

多目标免疫 GEP 算法 及其在多项式 NARMAX 模型辨识中的应用

周霞^{1,2}, 沈炯¹

(1. 东南大学 能源与环境学院, 南京 210096; 2. 金陵科技学院 机电工程学院, 南京 211169)

摘要: 基于免疫细胞亚群划分理论和生命周期理论, 提出一种用于多项式非线性自回归滑动平均 (NARMAX) 模型辨识的多目标免疫 GEP 算法, 并重新定义了适合于 NARMAX 模型辨识的细胞创建算子和基因操作算子. 该算法基于多目标优化的最优解通常为一个集合的特点, 通过各亚群的最优解集在总最优解集中的变化来判断各亚群所对应参数的优劣, 进而确定下一步的搜索方向. 仿真结果表明, 多目标免疫 GEP 算法可以同时正确地辨识出非线性系统的结构和参数.

关键词: 多目标; 免疫; 基因表达式编程; 非线性自回归滑动平均模型

中图分类号: TP13

文献标志码: A

An immune based multiobjective GEP algorithm for identifying polynomial NARMAX model

ZHOU Xia^{1,2}, SHEN Jiong¹

(1. School of Energy and Environment, Southeast University, Nanjing 210096, China; 2. School of Mechanical and Electrical Engineering, Jinling Institute of Technology, Nanjing 211169, China. Correspondent: SHEN Jiong, E-mail: shenj@seu.edu.cn)

Abstract: Based on the immune cell subsets division and immune life theory, an immune based multiobjective gene expression programming(GEP) algorithm for identifying the polynomial nonlinear auto regressive moving average with exogenous inputs(NARMAX) model is proposed. In the algorithm, the cell initialization method and gene operators are newly defined. Based on the multi-optimal-solutions characteristic of multiobjective optimization problems, the parameter performance of each subset is estimated by the optimal solutions of the subset in the total optimal solutions, and then the search direction for the next generation is determined. The simulation results show that the proposed algorithm performs well in the polynomial NARMAX modeling, and the structure and the parameters of the nonlinear system can be identified at the same time.

Key words: multiobjective; immune; gene expression programming(GEP); nonlinear auto regressive moving average with exogenous inputs model

0 引言

非线性广泛存在于各种控制系统中, 非线性系统辨识是进行非线性系统控制的前提和基础. Chen 等^[1]提出的非线性自回归滑动平均 (NARMAX) 模型提供了一个统一的有限可实现非线性系统表达式. NARMAX 模型逼近精度高、收敛速度快, 已逐渐成为应用领域所接受.

NARMAX 模型包括多种形式, 其中多项式是一

种常见的形式^[2]. 多项式 NARMAX 模型表达式为多个输入输出乘积项的和, 当输入输出的阶次较小、多项式的最大方次也比较小时, 可以应用最小二乘法进行模型参数的辨识^[3]; 但当多项式的方次和输入输出的阶次较大且缺乏先验知识时, 模型的项数会非常巨大, 应用最小二乘法将使辨识变得非常困难.

基因表达式编程 (GEP) 是 Ferreira 提出的一种新型进化计算方法^[4]. 该方法将基因型与表现型分开,

收稿日期: 2013-05-08; 修回日期: 2013-11-04.

基金项目: 国家自然科学基金项目(51306082).

作者简介: 周霞(1976—), 女, 讲师, 博士后, 从事智能控制算法及其应用的研究; 沈炯(1957—), 男, 教授, 博士生导师, 从事先进控制理论及其及应用等研究.

用简单编码处理复杂问题,在复杂问题求解方面具有极大的优越性.已有的研究成果表明,GEP在数据挖掘、时间序列预测、聚类等领域都取得了比传统优化算法更好的效果^[4].根据NFL定理^[5],将GEP应用于实际问题必须研究与实际应用相宜的优化策略才能达到良好的应用效果.

GEP的多基因特性使得它可以方便地表达多项式NARMAX模型.文献[6]创新性地提出了采用可变基因长度、可变头长、可变终端符集的GEP进行多项式NARMAX模型的辨识,并定义了复杂度与精度相结合的适应度计算方法.但是,由于各个体的阶次值与长度值可能不同,基因操作受到很多限制,且用加权方法处理多目标优化问题时,难以得到多目标意义的最优解.若能够结合多目标优化的特点,则一定能促进此问题的更好解决.

近年来,随着生命科学的发展,人工免疫算法已成为优化算法的重要研究方向之一.免疫细胞是参与免疫应答的细胞,根据免疫细胞的分化发育和功能特点的不同可将免疫细胞分为不同的亚群^[7].在免疫细胞的生命过程中,若遇抗原产生匹配,且累积足够的亲和力,则被激活,转变为记忆细胞,并进行克隆扩增,以抵御更多抗原的入侵.若免疫细胞在其生命周期内未能累积足够的亲和力,则会消亡,并被新的免疫细胞所代替^[7].

为将GEP更好地应用于多项式NARMAX模型的辨识,本文将免疫细胞亚群划分原理和生命周期理论引入GEP算法,提出了多目标免疫GEP算法.将所有细胞按方次和阶次的不同划分为多个亚群,而各亚群内的基因操作则不受任何限制.基于多目标优化的最优解通常为一个集合的特点,通过各亚群的最优解在总最优解集中的变化来判断各亚群所对应参数的优劣,进而确定下一步的搜索方向.应用多目标免疫GEP算法,在多项式的次数和阶次值较大且缺乏先验知识时,仍然可以保证在有限的时间复杂度内辨识出正确的模型.

1 基于多目标GEP的多项式NARMAX模型辨识

1.1 多项式NARMAX模型

多项式NARMAX模型的一般表达形式为

$$y(k) = \theta_0 + \sum_{i=1}^n \theta_i P_i(k) + e(k). \quad (1)$$

其中: θ_0 为常数项, $e(k)$ 为噪声, $P_i(k) = P_i(y(k-1), \dots, y(k-n_y), u(k-1), \dots, u(k-n_u))$ 为多项式模型的回归项, n_u 和 n_y 分别为控制量和被控量的阶次, θ_i 为待估计的未知参数.基于多项式NARMAX模型的

非线性系统辨识,本质上是确定模型由哪些非线性项组成,而且模型的精度和泛化能力在很大程度上依赖于模型结构的准确性.即使对于不太大的阶次和方次,模型项的数目也是非常巨大的,再加上系数的处理,辨识出正确的多项式函数形式是非常困难的.

1.2 基于GEP的NARMAX模型表达

将多项式NARMAX模型中相加的各个乘积项分别理解为一个基因,基因中所含运算符为相乘,基因连接符为相加,则可很方便地将多项式NARMAX模型描述为GEP的表达式形式.

文献[8]指出,GEP的特色体现在个体的构成方式和基因操作等细节方面.多项式NARMAX模型的基本项为各种阶次输入输出项的乘积,项中包含了阶次因素,因此需要研究新的个体表达方法.

关于基因操作,因为多项式NARMAX模型的函数运算符只包括乘法一种类型,而乘法是满足交换律的,故而GEP的倒串、插串、根插串等算子不再适用.因此,有必要研究新的适合于NARMAX模型辨识的基因操作方法.

1.3 多项式NARMAX模型的多目标辨识问题

常用的动态模型精度评价指标包括残余方差(VAR)、长期预测误差(LTPE)等^[9].VAR反映了模型的拟合能力,而LTPE则反映了模型的预测能力,VAR和LTPE皆为均方误差的形式.均方误差是对一组测试数据可靠性的估计,对于每个测试数据,可以保证其误差以一个百分比的概率落入均方误差求解出的区间范围,但仅根据VAR和LTPE并不能确切了解各个数据的误差上限.为了明确各个数据的精度状况,在多目标的情况下,可以将最大预测误差max_PE作为一个目标,这样可以更直观地进行评价.

综合考虑VAR、LTPE和max_PE这3个目标,多项式NARMAX模型的辨识问题可以描述为

$$\begin{aligned} & \min(\text{VAR}, \text{LTPE}, \text{max_PE}); \\ & \text{s.t. } n_u \in \{n_{u_{\min}}, \dots, n_{u_{\max}}\}, \\ & \quad n_y \in \{n_{y_{\min}}, \dots, n_{y_{\max}}\}. \end{aligned}$$

多目标优化问题是对多个目标同时进行优化,通常无法获得唯一的最优解.一种常见的多目标最优解形式为Pareto最优解,也称为非支配解.多目标优化的结果通常为为一组Pareto最优解.

2 多目标免疫GEP算法

为使GEP更好地适用于NARMAX模型的辨识,本文提出了新的适合于NARMAX模型辨识的个体创建方式,并新定义了适合于NARMAX模型辨识的基因操作算子.此外,为适应多目标优化的需要,在算

法执行时设立了记忆集,用于保存已求解出的 Pareto 最优解集。

下面首先给出算子定义和算法流程,然后对算法进行分析。

2.1 算子定义

在免疫学中,细胞含有特定条染色体,染色体中有多个 DNA,而 DNA 中有无数个基因。为算法描述方便,文中细胞简化为由基因组合而成。

定义 1(基因库创建算子) 根据基因项的最高方次确定基因库的内容,创建的基因库中包含运算符和终端符的所有合法组合模式。

为了方便基因操作,采用等长基因方式。记运算符 $F = \{f|*\}$,用“?”表示常数,终端符 $T = \{t|u, y, ?\}$,基因头部长度为 h ,尾部长度为 e ,则

$$\text{gene}(i) = g_{i_1}g_{i_2} \cdots g_{i_h}t_{i_1}t_{i_2} \cdots t_{i_e}. \quad (2)$$

式中: $g_{i_j} \in F \cup T, t_{i_k} \in T, j \in \{1, 2, \dots, h\}, k \in \{1, 2, \dots, e\}$ 。

定义 2(细胞创建算子) 在基因库中选择若干个基因,通过连接符连接生成细胞。将细胞所含的最大基因数目记为 $n_{g_{\max}}$,则各细胞所含的基因数目是在 $1 \sim n_{g_{\max}}$ 范围内动态变化的。

在创建初始细胞群时,对基因库中各个基因设置不同的选择概率。使用这种方式生成初始细胞,有以下 4 个方面的优点:

- 1) 使用基因库生成初始细胞群,可以保证细胞群的多样性;
- 2) 基因库中包括常数项,因此无需对常数项进行特殊处理;
- 3) 各细胞所含的基因数目是动态变化的,因此各细胞实际对应的表达式项数可能不同;
- 4) 基因采用了等长方式,一方面使得基因操作不受约束,方便基因操作的实现;另一方面,根据文献[10]的研究成果,在基因中保持一定的冗余(中性区),能使 GEP 算法具有更高的效率。

定义 3(解码算子) 对初始细胞群中的各基因进行解码,将参数项译为合适的参数形式,常数项译为合适的常数值。

参数项包括控制量和被控量,被控量解码后应包含阶次信息,控制量解码后应包含维数和阶次信息。对于常数项的解码,通过在初始常数数组中选择数据进行。

定义 4(基因变异算子) 在基因中随机选择一个非运算符的有效位置,然后对其进行变异。基因变异算子不改变基因类型,仅对 u 项的下标和阶次、 y 项的阶次或者常数项的值进行变异。

定义 5(转基因算子) 转基因算子改变基因类型,对随机选择的某基因位置执行操作,具体操作类型根据所选择的位置确定。如果选择的位置在基因的头,则执行运算符与非运算符之间的相互转换;如果选择的位置在基因的尾部,则执行参数项与常数项之间的相互转换。

定义 6(基因部分重组算子) 用于生成新的细胞,其重组的基因为解码后的基因。该算子在待重组的两个细胞中各随机选择若干个基因,然后重新连接生成基因数目小于 $n_{g_{\max}}$ 的新细胞。

定义 7(基因全部重组算子) 将一组细胞的所有基因作为一个库,在其中随机选择小于 $n_{g_{\max}}$ 个基因生成新的细胞。

基因全部重组算子中的基因亦是经过解码的,此算子在算法中仅作用于记忆细胞。

定义 8(克隆算子) 克隆是对各克隆源细胞繁殖给定数目的细胞的映射。算法中直接选取上代的非支配解集作为克隆源,并采用平均克隆的方式。

定义 9(细胞变异算子) 细胞变异算子按一定概率执行基因部分重组或者基因变异操作。执行基因变异操作时,随机选择细胞中的部分基因执行基因变异算子或转基因算子。

2.2 算法流程

多目标免疫 GEP 算法流程如图 1 所示。首先初始化产生各亚群,然后进行非支配集的求解和记忆集的处理,在未满足终止条件前,根据各细胞亚群的生命周期和记忆集中解的阶次和方次值生成下一代细胞的阶次和方次参数值,迭代搜索寻求最优解。

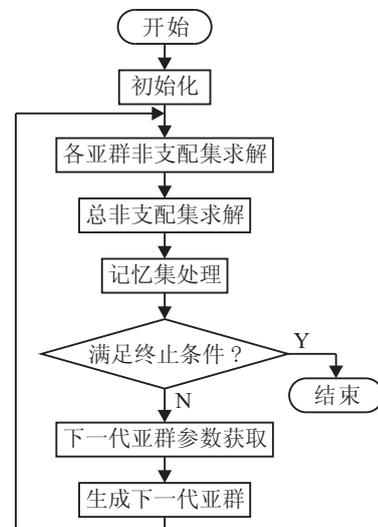


图 1 多目标免疫 GEP 算法流程

下面对算法流程中的主要计算步骤进行简要说明。

- 1) 初始化。初始化包括亚群初始化和记忆集初

始化两部分. 亚群初始化时, 在阶次和方次的取值范围内, 随机生成初始各亚群基因的阶次和方次参数值, 执行细胞创建、解码及目标值计算步骤; 记忆集则直接初始化为空集.

2) 非支配集求解. 非支配集求解算子计算给定细胞集合关于目标向量的非支配集, 具体方法参见文献 [11].

3) 记忆集处理. 进行记忆集处理时, 先获取当代非支配集与上代记忆集合的非支配集, 若数量小于记忆集设定的上限, 则直接将此非支配集作为新的记忆集; 否则, 需删除此非支配集中的部分个体, 直至集合大小等于记忆集设定的上限. 为了提高辨识结果的可信度, 优先删除 \max_PE 与 $LTPE$ 值相差较大的细胞.

4) 下一代亚群的参数获取. 每次迭代后, 若在记忆集中存在与该亚群最高方次和阶次相同的细胞, 则该亚群生命加 1, 否则该亚群生命减 1, 生命为 0 的亚群退出循环. 在每轮迭代结束前, 根据记忆集中解的方次和阶次值以及上代生命不为 0 的亚群来决定下一代亚群的参数. 此外, 为了提高算法的全局搜索性能, 另有部分下一代亚群的参数根据阶次和方次值的取值范围随机产生.

5) 下一代亚群的生成. 下一代亚群生成时包括 3 个组成部分: 第 1 部分在上代非支配解集执行细胞克隆算子和细胞变异算子后产生的集合中择优选取; 第 2 部分, 通过对上代记忆集执行基因全部重组算子和细胞变异算子来产生; 第 3 部分为根据记忆集中解的方次和阶次值、上代生命不为 0 的亚群的参数值及新随机产生的亚群参数按照初始细胞生成方案产生的细胞.

6) 算法的终止条件. 算法的终止有两种可能性: 一种可能性是达到一定的运行代数; 另一种可能性是连续若干代记忆集无变化.

2.3 算法分析

一个算法的适用性如何, 一方面取决于其计算结果的可信度, 另一方面与其计算过程的复杂性密切相关. 下面将对本文算法的收敛性和计算复杂度进行分析.

2.3.1 算法收敛性分析

用 C 表示候选细胞集, M 表示记忆细胞集, 多目标免疫 GEP 算法的一次迭代过程可以归纳为

$$\begin{cases} C(t) \rightarrow C^1(t+1), \\ M(t) \rightarrow C^2(t+1), \\ \text{initial para} \rightarrow C^3(t+1); \end{cases} \Rightarrow \begin{cases} C(t+1), \\ M(t+1). \end{cases} \quad (3)$$

其中

$$C(t+1) = C^1(t+1) + C^2(t+1) + C^3(t+1),$$

$$M(t+1) = \text{Pareto}(M(t) \cup C(t));$$

$C^1(t+1)$ 是对本代最优解周围的精细搜索, 通过求解本代的非支配集并执行克隆变异来产生; $C^2(t+1)$ 通过对记忆集中的解执行基因全部重组算子和细胞变异算子来产生; $C^3(t+1)$ 部分则根据本代记忆集中的相关参数值、上代生命不为 0 的亚群的参数值及部分随机产生的参数值按照初始化的方式来产生.

引理 1 假设基因变异的概率为 P_{mu} , 基因变异的 最大可能变化方案包括 N_{mu} 个, 转基因的概率为 P_{tr} , 转基因的最大可能变化方案包括 N_{tr} 个, gene_1 变异为 gene_2 需进行 k_1 次基因变异操作和 k_2 次转基因操作, 则

$$P\{\text{gene}_1 \rightarrow \text{gene}_2\} = (1 - P_{\text{mu}})^{h+e-k_1} (1 - P_{\text{tr}})^{h+e-k_2} \left(\frac{P_{\text{mu}}}{N_{\text{mu}}}\right)^{k_1} \left(\frac{P_{\text{tr}}}{N_{\text{tr}}}\right)^{k_2}. \quad (4)$$

引理 2 对于任意两个含基因数目为 n_{g_1} 和 n_{g_2} 的细胞 cell_1 和 cell_2 , 将 n_{g_1} 和 n_{g_2} 值中较大的记为 $n_{g_{\text{max}}}$, 较小的记为 $n_{g_{\text{min}}}$. 细胞 1 中各基因变异为细胞 2 中的各基因的最大、最小基因变异操作次数和转基因操作次数分别为 $k_{1_{\text{max}}}$ 、 $k_{1_{\text{min}}}$ 和 $k_{2_{\text{max}}}$ 、 $k_{2_{\text{min}}}$, 细胞 1 中基因经重组成为细胞 2 中基因的概率为 P_{rb} , 最大重组方案包括 N_{rb} 个, 则细胞 1 经变异重组转变为细胞 2 的概率为

$$P\{\text{cell}_1 \rightarrow \text{cell}_2\} \geq \left((1 - P_{\text{mu}})^{h+e-k_{1_{\text{min}}}} (1 - P_{\text{tr}})^{h+e-k_{2_{\text{min}}}} \times \left(\frac{P_{\text{mu}}}{N_{\text{mu}}}\right)^{k_{1_{\text{max}}}} \left(\frac{P_{\text{tr}}}{N_{\text{tr}}}\right)^{k_{2_{\text{max}}}} \right)^{n_{g_{\text{min}}}} \times (1 - P_{\text{rb}})^{n_{g_{\text{min}}}} \left(\frac{P_{\text{rb}}}{N_{\text{rb}}}\right)^{n_{g_{\text{max}}}-n_{g_{\text{min}}}}.$$

证明 在细胞 1 中 $n_{g_{\text{min}}}$ 个基因变异为细胞 2 中 $n_{g_{\text{min}}}$ 个基因的概率为

$$P_1 \geq \left((1 - P_{\text{mu}})^{h+e-k_{1_{\text{min}}}} (1 - P_{\text{tr}})^{h+e-k_{2_{\text{min}}}} \times \left(\frac{P_{\text{mu}}}{N_{\text{mu}}}\right)^{k_{1_{\text{max}}}} \left(\frac{P_{\text{tr}}}{N_{\text{tr}}}\right)^{k_{2_{\text{max}}}} \right)^{n_{g_{\text{min}}}}. \quad (5)$$

一细胞中 $(n_{g_{\text{max}}} - n_{g_{\text{min}}})$ 个基因经重组成为另一细胞中基因的概率为

$$P_2 = (1 - P_{\text{rb}})^{n_{g_{\text{min}}}} \left(\frac{P_{\text{rb}}}{N_{\text{rb}}}\right)^{n_{g_{\text{max}}}-n_{g_{\text{min}}}}. \quad (6)$$

因此细胞 1 转变为细胞 2 的概率为

$$P\{\text{cell}_1 \rightarrow \text{cell}_2\} = P_1 P_2 \geq \left((1 - P_{\text{mu}})^{h+e-k_{1_{\text{min}}}} (1 - P_{\text{tr}})^{h+e-k_{2_{\text{min}}}} \times \right.$$

$$\left(\frac{P_{\text{mu}}}{N_{\text{mu}}}\right)^{k_{1\text{max}}}\left(\frac{P_{\text{tr}}}{N_{\text{tr}}}\right)^{k_{2\text{max}}})^{n_{g\text{min}}}\times (1 - P_{\text{rb}})^{n_{g\text{min}}}\left(\frac{P_{\text{rb}}}{N_{\text{rb}}}\right)^{n_{g\text{max}} - n_{g\text{min}}}. \quad \square$$

定理 1 多目标免疫 GEP 算法求解出的记忆集序列是有限非齐次可约马尔可夫链。

证明 1) 在给定精度后,各变量将在一定范围内取值.对于方次、阶次和基因数目的取值,是有限的离散整数;对于常数部分,取值为离散实数.因此,变化范围都是有限的。

2) 迭代选择求解下一代记忆集时,仅与当代记忆集及当代的搜索过程有关,与前面的记忆集无关,因而求解出的记忆集序列构成了一个马尔可夫链.由于 $C^3(t+1)$ 部分的搜索与当时的细胞群状态有关,该马尔可夫链是非齐次的。

3) 由 Pareto 最优解集 (PS) 的定义可知,各成员细胞全部属于 PS 的所有记忆集组成的群体形成了一个正常返的闭集。

① $\forall X, Y \subset \text{PS}, p_{XY} > 0, p_{YX} > 0$, 因此 $X \leftrightarrow Y$;

② 若 $X \subset \text{PS}, Z \not\subset \text{PS}, p_{XZ} = 0$, 则 $X \nrightarrow Z$. \square

定理 2 假定马尔可夫链仅有一个正常返类 D , 则当 $Y \notin D$ 时, $\lim_{t \rightarrow \infty} P\{X_t = Y\} = 0$; 而当 $Y \in D$ 时, $\lim_{t \rightarrow \infty} P\{X_t = Y\} = \pi_Y$, 且

$$\lim_{t \rightarrow \infty} P\{X_t \in D\} = \sum_{Y \in D} \pi_\infty(Y) = 1^{[12]}.$$

定理 3 多目标免疫 GEP 算法的记忆集序列以概率 1 收敛到所求解问题的 PS, 即

$$\lim_{t \rightarrow \infty} P\{M(t) \subset \text{PS} | C(0) = C_0\} = 1. \quad (7)$$

证明 假设有另一个闭集, 该闭集中的解不属于 PS, 则该集合中的解不可能到达 PS. 而根据引理 2, 所有细胞之间都是变异可达的, 所以有可能到达 PS, 因而假设错误, 故只有一个闭集。

由定理 1 可知, 记忆集序列形成的马尔可夫链中存在正常返类闭集, 即各成员细胞全部属于 PS 的所有记忆集组成的群体; 由定理 2 可知, 只有一个正常返闭集的马尔可夫链必定收敛到该正常返闭集中. \square

2.3.2 时间复杂度分析

设亚群数目最大值为 ss , 各亚群的规模为 N_i , 目标向量的维数为 m , 各亚群非支配集的大小为 P_i , 最大克隆规模为 5 倍于群体数量, 最大运行迭代次数为 gen .

本算法的时间复杂度主要由 5 部分组成: 细胞初始化的时间复杂度 $O(N_i + mN_i)$, 非支配集求解的时间复杂度 $O(mN_iP_i)^{[11]}$, 克隆操作的时间复杂度 $O(C_i)$, 基因操作的时间复杂度 $O(C_i)$, 细胞更新操作的时间复杂度 $O(C_i + C_i^2 + P_i)$. 各亚群一次迭代的

时间复杂度为

$$\begin{aligned} &O(N_i + mN_i) + O(mN_iP_i) + O(C_i) + \\ &O(C_i) + O(C_i + C_i^2 + P_i) = \\ &O(N_i + mN_i + mN_iP_i + C_i + C_i + C_i + C_i^2 + P_i) = \\ &O(N_i + mN_i + mN_iP_i + 3C_i + C_i^2 + P_i) \leq \\ &O(N_i + mN_i + mN_iP_i + 3 \times 5N_i + (5N_i)^2 + P_i) = \\ &O(16N_i + mN_i + mN_iP_i + 25N_i^2 + P_i). \end{aligned}$$

假设各亚群总非支配集大小为 Q , 总目标向量维数为 n , 则获取各亚群总非支配集的时间复杂度为

$$O\left(nQ \sum_{i=1}^{ss} P_i\right).$$

运行一代总的时间复杂度为

$$\begin{aligned} &\sum_{i=1}^{ss} O(16N_i + mN_i + mN_iP_i + \\ &25N_i^2 + P_i + nQP_i) < \\ &\sum_{i=1}^{ss} O(16N_i + mN_i + mN_i^2 + \\ &25N_i^2 + P_i + nQP_i) = \\ &\sum_{i=1}^{ss} O(N_i^2). \end{aligned}$$

迭代运行 gen 代时, 总的时间复杂度为

$$gen \sum_{i=1}^{ss} O(N_i^2).$$

3 算例分析

为了验证本文算法的有效性, 本部分由简单到复杂, 选用文献 [6] 提供的 3 个算例进行仿真计算, 并与文献 [6] 的计算结果进行比较。

与文献 [6] 做法类似, 为了降低数值之间的相关性, $u(k)$ 采用零均值均匀分布, 最大幅值为 1.0 的随机序列; $e(k)$ 采用零均值均匀分布, 最大幅值为 0.1 的随机序列. 针对每个算例, 分别产生 100 组时间序列数据. 多目标免疫 GEP 算法的基本参数设置见表 1.

表 1 多目标免疫 GEP 算法参数设置

参数名称	参数值
各亚群规模	50
各亚群初始生命值	3
各亚群方次取值范围	{1, 2, ..., 6}
各亚群阶次取值范围	$n_y = \{1, 2, \dots, 6\}, n_u = \{1, 2, \dots, 6\}$
各细胞基因数取值范围	$n_g = \{2, 3, \dots, 10\}$
各亚群最大克隆数目	5 倍于群体规模
记忆集最大规模	20
最大运行代数	1 000

算例 1

$$\begin{aligned} &y(k) = \\ &- 0.8y(k-1) + 0.6y(k-2) + 0.5y(k-3) + \end{aligned}$$

$$0.9u(k-1) + 0.2u(k-2) + e(k).$$

应用本文提出的算法, 所获得的记忆集中包含多个互相无法比较的 Pareto 最优解, 记忆集中的最优解经合并同类项并整理后, 很多解与原系统非常相似. 限于篇幅, 对于每个算例, 这里仅列出如下 3 个计算结果:

$$\begin{aligned} y_1(k) = & -0.7983y(k-1) + 0.5913y(k-2) + \\ & 0.4827y(k-3) + 0.9123u(k-1) + \\ & 0.2082u(k-2); \\ y_2(k) = & -0.8052y(k-1) + 0.6115y(k-2) + \\ & 0.5218y(k-3) + 0.9067u(k-1) + \\ & 0.1868u(k-2); \\ y_3(k) = & -0.8104y(k-1) + 0.6083y(k-2) + \\ & 0.5161y(k-3) + 0.9028u(k-1) + \\ & 0.1961u(k-2). \end{aligned}$$

各最优解的 VAR、LTPE 及 max_PE 值如表 2 所示. 文献 [6] 结果对应的 VAR、LTPE 及 max_PE 值分别为 1.33%、3.22% 和 10.82%.

表 2 算例 1 的部分最优解精度指标值 %

最优解	VAR	LTPE	max_PE
y_1	1.55	5.17	9.55
y_2	1.81	5.02	11.88
y_3	1.20	4.17	12.15

算例 2

$$\begin{aligned} y(k) = & 0.5y(k-1) + 0.4y^2(k-2) + \\ & 0.6u(k-1) + 0.2u^2(k-2) + e(k). \end{aligned}$$

应用多目标免疫 GEP 算法的部分计算结果如下:

$$\begin{aligned} y_1(k) = & 0.5078y(k-1) + 0.3985y^2(k-2) + \\ & 0.6058u(k-1) + 0.2005u^2(k-2); \\ y_2(k) = & 0.5039y(k-1) + 0.3863y^2(k-2) + \\ & 0.5952u(k-1) + 0.2106u^2(k-2); \\ y_3(k) = & 0.4924y(k-1) + 0.4047y^2(k-2) + \\ & 0.6115u(k-1) + 0.1963u^2(k-2). \end{aligned}$$

各最优解的 VAR、LTPE 及 max_PE 值如表 3 所

示. 文献 [6] 结果对应的 VAR、LTPE 及 max_PE 值分别为 0.70%、0.86% 和 2.00%.

表 3 算例 2 的部分最优解精度指标值 %

最优解	VAR	LTPE	max_PE
y_1	0.41	1.01	2.44
y_2	0.50	0.67	1.62
y_3	0.75	0.76	1.85

算例 3

$$\begin{aligned} y(k) = & -0.4y(k-1) + 0.2y(k-2)y(k-3) + \\ & 0.4y(k-4) + 0.8u^3(k-1) + 0.2y^2(k-2) + \\ & 0.3u(k-3) + e(k). \end{aligned}$$

应用多目标免疫 GEP 算法的部分计算结果如下:

$$\begin{aligned} y_1(k) = & -0.4029y(k-1) + 0.1737y(k-2)y(k-3) + \\ & 0.3961y(k-4) + 0.8536u^3(k-1) + \\ & 0.2318y^2(k-2) + 0.3062u(k-3) + \\ & 0.0834y(k-1)y^2(k-2); \\ y_2(k) = & -0.3684y(k-1) + 0.2247y(k-2)y(k-3) + \\ & 0.4263y(k-4) + 0.7829u^3(k-1) + \\ & 0.1842y^2(k-2) + 0.2751u(k-3); \\ y_3(k) = & -0.3971y(k-1) + 0.1863y(k-2)y(k-3) + \\ & 0.4056y(k-4) + 0.8285u^3(k-1) + \\ & 0.2139y^2(k-2) + 0.2963u(k-3) + \\ & 0.1317u^3(k-2)u(k-1) + \\ & 0.0657u(k-1)y(k-2)u(k-2). \end{aligned}$$

各最优解的 VAR、LTPE 及 max_PE 值如表 4 所示. 文献 [6] 结果对应的 VAR、LTPE 及 max_PE 值分别为 3.43%、15.47% 和 21.72%.

表 4 算例 3 的部分最优解精度指标值 %

最优解	VAR	LTPE	max_PE
y_1	2.14	4.14	12.90
y_2	1.98	3.15	9.27
y_3	2.64	2.97	10.31

将本文的计算结果与文献 [6] 的结果相比较, 可以看出: 对于算例 1, 本文算法的结果有些优于文献 [6] 的结果, 有些劣于文献 [6] 的结果, 随着算例复杂度的增加, 本文算法的优异性能逐步显现; 对于算例 3, 文献 [6] 的结果明显劣于本文算法的运算结果,

这也直接表明了加权法得到的最优解不一定是多目标意义上的最优解。

4 结 论

为将GEP更好地应用于多项式NARMAX模型的辨识,本文将免疫细胞亚群划分和生命周期理论引入GEP算法,提出了多目标免疫GEP算法。新算法可以在缺乏先验知识的情况下辨识出正确的多项式NARMAX模型,与已有算法相比,新算法的主要创新点在于以下几方面:

1) 算法新定义了基于基因库的细胞创建算子和适合于NARMAX模型辨识的基因操作算子,细胞个体构成方式和基因操作方法非常灵活;

2) 算法采用了分群策略,并对各亚群赋予了生命指标,亚群的数目和各亚群的参数根据运行的中间结果进行调整,从而可以实现模型参数的自适应搜索;

3) 基于多目标优化的最优解不惟一这个特点,根据各亚群的解在最优解集中所占的比例来衡量各亚群对应参数的优劣,从而可以引导算法在缺乏先验知识的情况下,尽快将优化搜索集中到阶次和方次正确的亚群上。

仿真结果表明,算法可以很快集中到合适的模型结构上,而且多目标优化的结果,使得决策者可以综合考虑各目标的要求,提高了预测结果的可信度。

参考文献(References)

- [1] Chen S, Billings S A. Representations of nonlinear systems: The NARMAX model[J]. *Int J of Control*, 1989, 49(3): 1013-1032.
- [2] Mohamed M, M'hiri R. An approach to polynomial NARX/NARMAX systems identification in a closed loop with variable structure control[J]. *Int J of Automation and Computing*, 2008, 5(3): 313-318.
- [3] 王晓, 谢剑英, 贾青. 非线性NARMAX模型结构与参数一体化辨识的改进算法[J]. *信息与控制*, 2000, 29(2): 102-110.
(Wang X, Xie J Y, Jia Q. New modified integrated algorithm for structure determination and parameter estimation for nonlinear stochastic systems[J]. *Information and Control*, 2000, 29(2): 102-110.)
- [4] Ferreira C. Gene expression programming: A new adaptive algorithm for solving problems[J]. *Complex Systems*, 2001, 13(2): 87-129.
- [5] Wolpert D H, Macready W G. No free lunch theorems for optimization[J]. *IEEE Trans on Evolutionary Computation*, 1997, 1(1): 67-82.
- [6] Yan Bai, Yaochun Zhu, Yiheng Jiang. A new nonlinear system identification method using gene expression programming[C]. *Proc of the 2007 IEEE Int Conf on Mechatronics and Automation*. Harbin, 2007: 2951-2956.
- [7] 龚非力. 医学免疫学[M]. 北京: 科学出版社, 2003: 18-66.
(Gong F L. *Medical immunology*[M]. Beijing: Science Press, 2003: 18-66.)
- [8] 元昌安, 彭昱忠, 覃晓, 等. 基因表达式编程算法原理与应用[M]. 北京: 科学出版社, 2010: 21-69.
(Yuan C A, Peng Y Z, Qin X, et al. *Theory and applications for gene expression programming*[M]. Beijing: Science Press, 2010: 21-69.)
- [9] Vazquez K R, Fonseca C M, Fleming P J. Identifying the structure of nonLinear dynamic systems using multiobjective genetic programming[J]. *IEEE Trans on Systems, Man, and Cybernetics—Part A: System and Humans*, 2004, 34(4): 531-545.
- [10] 朱明放, 唐常杰, 代术成, 等. 基于中性突变的朴素基因表达式编程[J]. *计算机研究与发展*, 2010, 47(2): 292-299.
(Zhu M F, Tang C J, Dai S C, et al. Naive gene expression programming based on genetic neutrality[J]. *J of Computer Research and Development*, 2010, 47(2): 292-299.)
- [11] Zhou Xia, Shen Jiong, Sheng Jianxian. An immune recognition based algorithm for finding non-dominated set in multiobjective optimization[C]. *IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application*. Wuhan, 2008: 305-310.
- [12] 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, 2003: 68-77.
(Zhang W X, Liang Y. *Mathematical foundation of genetic algorithm*[M]. Xi'an: Xi'an Jiaotong University Press, 2003: 68-77.)

(责任编辑: 曹洪武)