程 2005,31(9):170 - 172.

ceedings of MASCOTS. Atlanta, GA: [s. n.], 2005: 425 - 434.

- [3] Calvert K, Zegura E. GT Internetwork Topology Models (GT - ITM) [R]. Georgia: College of Computing, Georgia Institute of Technology, 1996.
- [4] Joseph S. NeuroGrid: Semantically Routing Queries in Peer - to - Peer Networks [C]//Proceedings of the International Workshop on Peer - to - Peer Computing. Pisa, Italy: [s. n.], 2002.
- [5] Clarke I, Sandberg O, Wiley B, et al. Freenet: A Distributed Anonymous Information Storage and Retrieval System [C]//Proc. International Workshop on Design Issues in Anonymity and Unobservability, volume 2009 of LNCS. [s. l.]: Springer - Verlag, 2001: 46 - 66.
- [6] Mathis M, Semke J, Mahdavi J. Modeling TCP Throughput: A Simple Model and its Empirical Validation [J]. Computer Communication Review, 1997, 27(3): 21 - 24.

- [7] 袁勤勇,王福华. BizTalk Server 开发人员指南[M].北京:清华大学出版社,2002.
- [8] 刘 芳,肖铁军.XML 应用的基石: XML 解析技术[J].计算机工程与设计,2005(10):2823 - 2825.
- [9] 莫高习,傅秀芬.基于 XML 技术的异构数据集成系统框架设计与探讨[J].现代计算机,2005(10):9 - 12.