文章编号:1674-7070(2013)06-0553-04

基于均分法的小生境遗传算法

赵远东[1]　汪怡[2]

摘要

为了避免遗传算法种群中个体过早陷入局部最小,在以往随机初始种群的基础上提出一种均分法,使得初始种群随机平均地分为若干个子种群,形成小生境,这样既维持了种群的多样性,也使得种群中的个体不会过早出现早熟现象,更提高了算法的收敛速度.同时采用了自适应技术控制交叉和变异的概率,使得算法能更快速地找到最优解.仿真结果表明,与传统的遗传算法优化 RBF 网络相比较,新算法的迭代次数更少,精度更高,大大提高了收敛速度.

关键词

小生境技术;快速遗传算法;自适应技术

中图分类号 TP30L

文献标志码 A

0　引言

遗传算法[1](GA)由美国 Holland 教授于 1975 年根据生物进化理论和遗传变异理论提出,是一种建立在自然选择和群体遗传学机理基础上的并行的随机性全局寻优算法.虽然遗传算法有较强的全局搜索能力,但是也存在自身的局限性.由于初始种群是随机产生的,在寻找最优解时搜索方向不断发生变化,遗传算法搜索到全局最优解的代价较大.本文引用了小生境技术[2],在传统的小生境遗传算法的基础上提出了一种改进的方法,即在初始种群时采用均分法,使得初始种群能够随机平均地分为若干个子种群,然后在各子种群内部独立进行遗传算子操作,这样很大程度上避免了种群中的个体过早地陷入局部最小.改进后的算法(Niched Genetic Algorithms, NGA),可以更好地保持解的多样性,同时具有很高的全局寻优能力.本文在初始种群的时候,还引入了广义海明距离,使得各个个体尽可能均匀分布在整个解空间中,从而更快地收敛到全局最优解.将此方法与 RBF(Radical Basis Function)网络结合,对城市天燃气负荷进行建模预测,与传统的 RBF 网络相比,收敛速度更快,迭代次数更少.

1　改进的小生境遗传算法

本文在初始种群时引用广义海明距离[3],使各个个体尽可能地分散在解空间中,然后采用均分法,将一个种群随机平均地分为若干个小的子种群,即为小生境.经过一轮遗传算子的操作后,选出每个子种群中适应度高的个体组成新种群,再使用均分法,随机平均地分为若干个子种群,直到所有个体的适应度值都比较高为止.

设个体是以 b 为基的字符串,个体的长度为 k,种群的规模大小为 N,则要求入选种群的所有个体之间的广义海明距离 H_G 必须满足:

$$H_G \geqslant (k-b), \quad i \neq j, \tag{1}$$

其中:i,j 分别为 2 个个体,$i,j = 1,2,\cdots,N$;b 为常数,编码形式不同,b 的取值也不同,本文采用实数编码,$b \geqslant 2$.对于一个为 a 的基,长度为 k 的字符串,共有 a^k 个编码串.在这些编码串中,它们之间的海明距离 $\geqslant k-b$ 的编码串就有 $a^{k-(k-b)+1}$ 个.广义海明距离越大,字符间所包含的不相同模式就越多,种群中的模式就越多.然后采用均分法,即

$$c = \frac{N}{2^n}, \quad n = 1,2,3,\cdots,10, \tag{2}$$

收稿日期 2012-04-10

资助项目 公益性行业(气象)科研专项(GYHY201206070)

作者简介

赵远东,男,副教授,研究方向为人工智能与神经网络. ydzhao9@163.com

1　南京信息工程大学　电子与信息工程学院,南京,210044

2　南京信息工程大学　信息与控制学院,南京,210044

其中 c 为子种群,N 为种群数量,经过实验,n 取 3 最为合适.

2 遗传算子的设计

本文采用轮盘赌选择方法[4],每个个体进入下一代的概率取决于该个体的适应度值与整个种群适应度值总和的比例,适应度值越高的个体,被选中的可能性就越大,进入下一代的概率也就越大. 每个个体适应度值为 F_i,则该种群的适应度总和为

$$\sum_{i=1}^{N} F_i, \quad i = 1,2,3,\cdots,10, \tag{3}$$

每个个体被选中的概率为

$$P_i = F_i \Big/ \sum_{i=1}^{N} F_i, \quad i = 1,2,3,\cdots,N. \tag{4}$$

由于已经产生了子种群 c,所以交叉操作就是在这些已经产生的子种群内部独立进行. 交叉操作采用单点交叉方式(图 1).

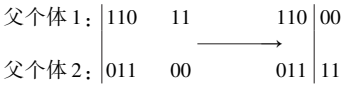


图 1 单点交叉

Fig. 1 Single point crossover

交叉操作[5]是以一定的概率 p_c 进行的,p_c 的大小反映了个体更新的快慢,p_c 过大会导致优良个体被过多破坏,一般为 0.4,在此,做个改进,采用自适应的方式来调节 p_c.

$$p_c = \begin{cases} p_{c,max} - \dfrac{(p_{c,max} - p_{c,min})(f_{c,max} - f_{avg})}{f_{max} - f_{avg}}, & f_{c,max} > f_{avg}, \\ p_{c,max}, & f_{c,max} \leqslant f_{avg}, \end{cases} \tag{5}$$

其中,f_{max} 表示群体中最大适应度,f_{avg} 表示每一代群体的平均适应度,$f_{c,max}$ 表示执行交叉操作的 2 个个体中较大的适应度. 另外,为了减少遗传算法中的退化现象,加快收敛的速度,在交叉操作中将父子 2 代 4 条染色体中适应度值高的 2 条保存到下一代中.

为了改善遗传算法的局部搜索能力以及维持群体的多样性,在遗传算法中要使用变异算子(图 2).



图 2 变异操作

Fig. 2 Mutation operation

变异操作以一定的概率 p_m 进行,一般为 0.1. 变异概率过大会增大搜索的随机性,过小则难以达到变异的目的. 本文采用如下自适应变异概率:

$$p_m = \begin{cases} p_{m,max} - \dfrac{(p_{m,max} - p_{m,min})(f_{max} - f_{m,max})}{f_{max} - f_{avg}}, & f_{m,max} > f_{avg}, \\ p_{m,max}, & f_{m,max} \leqslant f_{avg}, \end{cases} \tag{6}$$

其中,f_{max} 表示群体中最大的适应度,f_{avg} 表示每一代群体的平均适应度,$f_{m,max}$ 表示要执行变异操作的个体的适应度. 另外,由于变异操作可能会破坏种群中的最优解,使得变异结果更坏,因此,在变异前先保存原来的种群,变异后通过比较变异前后的适应度值函数来决定是否使用新个体,从而保证最优解不会被淘汰. 变异操作同样在子种群内部独立进行.

3 算法流程与复杂度分析

RBF 网络[6]输出权重(w_i)、隐含层单元中心(c_i)和宽度(σ_i)对网络性能的影响很大. 传统的 RBF 神经网络采用经验法或梯度下降法选择这些参数,就使得算法的精度不高、收敛速度慢,因此本文将改进后的小生境遗传算法应用在 RBF 神经网络的参数选择上. 算法流程如图 3 所示.

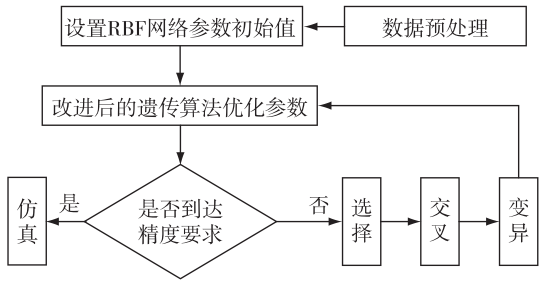


图 3 算法流程

Fig. 3 Flowchart of the proposed genetic algorithm

相比于传统的遗传算法,只是在初始种群时使用的均分法,所以只要分析式(2)的计算复杂度即可. 在式(2)中 n 的取值范围为 1 ~ 10,当 $n = 1$ 时,$c = \dfrac{N}{2}$,$T(n) = 1$,当 $n = 2$ 时,$c = \dfrac{N}{2^2}$,$T(n) = 2$,依此类推,所以最大计算复杂度[7]为

$$T(n) = \sum_{i=1}^{10} \frac{N}{2^n} = O(n), \tag{7}$$

小于传统的遗传算法的时间复杂度 $O(n^2)$.

根据文献[8],采用某城市 1—8 月天燃气的实时数据作为训练样本(表 1),9—11 月的实时数据作为预测样本(表 2).

表 1　训练样本

Table 1　Basic information of training sample

月份	平均气温/℃	锅炉负荷参数	空调负荷参数	发展趋势参数	月度负荷/(10^8 m^3)
1	4.3	1.0	1.0	2.804	465.490 6
2	5.6	2.0	1.5	4.951	633.592 8
3	9.5	1.5	1.0	6.744	608.868 9
4	15.8	1.0	0.8	8.501	560.563 7
5	20.7	0.8	0.5	11.030	510.023 5
6	24.3	0.8	0.8	13.723	531.610 1
7	28.4	0.8	0.8	16.369	514.798 0
8	27.9	0.8	1.5	18.812	518.414 1

表 2　测试样本

Table 2　Basic information of test sample

月份	平均气温/℃	锅炉负荷参数	空调负荷参数	发展趋势参数
9	23.4	1.0	1.5	21.052 2
10	18.3	1.0	1.0	23.752 9
11	12.4	1.0	0.5	28.323 9

仿真结果表明,用改进后的小生境遗传算法优化 RBF 网络的参数比标准的 RBF 网络在收敛速度上有明显优势[9]. 在同样的精度范围(10^{-3})内,改进后的算法(图 4)只要 50 次迭代即可达到目标精度,其标准的 RBF 算法(图 5)要 75 次迭代才能收敛. 另外,前者的误差明显小于后者,且时间复杂度明显小于后者. 利用遗传算法的全局搜索能力,并加以改进后,优化 RBF 网络的参数,克服了神经网络收敛速度慢、迭代次数多等缺点,加快了算法的收敛速度,提高了算法的精度.

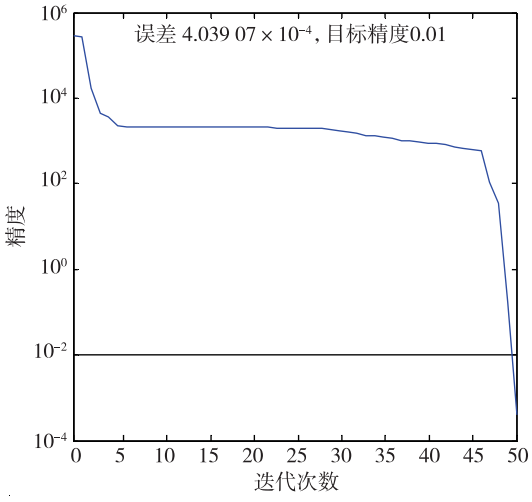


图 4　改进后的小生境遗传算法预测城市燃气负荷

Fig. 4　City gas load predicted by improved niche genetic algorithm

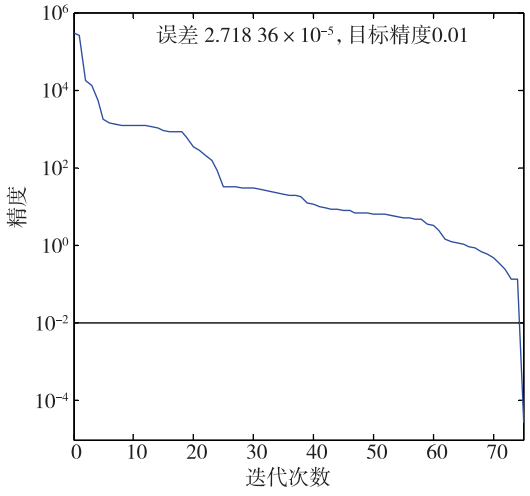


图 5　标准 RBF 算法预测城市燃气负荷

Fig. 5　City gas load predicted by standard RBF algorithm

4　结束语

本文提出的均分法,将初始种群平均地分为若干个子种群,这样可以避免适应度高的个体过度集中,有效地避免了早熟现象的发生,维持了种群的多样性. 从实例仿真可以看出,在目标精度相同的情况下,本文提出的改进算法收敛速度更快,时间复杂度更小、精度更高,是一种较优的算法.

参考文献

References

[1] 雷英杰,张善文,李续武,等. Matlab 遗传算法工具箱及应用[M]. 西安:西安电子科技大学出版社,2005
LEI Yingjie, ZHANG Shanwen, LI Xuwu, et al. Matlab genetic algorithm toolbox and its application[M]. Xi'an: Xidian University Publishing House, 2005

[2] 乔佩利,郑林,马丽丽. 一种小生境遗传算法研究[J]. 哈尔滨理工大学学报,2011,16(1):90-93
QIAO Peili, ZHENG Lin, MA Lili. Research on a niche

genetic algorithm[J]. Journal of Harbin University of Science and Technology, 2011, 16(1): 90-93
[3] 吴斌,吴坚,涂序彦. 快速遗传算法研究[J]. 电子科技大学学报,1999,28(1):49-53
WU Bin, WU Jian, TU Xuyan. Research of fast genetic algorithm[J]. Journal of University of Electronic Science and Technology of China, 1999, 28(1): 49-53
[4] 陈友文. 一种改进选择算子和基于小生境的遗传算法[J]. 计算机与数字工程,2009,37(6):21-24
CHEN Youwen. An improved selection operator of genetic algorithm based on niches[J]. Computer & Digital Engineering, 2009, 37(6): 21-24
[5] 张琰,谢娟英. 改进梯度算子的小生境遗传算法[J]. 计算机工程与科学,2009,31(9):71-73
ZHANG Yan, XIE Juanying. A new niche genetic algorithm based on improved gradient operators[J]. Computer Engineering and Science, 2009, 31(9): 71-73
[6] 代丽娴. 天燃气负荷预测的建模与应用研究[J]. 计算机仿真,2011,38(10):180-183
DAI Lixian. Modeling and application research on gas load forecasting[J]. Computer Simulation, 2011, 38(10): 180-183
[7] 张德富. 算法设计与分析[M]. 北京:国防工业出版社,2009
ZHANG Defu. Design and analysis of algorithms[M]. Beijing: National Defense Industry Press, 2009
[8] 高良军. 基于 BP 神经网络的城市燃气月度负荷预测研究[J]. 浙江海洋学院学报:自然科学版,2008,27(2):227-231
GAO Liangjun. The study on the monthly city gas load predicting based on BP neural network[J]. Journal of Zhejiang Ocean University: Natural Science, 2008, 27(2): 227-231
[9] 张德丰. Matlab 神经网络仿真与应用[M]. 北京:电子工业出版社,2009
ZHANG Defeng. Matlab neural network simulation and its application[M]. Beijing: Publishing House of Electronics Industry, 2009

Niche genetic algorithm research based on average method

ZHAO Yuandong[1] WANG Yi[2]

1 School of Electronic & Information Engineering, Nanjing University of Information Science & Technology, Nanjing 210044

2 School of Information and Control, Nanjing University of Information Science & Technology, Nanjing 210044

Abstract In order to avoid the population premature into local minimum, a new averaging method based on a random initial population was introduced into the genetic algorithm. The initial population is stochastically divided into several sub populations to form niches, with the purpose to maintain the population diversity, make the individuals in a sub population not display prematurity phenomenon, and improve the convergence speed of the algorithm as well. The adaptive technique is employed to control the crossover and mutation probability, therefore the algorithm can find the optimal solution quickly. Simulation results show that, compared with traditional RBF neural network optimized by genetic algorithm, the new algorithm is characterized by less iterations, higher precision, and greatly improved convergence speed.

Key words niche technique; genetic algorithm; adaptive technique