

文章编号 : 1004 - 7271(2004) 03 - 0339 - 04

遗传算法及其在渔饲料配方中的应用

王海峰¹, 张 健²

(1. 上海水产大学海洋学院, 上海 200090 ;

2. 上海水产大学图书馆, 上海 200090)

摘 要 : 首先阐述了参数编码、初始群体、适应度函数、遗传操作以及算法控制参数等遗传算法基本原理, 然后着重介绍其在渔饲料配方中的应用, 在该应用中采用了实数编码、基于线性排名的选择、算术杂交和非一致变异等遗传算法方法。结果表明, 在渔饲料配方优化中应用遗传算法要优于目前饲料配方软件中大多采用的常规应用数学的优化方法。主要体现在单位产量饲料配方中的成本价格前者要低于后者。由于遗传算法能够解决因子较多、非线性程度高的问题, 从而得出的饲料配方能更好的符合相关鱼种的营养含量标准。

关键词 : 遗传算法 ; 渔饲料配方 ; 遗传因子 ; 模式定理 ; 编码 ; 适应度函数

中图分类号 : TP301.6 文献标识码 : A

The genetic algorithm and its application to feed diet of fishing

WANG Hai-feng¹, ZHANG Jian²

(1. Ocean College, Shanghai Fisheries University, Shanghai 200090, China ;

2. Library of Shanghai Fisheries University, Shanghai 200090, China)

Abstract : After elaborating the theory of the genetic algorithm, including the coding, initial population, fitness function, genetic operation and controls parameter of arithmetic, this paper presents its application to feed diet of fishing in detail. It adopts real coding, linear ranking selection, arithmetical crossover and nonconforming mutation. The results show that this algorithm is better than the optimization of common application mathematics in the feed diet of fishing. It mainly embodies that the unitage cost price of the former is lower than that of latter. Due to resolving the long degree non-linear problem with more factors, the diet by the genetic algorithm more well and truly accords with the nutrition content criterion of corresponding fish.

Key words : genetic algorithm ; feed prescription of fishing ; genetic factor ; schemata theorem ; coding ; fitness function

鱼类养殖是水产业的重要组成部分, 而饲料是发展水产养殖业的物质基础。从目前水产养殖经济结构的分析来看, 以池塘养鱼生产为例, 养殖成本约占总产值的 80%, 物化成本约占总成本的 80%, 其中饲料成本约占 70% 左右^[1]。可见成本的投入在水产养殖业占有举足轻重的作用。提出饲料配方优化技术, 就是为了更好地利用原料资源, 减少资源浪费, 从而降低生产成本, 同时又能使得鱼类合理健康

收稿日期 : 2004-04-05

作者简介 : 王海峰 (1980 -) 男, 江苏南通人, 硕士研究生, 专业方向为计算机应用。

通讯作者 : 张 健 (1955 -) 男, 上海市人, 副教授, 研究方向为计算机应用, E-mail : jzhang@shfu.edu.cn

地生长。目前大多数优化方法比如运用数学中的单纯形法在饲料配方优化技术得到比较广泛的应用,取得了很好的效果,但是这类方法往往只能解决一些因子较少、较为简单的问题,面对非线性程度高、较为复杂的问题则很难适应。为此利用计算机技术,提出应用遗传算法来优化渔饲料配方。由于鱼类种类繁多,而每种鱼对蛋白质、脂肪、糖、维生素和矿物质等一些营养元素的要求各不相同,为了能把算法阐述明白同时尽量减少篇幅,研究以青鱼为例,其它鱼种的饲料配方只要在此算法中调整相应的饲料营养参数,亦能取得同样的效果。

1 遗传算法原理

遗传算法是一类借鉴生物界自然选择和自然遗传机制的随机化搜索算法,由美国 J. Holland 教授提出,是一个以适应度函数(或目标函数)为依据,通过对群体个体施加遗传操作实现群体内个体结构重组的迭代处理过程^[1]。遗传算法的实现涉及如下 5 个主要因素。

1.1 参数的编码

遗传算法不能直接处理问题空间的参数,必须把它们转换成遗传空间的由基因按一定结构组成的染色体或个体。DeJong^[2]提出有意义积木块编码规则和最小字符集编码规则两种较为客观明确的编码评估准则。在编码技术方面,主要有一维染色体编码、多参数映射编码、离散化编码、可变染色体长度编码、二维染色体编码和树结构编码等。

1.2 初始群体的设定

初始群体的设定可采用如下的策略:根据问题固有的知识,设法把握最优解所占空间在整个问题空间中的分布范围,在此分布范围内设定初始群体。在随机生成一定数目的个体后,从中挑出最好的个体加到初始群体中。这样不断迭代,直到初始群体中个体数达到了预定的规模。

1.3 适应度函数的设计

遗传算法在进化搜索中基本上不用外部信息,仅用目标函数即适应度函数为依据。这也是遗传算法优于其他算法的特征之一。对适应度函数的唯一要求是,针对输入可计算出能加以比较的非负结果。值得注意的是,适应度函数评估是选择操作的依据,适应度函数设计直接影响到遗传算法的性能。

1.4 遗传操作的设计

遗传算法操作包括以下三个基本遗传算子:(1)选择(2)交叉(3)变异。每个遗传算子的操作都是在随机扰动情况下进行的。遗传操作的效果和上述三个遗传算子所取的操作概率、编码方法、群体大小、初始群体以及适应度函数的设定密切相关。另外,每个遗传算子的操作方法或操作策略随具体求解问题的不同而异,是和个体的编码方式直接相关。

1.5 算法控制参数的设定

所谓控制算法也就是使得算法在许多复杂问题的求解中能够得到满意的结果,也就是控制算法的收敛性。目前普遍认为,标准遗传算法并不保证全局最优收敛。但是,在一定的约束条件下,遗传算法可以实现这一点。

2 遗传算法在渔饲料配方中的应用

2.1 数据来源

研究所用青鱼饲料营养参数标准为粗蛋白占 30%~35%、粗脂肪占 4%~6%、粗纤维占 8%~10%、糖类占 30%~35%、磷占 0.57%~0.62%、赖氨酸占 2.0%、蛋氨酸占 0.9%和可消化能占 13.4%^[3]。

2.2 数据编码

设计遗传算法的一个重要步骤是对所解问题的变量进行编码表示,而编码表示方案的选取很大程

度上依赖于问题的性质及遗传算子的设计。试验采用实数编码^[4]的方法来进行数据编码,实数编码表示比较自然,而且容易引入相关的领域知识。采用 10 个实数变量来分别表示玉米、小麦、豆粕、棉子粕、菜子粕、花生仁饼、芝麻饼、小麦麸、鱼粉和饲料酵母一些常用原料在配合饲料中所占百分比含量。

2.3 设定初始群体

群体规模 N 是遗传算法的很重要的一个参数,有时它对算法的性能起到关键性的作用,如果群体规模 N 太小,则群体内缺乏足够的多样性,算法可能收敛过快;而如果群体规模 N 太大,则算法可能会浪费掉很多计算资源。因为等待结果的更新要花费太长的时间,一般取 n 从 10 到 160 之间。试验取 n 为 100,通过随机方法产生 100 个可能解 $u_i (1 \leq i \leq n)$ 组成初始解群。

2.4 设计适应度函数

在自然界中,个体的适应值即它繁殖的能力。它将直接关系到其后代的数量,在遗传算法计算中,适应度函数是用来区分群体中个体好坏的标准,是进行自然选择的唯一依据。试验要降低饲料成本,适应度函数属于极小化问题范畴,研究采用标准函数极小化方法。设函数

$$y = \sum_{i=1}^{10} c_i x_i^2 + 1^{[5-6]}$$

其中 $c_i (i = 1, 2, \dots, 100)$, 分别代表其相应的饲料原料的市场价格, y 为饲料的成本价格。

2.5 遗传操作的设计

2.5.1 选择策略

选择策略对算法性能的影响起到举足轻重的作用,不同的选择策略将导致不同的选择压力,即下一代中父代个体的复制数目的不同分配关系,较大的选择压力使最优个体具有较高的复制数目,从而使得算法收敛速度较快,但也较容易出现过早收敛现象。相对而言,较小选择压力一般能使群体保持足够的多样性,从而增大了算法收敛到全局最优的概率^[7]。研究采用基于线性排名的选择方法,假设群体成员按适应值大小从好到坏依次排列为 u_1, u_2, \dots, u_{100} , 然后根据一个线性函数分配选择概率 p_i , 设 $p_i = (a + b_i/101)/100$, a, b 是常数, b_i 随机产生且满足 $b = \sum_{i=1}^{100} b_i (i = 1, 2, \dots, 100)$, 由于 $\sum_{i=1}^{100} p_i$, 易得 $b = 2(a - 1)$, 又要求对任意 $i = 1, 2, \dots, 100$ 有 $p_i \geq 0$ 且 $p_1 \geq p_2 \geq \dots \geq p_{100}$, 故限定 $1 \leq a \leq 2$, 取 $a = 1.1$ 。有了选择概率,即可类似于转盘式选择的方式来选择父体以进行遗传操作。通过这种选择方法,可以避免常常会出现的过早收敛和停滞现象。

2.5.2 杂交算子

采用算术杂交方法,把 10 个变量组合成一个 10 维的向量,并且每一个分量都在有限区间上定义。如设 $s = (v_1, v_2, \dots, v_{10})$ 是一个解向量,则有 $v_i \in [a_i, b_i], i = 1, 2, \dots, 10$ 。反之对任何满足 $v_i \in [a_i, b_i]$ 的向量 $s' = (v_1', v_2', \dots, v_{10}')$, s' 都是问题的解。算法中先生成 10 个 $(0, 1)$ 区间的随机数 $\alpha_1, \alpha_2, \dots, \alpha_{10}$, 则两后代 s_z 和 s_w 分别定义为

$$z_i = \alpha_i v_i^{(1)} + (1 - \alpha_i) v_i^{(2)} = v_i^{(2)} + \alpha_i (v_i^{(1)} - v_i^{(2)})$$

$$w_i = \alpha_i v_i^{(2)} + (1 - \alpha_i) v_i^{(1)} = v_i^{(1)} + \alpha_i (v_i^{(2)} - v_i^{(1)})$$

这里 $i = 1, 2, \dots, 10$ 。

2.5.3 变异算子

采用非一致变异的方法,设 $s = (v_1, v_2, \dots, v_{10})$ 是一个父解,分量 v_k 被选为进行变异,其定义区间是 $[a_k, b_k]$, 则变异后解为 $s' = (v_1, \dots, v_{k-1}, v_k', \dots, v_{10})$ 其中

$$v_k' = \begin{cases} v_k + \delta(t_1, b_k - v_k), & \text{如果 } \text{rnd}(2) = 0 \\ v_k - \delta(t_1, v_k - a_k), & \text{如果 } \text{rnd}(2) = 1 \end{cases}$$

这里 $\text{rnd}(2)$ 表示将随机均匀产生的正整数模 2 所得的结果, t 为当前演化代数,而函数 $\delta(t, y)$ 的值域为 $[0, y]$, 并使得当 t 增大时, $\delta(t, y)$ 接近于 0 的概率增加。即 t 的值越大, $\delta(t, y)$ 取值接近于 0 的可

能性越大,从而使得算法在演化初期能搜索到较大范围,而在后期主要是进行局部搜索了。函数 $\delta(x, y) = y \cdot (1 - r^{(1-1/T)^{\lambda}})$ 这里 r 是 $[0, 1]$ 上的随机数, T 表示最大代数, λ 是决定非一致性程度的一个参数,起着调整局部搜索区域的作用^[7]。

2.6 控制参数的设定

在设计遗传算法时,一方面需要针对具体问题选择适当的编码方案及相应的遗传算子;另一方面需要选择算法的控制参数。参数的不同选取常常会对算法的性能产生影响^[8]。试验群体规模为 100、杂交概率 P_c 为 0.90 和变异概率 P_m 为 0.15。

3 讨论与结论

试验中 10 个实数变量的演化数据如表 1(限于篇幅,选取其中 10 组数据):

表 1 演化中的实数变量
Tab.1 Real variant in the evolution

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
1	0.106	0.185	0.044	0.187	0.019	0.103	0.096	0.049	0.063	0.148
2	0.148	0.095	0.000	0.188	0.089	0.181	0.164	0.005	0.108	0.022
3	0.063	0.096	0.083	0.224	0.106	0.123	0.119	0.026	0.027	0.133
4	0.088	0.115	0.049	0.131	0.104	0.096	0.052	0.000	0.083	0.282
5	0.102	0.063	0.153	0.118	0.117	0.024	0.055	0.089	0.090	0.189
6	0.213	0.055	0.021	0.124	0.126	0.089	0.052	0.130	0.101	0.089
7	0.139	0.201	0.004	0.113	0.042	0.059	0.025	0.067	0.078	0.272
8	0.121	0.033	0.050	0.119	0.187	0.114	0.111	0.046	0.114	0.105
9	0.037	0.193	0.041	0.050	0.146	0.138	0.131	0.089	0.062	0.113
10	0.141	0.051	0.048	0.093	0.138	0.213	0.045	0.064	0.122	0.085

由此在 10 种饲料原料中得出了在满足青鱼饲料营养参数标准前提下比较经济的配方比例合成(1)小麦 17.21%、小麦麸 35.13%、豆粕 16.16%、鱼粉 30.5%和酵母 1%;与其对比,在通用配方软件得出的配方比例合成(2)玉米 28.34%、小麦 6.29%、豆粕 28.29%、棉子粕 2.54%、菜子粕 7.89%和鱼粉 26.65%。以单位饲料重量 1t 为例,则比例合成(1)的成本价格为 2373.04(17.21% * 1100 + 35.13% * 1300 + 16.16% * 2180 + 30.5% * 4500 + 1% * 1800)元,比例合成(2)的成本价格为 2412.37(28.34% * 1430 + 6.29% * 1100 + 28.29% * 2180 + 2.54% * 1850 + 7.89% * 950 + 26.65% * 4500)元,其中括号内表示的是各饲料原料在合成中所占比例与其市场价格(单位:元/吨)乘积的总和。试验引用的饲料原料价格来源于近期饲料网上。易见比例合成(1)的成本价格在单位饲料重量上要小于比例合成(2)的成本价格。另外在运用通用配方软件进行配方的过程中发现,由于该软件不能很好地解决因子较多、非线性程度高的问题,从而使得其在配出的饲料比例合成与青鱼的营养含量标准有失一致。

各地饲料原料价格存在差异,故试验所得比例合成也可能因地而异,另外此研究仅从降低水产养殖成本的角度,提出应用遗传算法来优化渔饲料配方,对于此配方下饲料中有机物的微量元素影响问题,饲料投放后可能影响水中含氧量问题等还需进一步探讨研究。

参考文献:

[1] 王道尊. 鱼用配合饲料[M]. 北京:中国农业大学出版社,1993.2-5.
 [2] 陈国良,王煦法,庄镇泉,等. 遗传算法及其应用[M]. 北京:人民邮电出版社,1999.1-53.
 [3] 关受江. 鱼类营养及饲料学[M]. 成都:电子科技大学出版社,1992.56-121.
 [4] 郑志军,郑守淇. 用基于实数编码的自适应遗传算法进化神经网络[J]. 计算机工程与应用,2000,9:36-37.
 [5] 王小平,曾立明. 遗传算法理论、应用与软件实现[M]. 西安:西安交通大学出版社,2002.272-332.
 [6] Koza J R. Genetic Programming II: Automatic Discovery of Reusable Programs[M]. The MIT Press,1995.51-132.
 [7] 潘正君,唐立山,陈毓屏,等. 演化计算[M]. 北京:清华大学出版社,1998.16-35.
 [8] 林霞光,刘轶,郑志军,等. 利用遗传算法求解线程分配问题[J]. 计算机工程与应用,2000,10:77-78.