

基于动物群落结构的改进病毒进化遗传算法

顾民 杨峰

(电子科技大学电工学院 成都 611731)

摘要 基于动物群落组织结构提出了一种改进的病毒进化遗传算法。主种群由一个父亲和若干个母亲及单身汉组成,父亲和母亲交叉产生后代,单身汉变异产生后代。病毒分为大病毒和小病毒,在前期迭代过程中,大病毒起作用,在后期迭代过程中,小病毒起作用。实例证明,改进的病毒进化遗传算法性能优于传统的病毒进化遗传算法。

关键词 病毒进化遗传算法,动物群落,组织结构,函数优化

中图分类号 TP301.6 **文献标识码** A

Improved Virus-evolutionary Genetic Algorithm Based on Fauna Structure

GU Min YANG Feng

(College of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)

Abstract An improved virus-evolutionary genetic algorithm based on fauna structure was proposed in this paper. The host population is consisted of a father and several mothers and several bachelors. Crossover is occurred between the father and the mothers to produce offspring. Mutation is occurred among the bachelors to produce offspring. Virus is divided into big virus and small virus. In the previous iteration, big virus works and in the latter iteration, small virus works. The instances verified that the improved virus-evolutionary genetic algorithm outperforms the tradition virus-evolutionary genetic algorithm.

Keywords Virus-evolutionary genetic algorithm, Fauna, Structure, Function optimization

遗传算法是一种借鉴生物界自然选择和生命遗传机制的全局随机搜索算法,它采用群体搜索策略,搜索过程中不依赖于梯度信息,只依赖适应度值,具有较强的鲁棒性,相比传统的优化算法具有很大的优越性,近年来广泛应用于优化和搜索、机器学习、智能控制、模式识别、人工生命、人工神经网络的权系数调整等领域^[1-3]。遗传算法从一组随机产生的初始解群体开始搜索,通过交叉和变异算子产生后代,通过选择算子进行优胜劣汰,经过若干代以后,算法收敛至最优个体。虽然遗传算法具有很多的优点,但传统遗传算法也存在不少缺点,例如易于早熟、易于陷于局部极值点等^[4,5]。许多学者因此提出许多改进的遗传算法^[6-8],其中 Kubota 于 1996 年基于病毒进化理论提出的病毒进化遗传算法较好地改善了传统遗传算法的性能,并成功应用于求解旅行商问题、自组织制造系统的调度问题和背包问题等^[9,10]。本文在传统病毒进化遗传算法的基础上提出了一种基于动物群落组织结构的改进的病毒进化遗传算法,实例验证了该算法的有效性。

1 病毒进化遗传算法

病毒进化遗传算法(简称 VEGA)是一种引入了病毒进化理论的混合遗传算法,它包含主种群和病毒种群两个种群,主种群等同于传统遗传算法的主群,对应问题的解空间,执行选择、交叉和变异遗传操作。病毒种群是主种群的子串集,主

要对主种群进行病毒感染操作。病毒进化遗传算法通过将主种群的全局进化和病毒种群的局部进化进行动态结合,加快了搜索速度。在实际应用中,病毒常常采取主个体染色体等长的染色体,其染色体常由有效符号和无效符号共同构成,无效符号不参与病毒感染操作^[11]。例如对于二进制编码,病毒染色体由 3 个符号{0,1,*}构成,符号“*”是通配符,为无效字符,符号{0,1}是有效字符。

1.1 病毒感染算子

病毒进化遗传算法引入了反向转录算子和转导算子两个病毒感染算子,如图 1 所示。

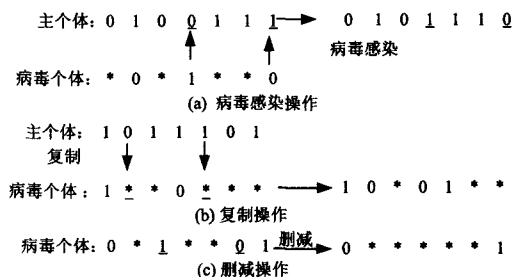


图 1 病毒感染算子

1) 反向转录算子(reverse transcription operator),也称为感染操作,即利用病毒染色体基因替换主群个体的相应基因从而生成新的主个体,如图 1(a)所示。

到稿日期:2010-01-26 返修日期:2010-03-09 本文受国家自然科学基金课题(编号 10876007)资助。

顾民(1971-),男,博士后,副教授,CCF 会员,主要研究方向为智能优化在天线中的应用,E-mail: gumind_cdut@126.com;杨峰(1962-),男,教授,博士生导师,主要研究方向为天线技术。

2) 转导算子(transduction operator), 包含两种操作, 一种是复制操作(copy), 即病毒从主种群个体的染色体复制一定长度的基因子串从而生成一个新的病毒, 如图 1(b) 所示。另一种是删减(cut)操作, 病毒染色体的每个基因以一定的概率用通配符“*”代替, 如图 1(c) 所示。初始病毒种群的个体是通过随机选择初始主种群中主个体进行复制操作生成的。

1.2 病毒进化遗传算法的参数

1) 病毒适应度函数(fitvirus)

病毒适应度函数通过主群体被感染前后的适应度值变化来衡量。假设病毒 i 感染的主群体为 U , fit_{host_j} 和 fit_{host_j}' 分别是主群体 U 中个体被感染前后的适应度函数值, 则病毒 i 的适应度函数为:

$$fitvirus_i = \sum_{j \in U} (fit_{host_j}' - fit_{host_j}) \quad (1)$$

2) 病毒感染率(infrate)

病毒感染率表征了病毒个体感染能力的强弱。病毒个体 i 的第 $t+1$ 代的感染率表示如下:

$$infrate_{i,t+1} = \begin{cases} (1+\alpha) \cdot infrate_{i,t}, & fitvirus_i > 0 \\ (1-\alpha) \cdot infrate_{i,t}, & otherwise \end{cases} \quad (2)$$

3) 病毒的生命力

病毒个体 i 的第 $t+1$ 代的生命力表示如下:

$$life_{i,t+1} = r \times life_{i,t} + fitvirus_i \quad (3)$$

式中, r 为生命力的衰减率。

1.3 传统病毒进化遗传算法

传统病毒进化遗传算法可简述如下。

步骤 1 初始化。设定初始参数, 形成初始主种群和初始病毒种群。

步骤 2 主种群遗传操作。主种群进行选择、交叉和变异操作。

步骤 3 病毒感染操作。病毒个体对主个体进行病毒感染操作, 如果感染后新的主个体适应度值大于感染前主个体的适应度值, 则用感染后的新个体替换该主个体, 否则不进行替换。

步骤 4 病毒的自我进化操作。计算每个病毒的感染力和生命力, 当病毒 i 的生命力小于零时, 该病毒自然死亡, 然后从感染后的主群体中随机寻找一个个体通过复制操作生成一个新的病毒; 当病毒 i 的生命力大于等于零且该病毒个体适应度大于等于零时, 则对该病毒以一定概率执行复制操作, 当病毒 i 的生命力大于等于零且该病毒个体适应度小于零时, 以一定概率执行删减操作。

步骤 5 判断收敛准则是否满足, 若满足, 则停止迭代, 否则转步骤 2。

2 基于动物群落结构的改进病毒进化遗传算法

在某些动物群落中, 如猴群, 存在着这样的组织结构, 即整个群体中有一个父亲、若干的母亲和一些单身汉, 最强壮的个体才能成为首领而享有父亲的权利。基于这种现象, 本文将主种群的主个体分成一个父亲、若干母亲和若干单身汉这样的结构, 父亲和母亲执行交叉操作, 父亲和母亲不进行变异, 单身汉执行变异操作。

遗传算法中主个体的染色体是由各个待寻优参数所对应的基因段组合而成, 每一个基因段上不同位置的基因的变动对于其相应寻优参数的影响是不一样的, 其中高位基因变化

对对应的参数在寻优空间产生较大的尺度变动, 低位基因变化使参数在寻优空间变动尺度较小。因此, 本文将每个寻优参数对应的染色体基因段分为两个部分, 前半部分称为该参数的大尺度基因段, 后半部分称为小尺度基因段。相应地将病毒也分为大病毒和小病毒两种, 大病毒具有这样的特征, 其染色体的小尺度基因座恒为通配符“*”, 对于小病毒, 其大尺度基因座恒为通配符“*”。图 2 为大病毒和小病毒的染色体示意图, 其对应的主个体染色体是由两个待寻优参数体构成, 对于大病毒, 下划线所对应的基因座为参数 1 和参数 2 的小尺度基因, 其值恒为通配符。对于小病毒, 下划线对应的基因座为参数 1 和参数 2 的大尺度基因座, 其值恒为通配符。

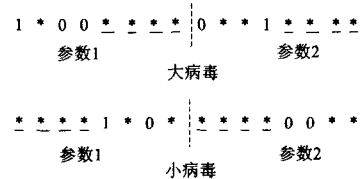


图 2 大病毒和小病毒染色体示意图

大病毒只能对主个体大尺度基因进行感染, 使主个体对应的显性值产生较大变化。小病毒只能对主个体的小尺度基因进行感染, 使主个体对应的显性值产生较小变化。在前期迭代过程中, 主个体感染大病毒而产生较大尺度的变化, 从而在全局范围内搜索, 不易陷入局部极值点; 在后期迭代中, 主个体感染小病毒而产生较小尺度的变化, 从而进行局部搜索, 加快收敛速度。

改进病毒进化遗传算法的步骤如下。

步骤 1 初始参数的设置, 包括主种群主个体数和病毒种群的病毒个体数、变异率、病毒感染率、病毒生命力衰减率、复制率以及删减率、母亲占总群体的百分比等参数。其中母亲占总群体的百分比 p_{mother} 是一个随迭代数逐渐减小的参数。第 $t+1$ 代 $p_{mother,t+1}$ 为:

$$\begin{cases} p_{mother,t+1} = p_{mother,t} - \alpha \times t, & p_{mother,t+1} > p_{min} \\ p_{mother,t+1} = p_{min}, & p_{mother,t+1} \leq p_{min} \end{cases} \quad (4)$$

式中, α 为比例系数, t 为迭代数, p_{min} 为母亲占总群体的最小百分比。

步骤 2 初始种群的产生。主种群随机产生, 病毒种群的个体由随机挑选的主个体执行复制操作产生, 大病毒个体由主个体染色体的大尺度基因按复制率执行复制操作生成, 小病毒个体由主个体染色体小尺度基因按复制率执行复制操作生成。

步骤 3 角色分配。计算每个主个体的显性值、适应度及种群中母亲数量 M 。主种群中适应度最高的主个体为父亲, 然后计算其余所有个体与父亲的显性值的欧式距离, 其中离父亲距离最远的 M 个个体作为母亲, 剩余的个体作为单身汉。

步骤 4 对主种群执行交叉和变异操作。其中父亲与每一位母亲执行交叉操作产生两个后代, 单身汉按照变异率执行变异操作产生一个后代, 父亲和母亲不参与变异。

步骤 5 竞争选择。父亲直接保留在主种群。母亲与自身产生的两个后代进行竞争, 三者中适应度最高的一个个体保留在主种群, 单身汉与其变异后产生的后代进行竞争, 两者中适应度值高的个体保留在主种群, 其余个体则消亡。

F1	-1.031628	-1.02799	-1.031607	70	37	31	98
F2	0	0.24×10^{-5}	0.12×10^{-9}	67	44	42	92

步骤6 病毒感染操作。病毒个体对主个体进行病毒感染操作,在前期迭代利用大病毒进行感染,在后期迭代,利用小病毒进行感染。如果感染后新的主个体适应度值大于感染前主个体的适应度值,则用感染后的新个体替换该主个体,否则不进行替换。

步骤7 病毒的进化操作。

步骤8 判断收敛准则是否满足,若满足,则停止迭代,否则转步骤3。

3 仿真和结果

选择文献[12]两个函数优化问题来比较改进的病毒进化遗传算法(简称 IVEGA)和传统的病毒进化遗传算法(VEGA)的性能。

$$F1(x, y) = (4 - 2.1x^2 + x^4/3)x^2 + xy + (-4 + 4y^2)y^2$$

$$x, y \in [-100, 100] \quad (5)$$

F1 有 6 个局部极小值和两个全局最小值,当 $(x, y) = (-0.0898, 0.7126)$ 和 $(x, y) = (0.0898, -0.7126)$ 时全局最小值 $F1 = -1.031628$ 。

$$F2(x, y) = 0.5 + \frac{\sin^2 \sqrt{x^2 + y^2} - 0.5}{[1 + 0.001 * (x^2 + y^2)]^2}$$

$$x, y \in [-10, 10] \quad (6)$$

F2 是 Schaffer 函数,当 $(x, y) = (0, 0)$ 时,有一个全局最小点 $F2 = 0$ 。

对于测试函数,当算法搜索到的解与其理论最优解之差的绝对值小于给定的允许误差时,认为算法已搜索到最优解。对于测试函数 $F1, F2$,给定的允许误差分别为 $1 \times 10^{-2}, 1 \times 10^{-5}$ 。

试验采用二进制编码,解的精度为 1×10^{-6} ,传统的病毒进化遗传算法中遗传操作采用最佳精英保留策略遗传操作,即轮盘赌选择,单点交叉和单点变异,保留最佳个体。两种算法终止条件采用预先设定的最大世代数,最大世代数为 100,对改进的病毒进化遗传算法,前 20 代采用大病毒感染主个体,后 80 代采用小病毒感染主个体。两种算法的参数如表 1 所列。

表 1 VEGA 和 IVEGA 算法的参数

主种群规模	100	母亲比例递减系数	0.05
病毒种群规模	10	交叉率	0.65
大病毒种群规模	10	变异率	0.05
小病毒种群规模	10	生命力衰减率	0.05
初始母亲比率	0.5	最大感染率	0.1
最小母亲比率	0.03	初始感染率	0.05

本文进行 100 次独立试验比较两种算法的性能,表 2 给出的是 100 次独立试验的统计结果。表中平均值只针对达到搜索精度的结果进行计算。

从表 2 可以看出,IVEGA 搜索到最优解的次数明显多于 VEGA,说明 IVEGA 全局搜索能力强。IVEGA 搜索到最优解的平均迭代次数少于 VEGA,说明 IVEGA 具有更快的收敛速度,IVEGA 搜索到的最优解的平均值更接近理论最优解,说明 IVEGA 有更高的搜索精度。

表 2 100 次独立试验统计结果

Func	理论最优解	搜索到最优解的平均值		搜索到最优解的平均迭代次数		搜索到最优解的次数	
		VEGA	IVEGA	VEGA	IVEGA	VEGA	IVEGA

为更直观地比较两种算法的进化过程,对于两个函数,随机选取两种算法中已搜索到的最优解中的一次运行结果,如图 3 和图 4 所示,从图中可以看出 IVEGA 有更快的收敛速度。

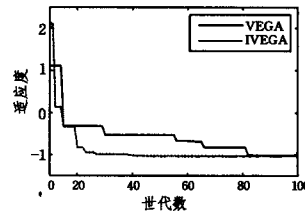


图 3 VEGA 和 IVEGA 对 F1 寻优仿真结果

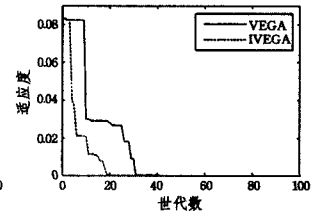


图 4 VEGA 和 IVEGA 对 F2 寻优仿真结果

结束语 本文根据动物群落的组织结构提出了改进病毒进化遗传算法,其将主种群的主个体分为一个父亲、若干个母亲及单身汉,父亲和所有的母亲执行交叉操作,单身汉进行变异操作。同时将病毒分为大病毒和小病毒,前期迭代过程采用大病毒进行感染,后期迭代过程采用小病毒进行感染,通过两个函数优化的实例证明,该改进病毒进化遗传算法比传统的病毒进化遗传算法有更快的收敛速度、更高的全局搜索能力并能得到更高的搜索解。

参考文献

- [1] Wu Ying-ying, Liu Jin, Nie De-ming. Optimal predicting PID controller based on elite-based hybrid genetic algorithm [C]// Proceedings-2009 IITA International Conference on Control, Automation and Systems Engineering. July 2009;345-347
- [2] Someya H, Sakamoto K, Yamamura M. Biologically-implemented genetic algorithm for protein engineering [C]// Proceedings of the 11th Annual conference on Genetic and evolutionary computation. July 2009;233-240
- [3] Maeda Y, Kajihara Y. Automatic generation method of twelve tone row for musical composition used genetic algorithm[C]// 2009 IEEE International Conference on Fuzzy Systems. August 2009;963-968
- [4] Goldberg D E. Genetic Algorithms in search, optimization and Machine Learning [M]. Addison Wesley Publishing Company, 1989;89-166
- [5] Zhao Han, Gao Xian-sheng, Zhu Ling-yun. Virus Co-evolution Genetic Algorithm Based on Niche Technology [C]// Proceedings of 2006 IEEE International Conference on Information Acquisition. August 2006;894-899
- [6] 李阳阳,焦李成.量子克隆遗传算法[J].计算机科学,2007,34(11);147-149
- [7] Zhang Jing-jun, Shang Yan-min, Gao Rui-zhen, et al. An Improved Genetic Algorithm Based on J1 Triangulation and Fixed Point [C]// Proceedings-2009 2nd IEEE International Conference on Computer Science and Information Technology. ICCSIT. August 2009;138-142
- [8] Zhang Jing-jun, Shang Yan-min, Gao Rui-zhen, et al. An improved multi-objective adaptive niche genetic algorithm based on pareto front[C]//2009 IEEE International Advance Computing Conference. March 2009;300-304
- [9] Kubota N,Shimajima K,Fukuda T. The Role of Virus Infection

in Virus-Evolutionary Genetic Algorithm[C]//Proceedings of the IEEE Conference on Evolutionary Computation, ICEC. May 1996;182-187

[10] Kubota N, Arakawa T, Fukuda T, et al. Fuzzy manufacturing scheduling by virus-evolutionary genetic algorithm in self-organizing manufacturing system[C]//Proceedings of the 1997 6th IEEE International Conference on Fussy Systems, FUZZ-IEEE'

97. Part 1 (of 3);1283-1288

[11] Kubota N, Fukuda T. Schema representation in virus-evolutionary genetic algorithm for knapsack problem [C]//Proceedings of the IEEE Conference on Evolutionary Computation, ICEC. May 1998;834-839
 [12] 焦李成,杜海峰,刘芳,等. 免疫优化计算学习与识别[M]. 北京: 科学出版社,2006

(上接第 228 页)

性能较佳。

实验二 测试式(11)中 α 取不同值对推荐准确度的影响,试验中 α 分别取 0.7,0.8,0.9,如图 5 所示。

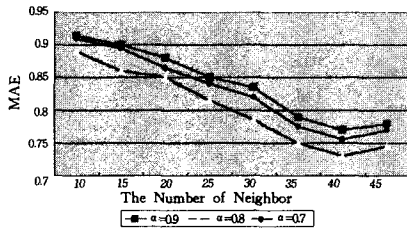


图 5 不同 α 取值对推荐准确度的影响图

从图 5 可以看出,引入时间遗忘函数、用户偏好度、用户特征向量能较大地提高推荐准确度,尤其当邻居数目较大时,准确度提高更为明显。由此可见,综合考虑时间效应、用户偏好有效地突出了用户近期访问数据对生成推荐的重要性,结合用户特征向量,使推荐结果能更好地反映用户的当前兴趣。同时可以看出, α 的取值对推荐准确度也有比较大的影响, α 取值过大或过小都会对推荐效果产生负面影响,当 $\alpha=0.8$ 时,推荐结果较佳。

实验三 对比本文提出的改进算法、传统的协同过滤算法、基于时间改进型协同过滤算法的推荐准确度。

从图 6 可以看出,本文算法与其他算法相比,推荐精度有一定程度的提高,尤其当邻居数目较大时,准确率提高更为明显。但最近邻居的选取对算法准确率有比较大的影响,其值过大或者过小都会对推荐效果产生负面影响。本实验取邻居数目为 35 到 45 之间比较合适。

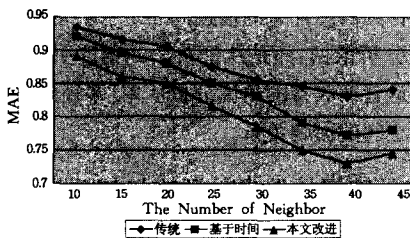


图 6 不同邻居数目对推荐准确度的影响图

实验四 测试本文提出的改进算法与传统的协同过滤算法、基于时间改进型协同过滤算法的时间性能($\alpha=0.8$,邻居数目=35,连续做 50 次实验平均结果)。

表 1 本文改进算法与其他协同过滤在效率的比较

推荐策略	MCT
传统的协同过滤算法	15006.32(ms)
基于时间的协同过滤算法	15194.45(ms)
本文改进协同过滤算法	15206.65(ms)

从表 1 可以看出,本文提出的改进算法在效率上并没有明显的差距。从图 3 可以看到,本文算法采用了离线和在线

计算相结合的方式,虽然引入了用户特征向量上的相似度计算和用户偏好度计算,二者的算法复杂度分别为 $O(m)$ 和 $O(n^2)$,但是用户数目 n_1 和用户评价过商品数目 n_2 都相对较小,所以对推荐的效率也没有明显的影响。

结束语 针对协同过滤算法中存在的用户兴趣变化、用户项目间联系单一和缺乏用户特征信息的问题,本文提出了适应用户兴趣变化和基于用户特征的个性化推荐算法,在协同过滤算法的基础上,引入了时间遗忘函数、黏度函数、用户特征向量,改进了寻找最近邻居集合过程,综合考虑了时间效应、用户特征和用户偏好程度,从而提高了寻找最近邻居的准确度。实验表明,本文提出的算法在一定程度上克服了传统协同过滤算法的弊端,能较大提高推荐的准确度,尤其在参数设置得当的情况下,算法性能有明显提高。

参考文献

- [1] 李聪,梁昌勇,董珂. 基于项目类别相似性的协同过滤推荐算法[J]. 合肥工业大学学报:自然科学版,2008(03)
- [2] 彭德巍,胡斌. 一种基于用户特征和时间的协同过滤算法[J]. 武汉理工大学学报,2009(03)
- [3] 杨怀珍,丛晓琪,刘枚莲. 于时间加权的个性化推荐算法研究[J]. 计算机工程与科学,2009(06)
- [4] 郑先荣,曹先彬. 线性逐步遗忘协同过滤算法的研究[J]. 计算机工程,2007(06)
- [5] Gong Songjie, Cheng Guanghua. Mining User Interest Change for Improving Collaborative Filtering[C]//Intelligent Information Technology Application 2008. Second International Symposium. Volume 3, Dec. 2008;24-27
- [6] 赵智,冯卓楠. 改进的基于相关相似性的协同过滤推荐算法[J]. 长春工业大学学报:自然科学版,2006(04)
- [7] Xia Weiwei, He Liang, Ren Lei, et al. A new collaborative filtering approach utilizing item's popularity[C]//Industrial Engineering and Engineering Management. IEEE International Conference, Dec. 2008;1480-1484
- [8] Su Xiaoyuan, Khoshgoftaar T M, Greiner R. A Collaborative Filtering Algorithm Based on Variance Analysis of Attributes-Value Preference[C]//IEEE/WIC/ACM International Conference. Volume 1, Dec. 2008;633-639
- [9] Dai Y, Ye Hongwu, Gong Songjie. Personalized Recommendation Algorithm Using User Demography Information. Knowledge Discovery and Data Mining [C]// Second International Workshop. Jan. 2009;100-103
- [10] Gong Song-jie, Ye Hongwu. Combining Memory - Based and Model-Based Collaborative Filtering in Recommender System [C]//Circuits, Communications and Systems. Pacific-Asia Conference. 2009;690-693
- [11] 张国忠. 遗忘规律的探索与运用[J]. 继续教育研究,2002(06)
- [12] 王元明,熊伟. 异常数据的检测方法[J]. 重庆工学院学报:自然科学版,2009,23(2);86-89