

# 二进制遗传算法编码长度的计算方法

朱会霞<sup>1</sup>, 郝志云<sup>1</sup>, 王福林<sup>2</sup>

(1. 辽宁工业大学 管理学院, 辽宁 锦州 121001; 2. 东北农业大学 工程学院, 黑龙江 哈尔滨 150030)

**摘要:** 针对现有文献二进制遗传算法编码长度选取存在的问题, 研究了二进制遗传算法编码长度确定的基本原理和具体实现过程, 给出了二进制遗传算法编码长度的计算公式, 使二进制遗传算法编码长度以公式的形式表述出来, 方便计算机编程实现。该公式可以根据决策变量的求解精度和区间长度准确确定所需的最小编码长度, 避免了因为编码长度过大, 搜索空间呈指数增加而造成效率低的问题, 也避免了因为编码长度太短而造成的离散误差问题, 提高了二进制遗传算法的优化效率。

**关键词:** 二进制遗传算法; 编码长度公式; 求解精度

**中图分类号:** TP301.6    **文献标识码:** A    **文章编号:** 1674-3261(2016)02-0138-03

## Calculation Method of Encoding Length for Binary Encoded Genetic Algorithm

ZHU Hui-xia<sup>1</sup>, HAO Zhi-yun<sup>1</sup>, WANG Fu-lin<sup>2</sup>

(1. Management College, Liaoning University of Technology, Jinzhou 121001, China;

2. School of Engineering, Northeast Agricultural University, Harbin 150030, China)

**Abstract:** There are problems in selecting the encoding length for binary encoded genetic algorithm in existing literature; the principles and the detailed realization procedure of determining the encoding length for binary encoded genetic algorithm are studied. The encoding length formula for binary encoded genetic algorithm is given, so the encoding length for binary encoded genetic algorithm is expressed in the form of formula which facilitates the realization of computer programming. The formula can determine the shortest encoding length precisely according to the solution precision and the interval length of decision variables. It avoids the problem of low efficiency for the sake of the encoding length being too long which makes the search space increase exponentially; it also avoids the problem of discrete errors because of the too short encoding length. Thus the optimization efficiency is improved for binary encoded genetic algorithm.

**Key words:** binary encoded genetic algorithm; encoding length formula; solution precision

遗传算法 (genetic algorithm, GA) 是用概率变迁规则对搜索方向进行引导的全局优化算法。因二进制编码方法符合最小字符集原则和有意义积木块编码规则<sup>[1-5]</sup>, 所能处理的模式数最多, 与生物染色体的组成类似, 便于用生物遗传理论来解释。编码简单, 基因表达细腻, 交叉、变异等遗传操作

易于实现, 操作过程清晰, 可以用模式定理对算法进行理论分析, 且已有理论证明, 低进制编码遗传算法在搜索效率和优化结果鲁棒性方面普遍优于高进制编码遗传算法<sup>[6-8]</sup>, 故二进制编码仍是用得最多、最广的编码方法, 几乎任何问题的优化都可以用二进制编码来表达<sup>[9-13]</sup>。

收稿日期: 2015-02-07

基金项目: 辽宁省社会科学规划基金项目 (L13DJY092), 辽宁工业大学校基金项目 (X2013023)。

作者简介: 朱会霞 (1978-), 女, 黑龙江佳木斯人, 讲师, 博士。

二进制遗传算法与实际应用问题结合的第一步是将决策变量映射为遗传空间中的基因串<sup>[14]</sup>，所以在使用二进制遗传算法时，首先要确定编码长度，编码长度直接影响算法搜索空间的大小和寻优能力，故本文提出一种计算二进制编码所需最小长度的方法，提高二进制遗传算法的性能。

## 1 现有编码长度确定方式存在问题

二进制遗传算法对连续型决策变量进行优化时，实际是对决策变量的一种离散化处理。二进制遗传算法需要频繁地编码和解码，不可避免地存在解空间与二进制域搜索空间的量化误差<sup>[15]</sup>。决策变量的定义域及计算所需的精度决定二进制遗传算法的编码长度。对于长度相同的区间，编码长度直接影响到决策变量可识别的精度。现有二进制遗传算法的文献，编码长度一般有两种确定方法：一种是在个人经验的基础上随机选取；另一种是预先给一个较大的编码长度值，根据实验结果进行反复调试，逐渐缩小数值的试验方法确定。

对于根据经验指定的这种方法，容易产生两个问题：

(1)指定的二进制编码长度较短。当指定的二进制编码长度较短时，二进制编码所能表示的变量精度就会较低，可能达不到精度要求，就会产生有限的离散点阵，在从二进制域到实数域的解码过程中，因为映射误差而损失精度，导致接近最优值的个体被漏掉。如果被优化的目标函数对这种“损失”是敏感的，那么算法所求得的较优解的精度就会大打折扣，且易陷入局部极值点，造成算法早熟收敛。

(2)指定的二进制编码长度较大。解决精度较低的办法只能是增加编码长度，这种方法确实可以提高变量的表示精度，但却使算法的搜索空间呈指数级增长，增加了算法的计算量。当遇到多参数优化问题时，二进制编码位串会更长。一个长的二进制编码个体一直占据内存，会造成计算机内存极度浪费和低效使用。也会使交叉、变异等遗传操作实现起来变得困难，而且解码也很繁琐，导致算法寻优速度变慢，甚至无法运行，影响算法搜索效率。

对于另一种预先给一个较大编码长度值再缩小的试验方式确定编码长度的方法，二进制编码长度也取决于问题的求解精度。例如，设变量所在区间为 $[-1, 2]$ ，求解精度精确到小数点后6位，则区间长度为 $2 - (-1) = 3$ ，则需将区间 $[-1, 2]$ 分成 $3 \times 10^6$ 等份。则有：

$$2097152 = 2^{21} \quad 3 \times 10^6 \quad 2^{22} = 4194304 \quad (1)$$

由式(1)可以看出，二进制编码位串长度至少需要22位。

这种方法存在的问题是需要多次试验，才能求得一个准确的编码长度，用计算机编程实现的时候需要花费很长的实验及调试时间，浪费了时间，降低了算法的性能。编码长度无法用公式来实现。而过长的编码长度也存在搜索空间扩大，增加算法的计算量的问题，影响算法的寻优效率。

由此可见，适合的编码长度十分重要，会影响二进制遗传算法的收敛性、复杂度及效率。故二进制编码长度的值应取满足编码精度要求的最小值。

## 2 编码长度的计算方法

### 2.1 编码长度对决策变量精度的影响

设某一决策变量的可行区域是 $[U_{\min}, U_{\max}]$ ，用 $k$ 位二进制编码表示该决策变量，共有 $2^k$ 种不同的编码，参数编码与决策变量的关系如下：

$$\begin{aligned} 00000000 \cdots 0000 &= 0 \rightarrow U_{\min} \\ 00000000 \cdots 0001 &= 1 \rightarrow U_{\min} + \Delta x \\ 00000000 \cdots 0010 &= 2 \rightarrow U_{\min} + 2\Delta x \\ &\vdots \\ 11111111 \cdots 1111 &= 2^k - 1 \rightarrow U_{\max} \end{aligned}$$

其中，编码精度表示为：

$$\Delta x = \frac{U_{\max} - U_{\min}}{2^k - 1} \quad (2)$$

例如，决策变量 $x \in [2, 4]$ ，现在用5位二进制编码符号对 $x$ 进行编码，得 $2^5 = 32$ 个二进制个体，可表示的编码精度为

$$\Delta x = \frac{4 - 2}{2^5 - 1} = 0.064516$$

如果用4位二进制编码符号对 $x$ 进行编码，得 $2^4 = 16$ 个二进制个体，可表示的编码精度为

$$\Delta x = \frac{4 - 2}{2^4 - 1} = 0.133333$$

由此可见，决策变量在相同区间长度的条件下，二进制编码长度每增加一位，决策变量精度就可以提高一倍以上。

### 2.2 编码长度的计算公式

设 $[a, b]$  ( $a < b$ ) 区间内待优化的决策变量 $x$ ，求解精度为小数点后 $c$ 位， $m$ 为可能需要的基因数， $l$ 为待优化决策变量的二进制编码长度，则 $l$ 计算方法的基本原理和具体实现过程如下：

$$\frac{(b-a)}{10^{-c}} \quad 2^m \quad (3)$$

将式(3)变形为

$$(b-a) \times 10^c \quad 2^m \quad (4)$$

将式(4)变为等式:

$$(b-a) \times 10^c = 2^m \quad (5)$$

两边取以 2 为底的对数, 得到式(6):

$$m = \log_2(b-a) + c \log_2 10 \quad (6)$$

则对于一个决策变量  $x$  来讲, 二进制编码长度  $l$  的计算公式为:

$$l = \begin{cases} m & \text{当 } m \text{ 是整数时} \\ \text{int}(m) + 1 & \text{当 } m \text{ 不是整数时} \end{cases} \quad (7)$$

式中,  $\text{int}(m)$  表示对当  $m$  不是整数时, 取其整数部分。

对于对于多维优化问题  $f(x), x = (x_1, x_2, \dots, x_n)$ , 每一维决策变量二进制编码长度的计算方法与上述相同。若种群中每个个体编码采用累积二进制序列方式表示, 则每个个体的二进制编码长度  $L$  的计算公式为:

$$L = \sum_{i=1}^n l_i \quad (8)$$

由此可见, 二进制遗传算法种群中每个个体的编码长度  $L$  与决策变量的维数、决策变量的取值范围及求解精度有关。

### 3 结 论

(1) 分析了现有二进制遗传算法编码长度选取存在的问题。

(2) 给出了一种根据求解精度和区间长度确定编码长度的计算公式, 使待优化问题的编码长度以公式的形式表述出来, 方便计算机编程实现。

(3) 本文给出的计算公式, 可以得到以求解精度和区间长度为依据的所需的最小编码长度, 避免了因为搜索空间增加而造成效率降低的问题, 也避免了因为编码长度太短而造成的离散误差问题, 提高了算法的搜索效果和效率。

### 参考文献:

- [1] Goldberg D E. Genetic algorithms in search, optimization and machine learning[M]. Boston: Addison Wesley Longman Publishing Company, INC, New York, 1989.
- [2] Bertoni A, Dorigo M. Implicit parallelism in genetic algorithms[J]. Artificial Intelligence, 1993, 61(2): 307-314.
- [3] 张强, 王宾, 张锐, 等. 基于动态遗传算法的 DNA 序列集合设计[J]. 计算机学报, 2008, 31(12): 2193-2199.

- [3] 张晓绩, 方浩, 戴冠中. 遗传算法的编码机制研究[J]. 信息与控制, 1997, 26(2): 134-139.
- [4] Sarkar B K, Sana S S. A Genetic Algorithm-based Rule Extraction System[J]. Applide Soft Computing, 2012, 12(1): 238-254.
- [5] 张晋, 李冬黎, 李平. 遗传算法编码机制的比较研究[J]. 中国矿业大学学报, 2002, 33(1): 637-640.
- [6] Serkan Bekiroğlu, Tayfun Dede, Yusuf Ayvaz. Implementation of different encoding types on structural optimization based on adaptive genetic algorithm [J]. Finite Elements in Analysis and Design, 2009, 45(11): 826-835.
- [7] 岳崧, 冯珊. 遗传算法的计算性能的统计分析[J]. 计算机学报, 2009, 32(12): 2389-2392.
- [8] Wei Dong, Zanchetta P, Thomas D W P. Identification of electrical parameters in a power network using genetic algorithms and transient measurements[J]. International Journal for Computation and Mathematics in Electrical and Electronic Engineering, 2010, 29(1): 235-249.
- [9] Deming Lei. Simplified multi-objective genetic algorithms for stochastic job shop scheduling[J]. Applied Soft Computing, 2011, 11(8): 4991-4996.
- [10] ChienLung Chan, HsienWei Ting. Constructing a novel mortality prediction model with Bayes theorem and genetic algorithm [J]. Expert Systems with Applications, 2011, 38(7): 7924-7928.
- [11] Tayfun Dede, Serkan Bekiroğlu, Yusuf Ayvaz. Weight minimization of trusses with genetic algorithm[J]. Applied Soft Computing, 2011, 11(2): 2565-2575.
- [12] Castillo L, Dorao c a. Decision-making in the oil and gas projects based on game theory: Conceptual process design[J]. Energy Conversion and Management, 2013(66): 48-55.
- [13] 张超群, 郑建国, 钱洁. 遗传算法编码方案比较[J]. 计算机应用研究, 2011(03): 819-822.
- [14] 张文修, 梁怡. 遗传算法的数学基础[M]. 西安: 西安交通大学出版社, 2000.

责任编辑: 刘亚兵