

信息熵动态变异概率RNA遗传算法

王康泰^{1,2}, 王宁¹

(1. 浙江大学工业控制技术国家重点实验室, 浙江杭州 310027; 2. 杭州电子科技大学电子信息学院, 浙江杭州 310018)

摘要: 约束优化问题是科学研究和工程应用的热点和难点. 受生物RNA分子遗传信息表达机制和信息熵概念的启发, 本文提出了一种信息熵动态变异概率RNA遗传算法来求解这类问题, 算法采用碱基序列的个体编码方式, 并用RNA分子再编码和蛋白质折叠操作代替传统遗传算法的交叉操作, 在变异概率的设置中, 借鉴信息熵对系统有序程度度量的概念, 根据当前种群个体每一位的碱基分布情况对变异概率进行自适应调整. 测试函数的仿真结果表明所提出的算法具有收敛速度快、搜索精度高的特点. 将该算法用于求解短期汽油调合调度问题, 能得到比其他几种算法更高的调合利润.

关键词: RNA遗传算法; 信息熵; 动态变异概率; 汽油调合短期调度

中图分类号: TP273 **文献标识码:** A

A RNA genetic algorithm with entropy based dynamic mutation probability

WANG Kang-tai^{1,2}, WANG Ning¹

(1. State Key Laboratory of Industrial Control Technology, Zhejiang University, Hangzhou Zhejiang 310027, China;

2. College of Electronics & Information Engineering, Hangzhou Dianzi University, Hangzhou Zhejiang 310018, China)

Abstract: The constrained optimization problem becomes a focus and difficulty in science and engineering filed. Inspired by the expression of bio-genetic information of RNA molecular and entropy concept, a RNA genetic algorithm with entropy based dynamic mutation probability (edmp-RGA) is proposed. The algorithm adopts nucleotide base encoding, and RNA recoding operation and protein folding operation are designed to replace the conventional crossover operation. In the algorithm, the values of mutation probability are decided by nucleotide base distribution of the current bits of population. The numerical experiments on four benchmark functions show the effectiveness of the proposed algorithm. The solution to the short-time gasoline blending scheduling problem shows that the proposed algorithm gains a higher profit.

Key words: RNA genetic algorithm; entropy; dynamic mutation probability; short-time gasoline blending scheduling

1 引言(Introduction)

遗传算法是模拟达尔文的生物进化论而产生的一种智能优化算法, 具有较强的全局寻优能力和通用性. 遗传算法由美国密歇根大学的John H. Holland教授于20世纪70年代提出^[1], 其后De Jong和Goldberg等人对它进行了完善和补充^[2-3]. 尽管遗传算法在许多领域得到了成功的应用, 但是它仍然存在着一些缺点和不足, 例如局部搜索能力差, 进化过程中种群个体快速趋同而使得算法容易早熟收敛等. 传统遗传算法用染色体代表个体, 并采用双亲单点或多点交叉操作方式, 这些只是对生物进化机理的简单模拟, 没有体现出自然遗传信息表达过程. 随着生物科学技术的不断进步, 人们对RNA分子的结构和遗传信息表达机理的认识也越来越深刻. 为了提高遗传算法的寻优性能, 基于碱基编

码和生物分子操作的遗传算法受到了人们的极大关注. 将个体通过碱基编码成生物分子, 借鉴生物分子操作来提高遗传算法的搜索效率和寻优性能^[4-10].

信息熵是系统有序程度的度量. 在信息论中, 信息熵越高表明信号不确定性越大, 所含信息量越大^[11]. 对系统而言, 一个系统越混沌无序, 则该系统信息熵越大. 与之相反, 一个系统有序则信息熵相对较小^[12-13]. 对于遗传算法的变异概率而言, 如果种群个体多样性好, 则只需要较小的变异概率; 而对于种群多样性差的情况, 则需要较大的变异概率才能使算法摆脱局部极值. 借鉴信息熵对系统有序程度的度量原理, 将其用于遗传算法种群离散程度度量, 并以此为依据来动态调整变异概率. 如果种群信息熵大, 则说明种群多样性较好, 可以设置较小的变异概率, 如果信息熵小, 则说明种群多样性差, 需要设

置较大的变异概率。

本文受RNA生物分子和遗传信息表达过程启发, 在文献[9]的基础上, 结合信息熵概念提出了一种信息熵动态变异概率RNA遗传算法(edmp-RGA)。通过对遗传算法编码和操作以及变异概率进行改进, 以提高遗传算法的寻优性能。

2 信息熵动态变异概率RNA遗传算法 (RNA genetic algorithm with entropy based dynamic mutation probability)

2.1 编码(Encoding)

RNA分子由4种核苷酸构成, 不同的核苷酸之间的区别主要是所含的碱基不同, 因此, 采用核苷酸所含的碱基类型来代表不同的核苷酸。在本文提出的edmp-RGA算法中, 为了便于借鉴RNA生物分子操作, 采用基于碱基的编码方式。RNA的碱基主要有4种, 分别为: 腺嘌呤(A)、尿嘧啶(U)、鸟嘌呤(G)和胞嘧啶(C)。个体编码空间可以表示为 $E = \{A, U, G, C\}^l$, 其中 l 表示RNA链的长度。edmp-RGA采用类似的四进制编码 $E' = \{0, 1, 2, 3\}^l$ 来表示 $E = \{A, U, G, C\}^l$ 。个体编码通常采用的对应关系为: CGAU \leftrightarrow 0123, 即0对应碱基C, 1对应碱基G, 2对应碱基A, 3对应U。生物分子中每3个碱基决定了一个组成蛋白质的基本单位——氨基酸, 因此可以将个体中3个连续的碱基当成一个氨基酸单元并进行相应的生物操作。由于edmp-RGA的交叉操作借鉴了蛋白质分子的生物操作, 因此要求个体的碱基长度取为3的整数倍。

2.2 选择和交叉(Selection and crossover)

选择操作根据当代种群个体适应度函数值, 选择适当个体组成下一代种群。通常, 适应度函数值较高的个体拥有更大的生存机会。本文算法采用了轮盘赌选择方法, 为了避免算法由于失去已经得到的最优值而出现退化现象, 算法中还采用了精英保留策略。

交叉操作是遗传算法中最重要的一个操作。为了更好地利用当前个体的有用信息使算法有更高的搜索效率和搜索性能, 受生物分子从RNA到蛋白质的生物信息表达过程的启发, 文献[9]设计了新的交叉操作——RNA再编码操作和蛋白质折叠操作。具体过程描述如下:

1) RNA再编码操作。

在生物学里, 在从RNA翻译成蛋白质之前, RNA链通常会改变形态来修正复制误差, 这个修正过程就叫做RNA再编码^[14]。模拟这种生物现象, 文献[9]设计的RNA再编码操作如图1所示。

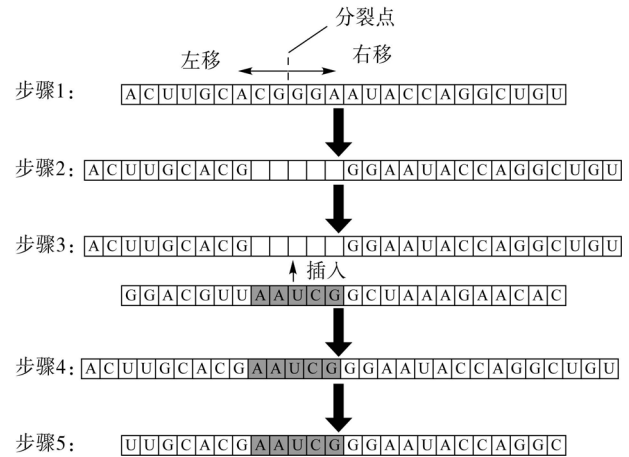


图 1 RNA再编码操作

Fig. 1 RNA recoding operator

当前的RNA链在一个随机点处分裂为两部分, 左半部分左移 n_1 步, 右半部分右移 n_2 步。随后将另一个随机选择的RNA链对应位置的碱基链插入该链中间的空缺处。为了保持个体长度一致, 超出个体长度的前面和后面的部分被切除。

2) 蛋白质折叠操作。

在生物学中, 当RNA链翻译成蛋白质链以后, 整个翻译过程并没有全部完成, 还会发生一些如蛋白质修饰和折叠等操作^[14]。蛋白质折叠可以分成两类: 链内自折叠和链间互相折叠。链间互相折叠发生在两个或两个以上的蛋白质分子之间, 而蛋白质自折叠则发生在单个蛋白质分子内部。文献[9]设计的蛋白质折叠操作如下:

折叠操作: 两个氨基酸链互相折叠靠近, 进而发生相互交流和相互影响。对于两个相互影响的碱基来说, 如果它们是相同的, 则分别用新的不同于它们的碱基替代。如果它们为不相同碱基, 则直接发生交换。互折叠操作示意图如图2所示。

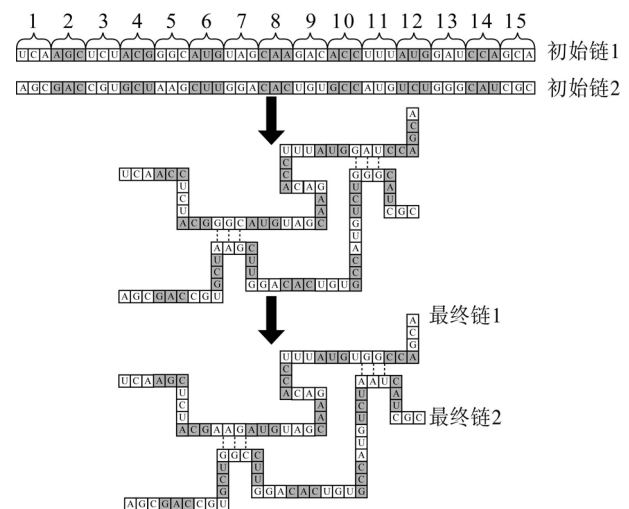


图 2 蛋白质互折叠操作

Fig. 2 Protein mutual-folding operator

由图2可知,两个链的第5个和第13个氨基酸是相互影响的,以第13个氨基酸为例进行分析.两个链的第13个氨基酸原来的碱基分别是GAU和GGG.这两个氨基酸的第一个碱基是相同的,均为G,则用两个不同的、随机产生的碱基U和A来代替.由于第2个和第3个核苷酸是不相同的,则直接交换它们的位置.

自折叠操作:一个RNA链内部的氨基酸链发生折叠从而发生相互影响和作用,其作用机理和互折叠是相同的,即如果发生作用的碱基相同则用随机的不同的碱基进行替换;如果碱基不同则直接交换这两个碱基位置.自折叠操作示意图如下图3所示.

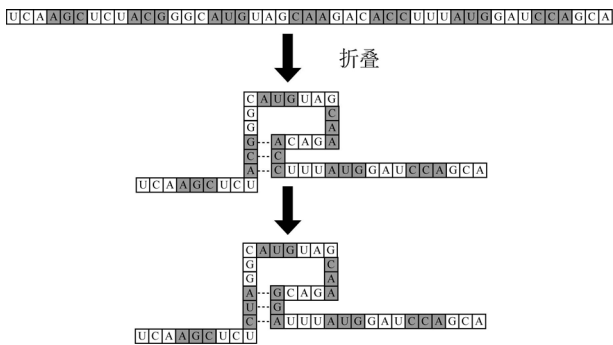


图3 蛋白质自折叠操作
Fig. 3 Protein self-folding operator

在整个交叉过程中,采用不替换策略,即新产生的个体直接加入种群中,而不替代父个体.执行交叉操作后,种群中个体数量会增加.为了使RNA再编码操作在遗传算法中发挥更大作用,对RNA再编码操作赋予比蛋白质折叠操作更大的概率.

为了减少计算量并控制种群规模在一定的范围,算法中设置蛋白质折叠操作只发生在没有执行RNA再编码操作的个体中.假设初始种群规模是 N_{pop} ,则执行完交叉操作以后,种群个体数量变成 $2N_{pop}$.

2.3 变异(Mutation)

变异操作可以提高遗传算法种群多样性,防止早熟收敛.变异概率是影响变异操作性能的主要因素.变异概率过小,种群中发生变异的位数过少,变异操作的作用就减弱.变异概率过大,则有可能使种群失去原有的优良个体,遗传算法会陷入一种盲目的跳变搜索状态,从而影响算法的收敛性.

为了提高遗传算法的性能,根据进化后期比进化初期需要更大变异概率的观点,研究者们提出了一些变异概率自适应设定方法^[15-16].但是这些方法都存在共同的弊端:变异概率的设定只考虑了进化代数因素,没有考虑实际种群中个体的分布情况.

借鉴信息熵对系统有序程度的度量原理,将其用于对RNA遗传算法种群离散程度度量,并以此为依据来动态调整变异概率.如果种群信息熵大,则说明

种群多样性较好,可以设置较小的变异概率,如果信息熵小,则说明种群多样性差,需要设置较大的变异概率.

假设种群中每个个体长度为 L ,种群中任意一个个体可以表示为 $S = s_1, s_2, \dots, s_L$,则定义当前种群个体第 i 位的信息熵为

$$H_i = -(N_i(A)P_i(A) \log_2(P_i(A)) + N_i(U)P_i(U) \times \log_2(P_i(U)) + N_i(G)P_i(G) \log_2(P_i(G)) + N_i(C)P_i(C) \log_2(P_i(C))), 0 < i \leq L, \quad (1)$$

式中: $P_i(X)$ 表示当前种群中所有个体第 i 位出现元素 X 的概率; $N_i(X)$ 表示当前种群中第 i 位出现元素 X 的个数.假设种群规模为Size,则上述信息熵公式可以变化为

$$H(i) = -\left(\frac{N_i^2(A)}{\text{Size}} \log_2\left(\frac{N_i(A)}{\text{Size}}\right) + \frac{N_i^2(U)}{\text{Size}} \times \log_2\left(\frac{N_i(U)}{\text{Size}}\right) + \frac{N_i^2(G)}{\text{Size}} \log_2\left(\frac{N_i(G)}{\text{Size}}\right) + \frac{N_i^2(C)}{\text{Size}} \log_2\left(\frac{N_i(C)}{\text{Size}}\right)\right), 0 < i \leq L. \quad (2)$$

从上面信息熵公式可以得到,当种群中该位的元素都相同时,该位离散程度最小,信息熵值最小等于零.而且,随着该位拥有的元素种类增多,各元素的数量增大,信息熵增大.根据种群中每一位的信息熵值,设置动态变异概率为

$$P_i = P_{\max} - \frac{P_{\max} - P_{\min}}{H_{\max} - H_{\min}} \times H_i, \quad (3)$$

式中: P_{\max} , P_{\min} 为常数表示变异概率的最大值与最小值,该取值决定了变异概率的变化范围; H_{\max} , H_{\min} 表示信息熵的最大值和最小值.假设动态概率的变化范围为 $[0.01, 0.2]$,即 $P_{\min} = 0.01$, $P_{\max} = 0.2$,种群规模 $\text{Size} = 80$, $H_{\max} = 38$ 时,动态变异概率与信息熵之间的关系曲线如图4所示.

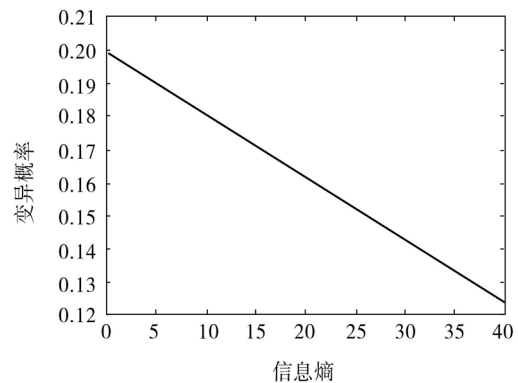


图4 动态变异概率曲线

Fig. 4 The curve of dynamic mutation probability

2.4 edmp-RGA实现过程(The procedure of edmp-RGA)

首先根据碱基编码初始化种群,然后采用交叉、

变异等操作对种群个体进行变换, 每一代结束后, 判断是否满足终止条件, 如不满足则选择适当的个体组成下代种群开始下代操作. 该算法的流程图如图5所示.

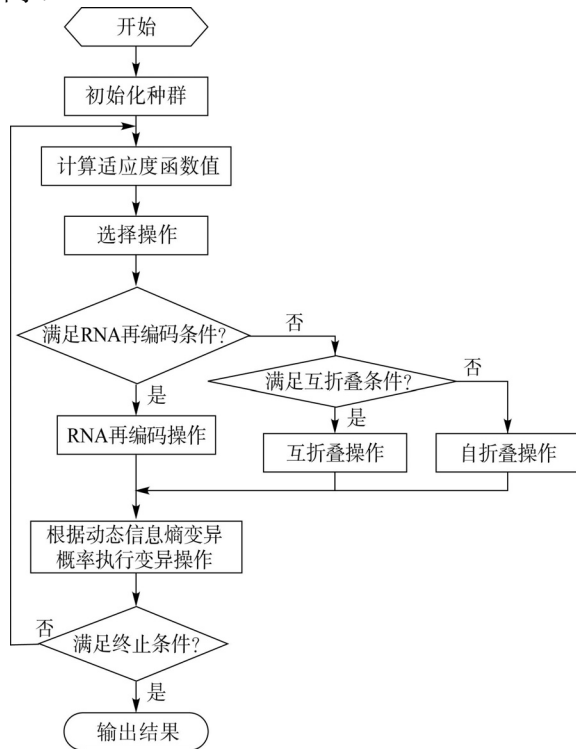


图 5 edmp-RGA算法流程图

Fig. 5 The flowchart of edmp-RGA

edmp-RGA实现步骤如下:

步骤 1 初始化一个含 N_{pop} 个个体的初始RNA种群.

步骤 2 计算个体的适应度函数值.

步骤 3 采用选择操作从初始种群或上代种群中复制 N_{pop} 个个体组成当代种群.

步骤 4 判断是否满足RNA再编码的条件, 如果满足, 则执行RNA再编码操作, 然后跳转到步骤6. 否则直接跳转到步骤5.

步骤 5 判断个体是否满足蛋白质互折叠条件, 如果满足则执行蛋白质互折叠操作. 否则执行蛋白质自折叠操作.

步骤 6 对所有交叉池中的个体根据变异概率执行变异操作.

步骤 7 判断是否满足算法终止条件, 如果满足则停止算法运行, 输出最终结果, 否则重复步骤2至步骤6.

3 测试函数寻优实验(Experiments on test problems)

3.1 测试函数(Test functions)

为了检验所提出算法的有效性, 选取4个典型的

具有约束的测试函数进行寻优实验. 函数的具体形式描述如下:

$$\begin{aligned} \min f_1(x) &= (x_1^2 + x_2 - 1)^2 + (x_1 + x_2^2 - 7)^2, \quad (4) \\ \text{s.t. } g_1(x) &\equiv 4.84 - (x_1 - 0.05)^2 - (x_2 - 2.5)^2 \geq 0, \\ g_2(x) &\equiv x_1^2 + (x_2 - 2.5)^2 - 4.84 \geq 0, \\ 0 &\leq x_1 \leq 6, 0 \leq x_2 \leq 6. \end{aligned}$$

最优解为

$$X^* = (2.246826, 2.381865), f_1(x^*) = 13.59085.$$

$$\min f_2(x) = x_1 + x_2 + x_3, \quad (5)$$

$$\begin{aligned} \text{s.t. } g_1(x) &\equiv 1 - 0.0025(x_4 + x_6) \geq 0, \\ g_2(x) &\equiv 1 - 0.0025(x_5 + x_7 - x_4) \geq 0, \\ g_3(x) &\equiv 1 - 0.01(x_8 - x_5) \geq 0, \\ g_4(x) &\equiv x_1x_6 - 833.33252x_4 - 100x_1 + \\ &\quad 83333.333 \geq 0, \\ g_5(x) &\equiv x_2x_7 - 1250x_5 - x_2x_4 + 1250x_4 \geq 0, \\ g_6(x) &\equiv x_3x_8 - x_3x_5 + 2500x_5 - 1250000 \geq 0, \\ 100 &\leq x_1 \leq 10000, 1000 \leq (x_2, x_3) \leq 10000, \\ 10 &\leq x_i \leq 1000, i = 4, \dots, 8. \end{aligned}$$

最优解 $f_2(x^*) = 7049.330923$ 位于

$$\begin{aligned} x^* &= (579.3167, 1359.943, 5110.071, 182.0174 \\ &\quad 295.5985, 217.9799, 286.4162, 395.5979). \end{aligned}$$

$$\begin{aligned} \min f_3(x) &= x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + \\ &\quad (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2 + \\ &\quad 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + \\ &\quad 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45, \quad (6) \end{aligned}$$

$$\begin{aligned} \text{s.t. } g_1(x) &\equiv 105 - 4x_1 - 5x_2 + 3x_7 - 9x_8 \geq 0, \\ g_2(x) &\equiv -10x_1 + 8x_2 + 17x_7 - 2x_8 \geq 0, \\ g_3(x) &\equiv 8x_1 - 2x_2 - 5x_9 + 2x_{10} + 12 \geq 0, \\ g_4(x) &\equiv -3(x_1 - 2)^2 - 4(x_2 - 3)^2 - 2x_3^2 + \\ &\quad 7x_4 + 120 \geq 0, \\ g_5(x) &\equiv -5x_1^2 - 8x_2 - (x_3 - 6)^2 + 2x_4 + 40 \geq 0, \\ g_6(x) &\equiv -x_1^2 - 2(x_2 - 2)^2 + 2x_1x_2 - 14x_5 + x_6 \geq 0, \\ g_7(x) &\equiv -0.5(x_1 - 8)^2 - 2(x_2 - 4)^2 - 3x_5^2 + \\ &\quad x_6 + 30 \geq 0, \\ g_8(x) &\equiv 3x_1 - 6x_2 - 12(x_9 - 8)^2 + 7x_{10} \geq 0, \\ -10 &\leq x_i \leq 10, i = 1, \dots, 10. \end{aligned}$$

最优解 $f_3(x^*) = 24.3062$ 位于

$$\begin{aligned} x^* &= (2.171996, 2.363683, 8.773926, 5.095984, \\ &\quad 0.9906548, 1.430574, 1.321644, 9.828726, \\ &\quad 8.280092, 8.375927). \end{aligned}$$

$$\min f_4(x) = -\frac{|\sum_{i=1}^n \cos^4(x_i) - 2 \prod_{i=1}^n \cos^2(x_i)|}{\sqrt{\sum_{i=1}^n ix_i}}, \quad (7)$$

$$\text{s.t. } \prod_{i=1}^n x_i \geq 0.75, \quad \prod_{i=1}^n x_i \leq 7.5n, \\ 0 \leq x_i \leq 10, \quad i = 1, \dots, n.$$

目前已知最优解为 $f_4(x^*) = -1.1314$, 位于 $x^* = (3.1520, 9.4137, 6.2666, 4.6805, 3.1172, 3.1085, 3.0997, 3.0910, 3.0821, 3.0733, 0.2080, 3.0554, 0.2003, 0.1965, 0.1929, 0.1894, 0.1859, 0.1826, 0.1793, 0.1762)$.

3.2 计算结果与分析(Computational results and analysis)

edmp-RGA算法的参数设置为 $N_{\text{pop}}=80, G_{\text{max}}=1000, P_{\text{RNA-recoding}}=0.8, P_{\text{mutual-folding}}=0.1, P_{\text{self-folding}}=0.1$. 为了获取算法的统计性能, 对每个函数分别独立运行50次, 并将优化结果与文献 [4]和文献[17]的进行比较, 计算结果如表1-3所示.

表1为3种算法在50次独立运行中的成功次数比较, 表2为3种算法的平均进化代数比较, 表3为3种算法所得到的最优值比较. 定义算法找到最优解与已知最优解之间的相对误差 $\varepsilon = |(f - f^*)/f^*|$, 当 $\varepsilon < 1\%$ 表示搜索成功, 此时的迭代次数就表示该次运行找到最优解的代数.

从表1-3可以看出, 本文所提出的edmp-RGA算法在各方面都要明显优于小生境遗传算法, 在平均进化代数上也优于混合遗传算法(HGA).

为了进一步考察edmp-RGA所采用的动态变异概率在进化过程中的变化情况, 对函数 f_1 在初始和进化终止两个时期的变异概率进行对比试验. 图6为第1代时种群变异概率曲线. 图7为第1000代时种群变异概率曲线.

表 1 3种算法成功次数比较

Table 1 The success number of three algorithms

测试函数	小生境GA ^[17]	HGA ^[4]	edmp-RGA
f_1	29	50	50
f_2	23	50	50
f_3	41	50	50

表 2 3种算法平均进化代数比较

Table 2 The evolutionary number of three algorithms

测试函数	小生境GA ^[17]	HGA ^[4]	edmp-RGA
f_1	1000	80	295
f_2	4000	532	282
f_3	—	650	358

表 3 3种算法最优结果比较

Table 3 The best solution of three algorithms

测试函数	小生境GA ^[17]	HGA ^[4]	edmp-RGA
f_1	13.59085	13.5908	13.5908
f_2	7060.221	7049.248	7049.35
f_3	24.3725	24.3062	24.3120
f_4	-0.8011	-1.1314	-1.1039

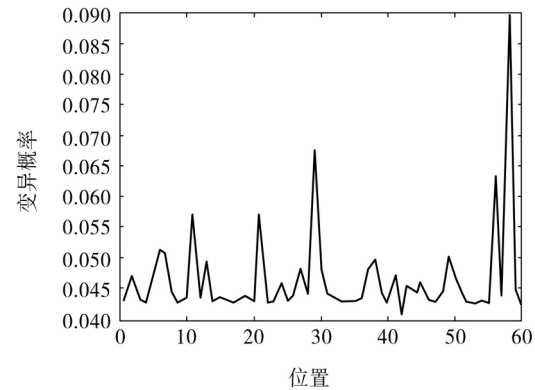


图 6 函数 f_1 第1代时的变异概率曲线

Fig. 6 Mutation probability curve of the first generation on function f_1

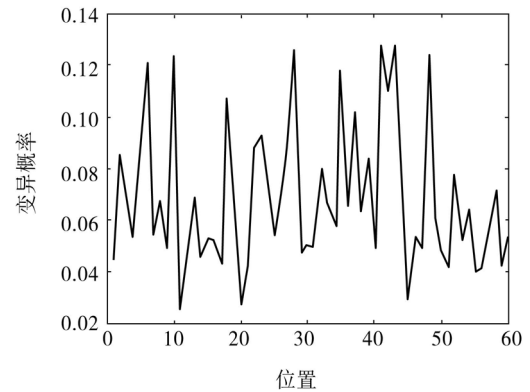


图 7 函数 f_1 第1000代时的变异概率曲线

Fig. 7 Mutation probability curve of the 1000th generation on function f_1

从图6和图7中可以看出, 在进化的初始阶段, 由于种群是随机产生的, 因此各个位置的熵差别较大, 变异概率大小差别也大, 但平均变异概率较小, 大概在0.045附近. 而在第1000代时, 由于种群多样性变差, 信息熵变小, 总体变异概率均值达到0.06左右, 符合预期设想.

4 短期汽油调合调度问题(Short-time gasoline blending scheduling problem)

4.1 问题描述(Problem description)

汽油调合过程, 就是在保证各种汽油产品质量指标的情况下, 根据现有的组分油和各种约束, 使得调合得到的产品汽油总利润最大. 评价油品质量最重要的油品质量指标研究法辛烷值(RON)、马达法辛

烷值(MON)和雷德蒸气压(RVP)计算公式如下^[18]:

$$P_{RON} = c_r^T c_x + \alpha_1 \left(c_r^T \text{diag}\{s\} c_x - \frac{(c_r^T c_x)(s^T c_x)}{e^T c_x} \right) + \alpha_2 \left(o_s^T c_x - \frac{(o^T c_x)^2}{e^T c_x} \right) + \alpha_3 \left(a_s^T - \frac{(a^T c_x)^2}{e^T c_x} \right), \quad (8)$$

$$P_{MON} = c_m^T c_x + \alpha_4 \left(c_m^T \text{diag}\{s\} c_x - \frac{(c_m^T c_x)(s^T c_x)}{e^T c_x} \right) + \alpha_5 \left(o_s^T c_x - \frac{(o^T c_x)^2}{e^T c_x} \right) + \frac{\alpha_6}{10000(e^T c_x)} \cdot \left(e_s^T - \frac{(a^T c_x)^2}{e^T c_x} \right)^2, \quad (9)$$

$$P_{RVP} = \left(\sum_{i=1}^n u_i (RVP_i)^{1.25} \right)^{0.8}, \quad (10)$$

式中: c_r 为每种组分的RON值, c_x 为组分流率, c_m 为每种组分的MON值, $s = c_r - c_m$, o 是每种组分的蜡含量, o_s 是蜡含量的平方, e 为单位列向量, a 为参调组分的芳烃含量, a_s 是芳烃含量的平方, α_i 为模型参数, P_{RON} 为产品汽油的RON值, P_{MON} 为产品汽油的MON值, P_{RVP} 为产品的RVP值, n 为参加该产品调合的组分个数, RVP_i 为组分 i 的RVP值, u_i 为组分 i 在产品中的比例含量。

在汽油调合过程中, 考虑下列约束^[19-20]:

1) 产品和调合组分之间的物料平衡约束

$$\sum_{n=1}^{N_p} V_{P_n} = \sum_{n=1}^{N_p} \sum_{m=1}^{N_{C_n}} V_{C_{n,m}}. \quad (11)$$

2) 调合产品的物料平衡约束

$$V_{P_{n,j}} = V_{P_{n,j-1}} + \sum_{m=1}^{N_{C_n}} V_{C_{n,m,j}}. \quad (12)$$

3) 组分储量约束

$$V_{C_{m,L}} \leq \sum_{n=1}^{N_p} V_{C_{n,m}} \leq V_{C_{m,U}}. \quad (13)$$

4) 产品需求量约束

$$V_{P_n} = M_{P_n}. \quad (14)$$

5) 产品辛烷值指标约束A

$$P_{RON,n} \geq RON_{n,\min}. \quad (15)$$

6) 产品辛烷值指标约束B

$$P_{MON,n} \geq MON_{n,\min}. \quad (16)$$

7) 产品蒸汽压指标约束

$$P_{RVP,n} \leq RVP_{n,\max}. \quad (17)$$

上述公式中: N_p 为调合产品的个数, N_{C_n} 为参与调合产品的组分个数, $V_{P_{n,j}}$ 为产品 n 第 j 天的产量, $V_{C_{n,m,j}}$ 为第 j 天参与调合产品 n 的组分 m 的体积, $V_{C_{m,L}}$ 为组分 m 的罐存储量下限, $V_{C_{m,U}}$ 为组分 m 的罐存储量上限, M_{P_n} 为产品 n 的市场需求量, $P_{RON,n}$ 为产品 n 的RON值, $RON_{n,\min}$ 为产品 n 的RON最小

值, $P_{MON,n}$ 为产品 n 的MON值, $MON_{n,\min}$ 为产品 n 的MON最小值, $P_{RVP,n}$ 为产品 n 的RVP值, $RVP_{n,\max}$ 为产品 n 的RVP最大值, V_{P_n} 为产品 n 的调合量, $V_{C_{n,m}}$ 为参与调合产品 n 的组分 m 的体积。

4.2 调合结果及分析(Results of gasoline blending and analysis)

采用本文提出的edmp-RGA对24小时内的短期汽油调合问题进行优化, 并与其他算法得到的结果进行比较. 相关成品油质量指标和组分油参数见文献[21-22].

使用本文的edmp-RGA得到的调合高级汽油的组分油优化配方如表4所示, 调合普通汽油的组分油优化配方如表5所示. 表6列出了基于此调合配方得到的产品汽油质量指标结果, 表7为最终总利润. 表4至表7中同时列了文献[19]的粒子群优化(PSO)算法和文献[23]的新型变异算子DNA遗传算法(nmDNA-GA)对该油品调合问题的优化结果, 以便进行比较。

从表7可以看出, 由edmp-RGA优化配方得到的产品汽油总利润为\$75916.9507, 比文献[19]的粒子群优化算法得到的结果\$74618高出了\$1298.9507, 这个效益是非常可观的. 同时, 该调合利润也大于nmDNA-GA得到的利润75112. 从表6的成品油质量指标可以看出, edmp-RGA与nmDNA-GA都取得了较好的卡边效果, 各项指标基本上都在边界附近, 优于文献[19]的PSO算法。

表 4 调合高级汽油组分油配方结果

Table 4 The consuming of components for higher quality gasoline

组分油种类	edmp-RGA	nmDNA-GA ^[23]	PSO ^[19]
重整油	3761.76	3842.5	5472.7
直馏油	1629.76	1546.3	1678
丁烷	219.72	220.2	225.3
催化油	4288.99	4331.1	1968
烷基化油	109.76	59.9	656

表 5 调合普通汽油组分油配方结果

Table 5 The consuming of components for lower quality gasoline

组分油种类	edmp-RGA	nmDNA-GA ^[23]	PSO ^[19]
重整油	5148.91	5000.2	2776.1
直馏油	2549.27	2488.30	2555.6
丁烷	150.69	145.8	136.3
催化油	161.14	165.6	2532
烷基化油	0	0	0

表6 产品汽油指标
Table 6 Oil product index

方法	汽油类型	RON	MON	RVP
edmp-RGA	普通汽油	88.51457	77.902265	10.79993
	高级汽油	91.83916	80.000004	10.79803
nmDNA-GA ^[23]	普通汽油	88.5000	77.8960	10.7823
	高级汽油	91.5390	80.0373	10.7449
PSO ^[19]	普通汽油	88.7	78.3841	10.8004
	高级汽油	91.5002	80.0406	10.8006

表7 调合总利润
Table 7 Profits of three algorithms

优化方法	edmp-RGA	nmDNA-GA ^[23]	PSO ^[19]
总利润/\$	75916.9507	75112	74618

5 结论(Conclusions)

本文提出一种信息熵动态变异概率RNA遗传算法,该算法在文献[9]的基础上在变异概率的设置中借鉴了信息熵的概念,根据当前种群个体每一位的碱基分布情况对变异概率进行动态调整.通过测试函数验证了所提出算法在收敛速度、搜索精度等方面的优越性能,将该算法用于求解汽油调合优化调度问题,结果表明使用该算法得到的调合利润优于其他方法.

参考文献(References):

- [1] HOLLAND J H. *Adaptation in Natural and Artificial Systems* [M]. Ann Arbor: University of Michigan Press, 1975.
- [2] DE JONG K A. *An analysis of the behavior of a class of genetic adaptive systems* [D]. Ann Arbor: University of Michigan, 1975.
- [3] GOLDBERG D E. *Genetic Algorithms in Search, Optimization and Machine Learning* [M]. Boston: Addison Wesley Longman, 1989.
- [4] TAO J L, WANG N. DNA double helix based hybrid genetic algorithm for the gasoline blending recipe optimization problem [J]. *Chemical Engineering & Technology*, 2008, 31(3): 440 – 451.
- [5] TAO J L, WANG N. DNA computing based RNA genetic algorithm with applications in parameter estimation of chemical engineering processes [J]. *Computers & Chemical Engineering*, 2007, 31(12): 1602 – 1618.
- [6] CHEN X, WANG N. A DNA based genetic algorithm for parameter estimation in the hydrogenation reaction [J]. *Chemical Engineering Journal*, 2009, 150(2/3): 527 – 535.
- [7] 陈霄, 王宁. 基于混沌DNA遗传算法的模糊递归神经网络建模[J]. *控制理论与应用*, 2011, 28(11): 1589 – 1594.
(CHEN Xiao, Wang Ning. Fuzzy recurrent neural network modeling based on chaos DNA genetic algorithm [J]. *Control Theory & Applications*, 2011, 28(11): 1589 – 1594.)
- [8] CHEN X, WANG N. Optimization of short-time gasoline blending scheduling problem with a DNA based hybrid genetic algorithm [J]. *Chemical Engineering and Processing*, 2010, 49(10): 1076 – 1083.
- [9] WANG K T, WANG N. A protein inspired RNA genetic algorithm for parameter estimation in hydrocracking of heavy oil [J]. *Chemical Engineering Journal*, 2011, 167(1): 228 – 239.
- [10] WANG K T, WANG N. A novel RNA genetic algorithm for parameter estimation of dynamic systems [J]. *Chemical Engineering Research & Design*, 2010, 88(11): 1485 – 1493.
- [11] SIGALOV D, SHIMKIN N. Cross entropy algorithms for data association in multi-target tracking [J]. *IEEE Transactions on Aerospace and Electronic Systems*, 2011, 47(2): 1166 – 1185.
- [12] WANG K H, YANG D Y, PEARN W L. Comparative analysis of a randomized N-policy queue: an improved maximum entropy method [J]. *Expert Systems with Applications*, 2011, 38(8): 9461 – 9471.
- [13] HSIAO B, CHERN C C, CHIU C R. Performance evaluation with the entropy-based weighted Russell measure in data envelopment analysis [J]. *Expert Systems with Applications*, 2011, 38(8): 9965 – 9972.
- [14] CLARK D P. *Molecular Biology: Understanding the Genetic Revolution* [M]. New York: Academic Press, 2005.
- [15] LIS J. Genetic algorithm with the dynamic probability of mutation in the classification problem [J]. *Pattern Recognition Letters*, 1995, 16(12): 1311 – 1320.
- [16] SERPELL M, SMITH J E. Self-adaptation of mutation operator and probability for permutation representations in genetic algorithms [J]. *Evolutionary Computation*, 2010, 18(3): 491 – 514.
- [17] DEB K. An efficient constraint handling method for genetic algorithms [J]. *Computer Methods in Applied Mechanics and Engineering*, 2000, 186(2/4): 311 – 338.
- [18] ZHANG Y, MONDER D, FORBES J F. Real-time optimization under parametric uncertainty: a probability constrained approach [J]. *Journal of Process Control*, 2002, 12(3): 373 – 389.
- [19] 张建明, 冯建华. 两群微粒群算法及其在油品调合优化中的应用 [J]. *化工学报*, 2008, 59(7): 1721 – 1726.
(ZHANG Jianming, FENG Jianhua. Gasoline blending recipe optimization based on two particle swarms optimization [J]. *Journal of Chemical Industry and Engineering*, 2008, 59(7): 1721 – 1726.)
- [20] ZHAO X, RONG G. Blending scheduling under uncertainty based on particle swarm optimization algorithm [J]. *Chinese Journal of Chemical Engineering*, 2005, 13(4): 535 – 541.
- [21] SINGH A, FORBES J F, VERMEER P J, et al. Model-based real-time optimization of automotive gasoline blending operations [J]. *Journal of Process Control*, 2000, 10(1): 43 – 58.
- [22] FORBES J F, MARLIN T E. Model accuracy for economic optimizing controllers: the bias update case [J]. *Industrial & Engineering Chemistry Research*, 1994, 33(8): 1919 – 1929.
- [23] 陈霄. DNA遗传算法及应用研究[D]. 杭州: 浙江大学, 2010.
(CHEN Xiao. *Research on DNA genetic algorithms and applications* [D]. Hangzhou: Zhejiang University, 2010.)

作者简介:

王康泰 (1980–), 男, 博士, 讲师, 从事智能优化算法、信号处理、嵌入式系统的研究, E-mail: kangtaiwang@gmail.com;

王宁 (1961–), 男, 教授, 博士生导师, 从事生物计算与智能系统设计、计算机控制与优化的研究, E-mail: nwang@iipc.zju.edu.cn, 通讯作者。