

基于改进的基因表达式编程在重金属形态获取中的应用

李俊霞, 刘富军, 张智攀

(河北工程大学信息与电气工程学院, 河北 邯郸 056038)

摘要: 本文研究了基因表达式编程在重金属形态获取中的应用问题. 利用改进的基因表达式编程和 Shannon 信息熵方法, 验证了算法的高效性, 获得了重金属形态预测模型并从该模型中获取了重金属的形态知识的结果. 该新模型方法还可广泛用于其他时间序列预测问题的研究.

关键词: 基因表达式编程; 重金属形态预测建模; 跳跃基因表达式编程; 信息熵; 知识获取

MR(2010) 主题分类号: 68Q05; 68Q30

中图分类号: N945.12

文献标识码: A

文章编号: 0255-7797(2016)04-0831-10

1 引言

焚烧作为城市生活垃圾 (MSWI) 的处理方法之一, 产生的垃圾底灰中存在微量的重金属可对周围的环境造成危害. 这时为探索城市生活垃圾焚烧底灰的安全处置和再利用提供理论依据就变得很重要. 提出一种精确评估和预测重金属形态的模型来预测重金属的形态从而减少重金属对环境的 P 浓度从街區研究重金属的形态随时间变化的机制有如下几种方法: 直接测定法^[1,2]、模型计算法^[3]、化学提取 (连续提取)^[4]. 其中文献 [1] 通过 Anodic Stripping Square Wave Voltammetry 方法直接测定 HA 和 FA 和 Pb (II) 的含量来预测土壤中的重金属含量多少, 文献 [2] 是直接测定 HA 和 FA 和 Cu (II)Ph (II) 的浓度来预测重金属的形态对环境的影响. 文献 [3] 是通过溶液化学平衡计算模型来研究重金属在环境中的形态和迁移转化机理. 文献 [4] 主要通过观察重金属 Lead, zinc and copper 在 12 周后的浓度和 PH 的减少来测量重金属的形态在环境中的变化. 但由于影响重金属形态变化的因素众多 (PH 值、氧化还原电位 (ORP)、有机/无机吸附质等), 这三种方法的研究结果存在不唯一性和不完备性, 这给重金属迁移转化机理的研究带了不可逾越的困难.

本文通过对城市生活垃圾焚烧底灰中的重金属形态随时间变化的数据进行挖掘和形态变化的时间序列建模, 提出一种新型的基于 JM-GEP 和信息熵结合的重金属形态预测模型. 新模型适合应用在重金属 (Cr 的 PH 和含量) 形态随时间变化的时间序列问题上, 克服了以往的不足. 为了更好预测重金属的形态随时间变化的特点, 本文首次将跳跃基因融入 GEP 中使算法具有更高、更快的搜索能力, 并用信息熵理论来验证算法的合理性. 通过以下三方面对改进算法进行分析: 算法的种群多样性; JM-GEP 对重金属随时间变化的数据序列进行挖掘计算模型; 和传统 GEP 和其他模型对比来评价新模型.

*收稿日期: 2014-07-29 接收日期: 2014-11-04

基金项目: 国家自然科学基金资助 (41373101).

作者简介: 李俊霞 (1989-), 女, 河北邯郸, 硕士, 主要研究方向: GEP 算法, 重金属形态预测模型.

2 GEP 基础

GEP 是结合了 GAs(简单线性染色体串) 和 GP(树形结构表示计算机程序) 的简单思想并在此基础上发展用表达树来表示算法的计算机程序等. 但 GEP 算法的核心主要包括适应度函数、选择算子、变异算子、交叉和转座算子等, 每个算子对保持种群多样性来说都是必不可少的. 但它仍存在种群多样性减弱的缺陷. 为了保持种群多样性. 本文首次提出将跳跃基因(算子) 融入基因表达式编程中. GEP 的适应度函数采用相对误差, 如下式所示:

$$f_i = \sum_{j=1}^n (R - |\frac{P_{(i,j)} - T_j}{T_j} \times 100|) = 0. \quad (2.1)$$

2.1 跳跃基因

跳跃基因(也叫转座子) 是 1940 年代美国的 McClintock 对玉米的研究时发现的. 她观察到, 并非所有的染色体中都能跳跃, 其中的一条 9 号染色体往往能被打破且代代相传下去, 但是断裂的地方总是相同的. 她还发现非自治转座子(Ds) 在激活情况下能够从一个地方跳到另一个地方在相同的或不同的染色体之间, 这种自治转座子被称为激活转座子(Ac). Ds 本身不能跳, 除非 Ac 激活它. 但是, 激活转座子具有自身跳跃的能力. 根据实验观察, 跳跃基因可以通过两种跳跃方式: 剪切和粘贴, 复制和粘贴. 前者是剪切一段 DNA 片段, 粘贴到其他位置. 后者是把选中的 DNA 片段复制到 RNA, 然后粘贴到另一个位置, 原有的基因位保持不变. 无论哪种跳跃方式, 跳跃基因在一个染色体间的跳跃都会影响其他基因组的遗传行为.

2.2 跳跃基因图例

基因的跳跃往往是随机产生的, 而不是确定的, 跳的过程既不是流线型也不能提前计划. 因此, 跳跃基因的行为类似于许多其他基因操作, 都是随机发生的. 跳跃基因换位发生在选择操作之后变异之前. 选择算子采用轮盘赌或排序策略. JM-GEP 的实现是这样的, 每条染色体都有一些被选中的连续基因叫做转座子. 转座子的数量在一条染色体中可以大于 1. 本文选择每种跳跃方式转座子的长度均为 3, 且转座子的位置是随机分配的, 但可以在相同或者不同的染色体之间进行跳跃. 剪切和粘贴的实际操作是剪切被选中位置的转座子, 粘贴到一个新的插入位置. 复制和粘贴操作是复制被选中的转座子, 插入到一个新位置, 原基因位保持不变. 跳跃操作如图 1, 2 所示, 一个完整的程序流程图 JM-GEP 算法如图 3 所示.

图 1 演示了同条染色体上父代染色体 P 经过 CYP 元素跳跃产生子代 S 的过程. 这里也选择长度为 3 的 CYP 元素(第 9-11 位, 粗体部分) 产生子代 S, P 的第 3-5 位置被覆盖^[5]. 同时也演示了 P 经过 CTP 元素跳跃产生子代 S, 插入到第 3 个位置, P 的第 9-11 位置被截掉同时第 3 个位置之后的元素依次后移. 跳跃基因的跳跃位置、插入位置均是随机产生的, 以及跳跃基因的个数、长度要根据实际问题设定.

图 2 演示了不同染色体上 P 经过 CYP 元素跳跃产生子代 S 的过程, 分别选择父代染色体 P1 的长度为 3 的元素(第 5-7 位置, 粗体部分) 和染色体 P2(第 0-2 位置的元素, 粗体部分), p1 的第 1-3 位置被覆盖及 p2 的第 0-2 位置被覆盖. 同理, 剪切粘贴操作是截掉 p1 和 p2 长度为 3 的元素, 插入到第 1 个位置和第 0 个位置, p1 和 p2 被截掉的部分由后续位补上.



图 1: 同条染色体上的跳跃操作



图 2: 不同染色体上的跳跃操作

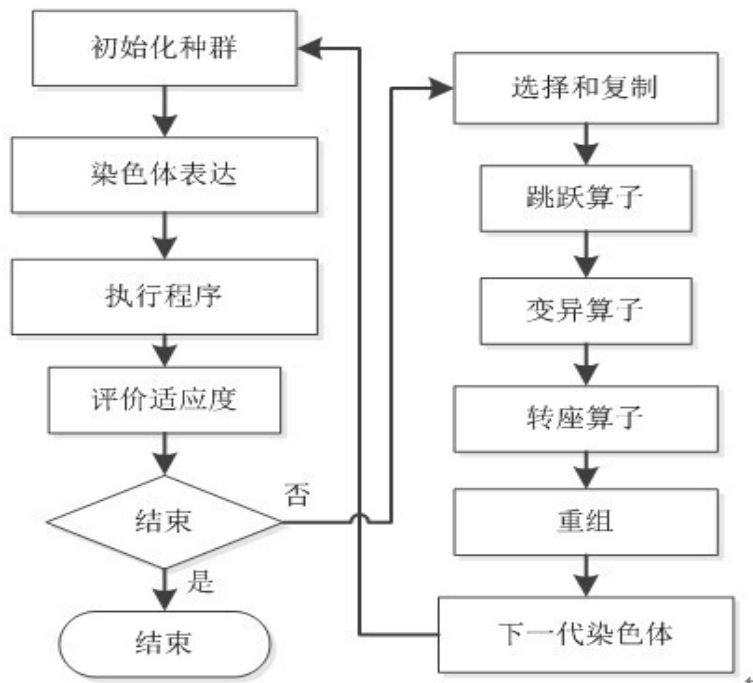


图 3: JM-GEP 算法总体流程图

2.3 跳跃算子的影响机制

历年来遗传算子的研究大都集中在纵向和横向遗传两个方面^[7], 每一个传统的遗传算子只采用纵向传播的基因形式代代相传 (即从父母到孩子). 然而, 跳跃算子引入一种横向传播的形式. 这种类型的基因传播可以作用在同条染色体或不同染色体之间. 下面描述的两中跳方式的影响.

剪切和粘贴操作通过大量染色体变化的可能性来得出一个新的遗传算子, 此方法是创造和尝试新的基因的更有效的策略. 更重要的是, 转座子通过剪切和粘贴的方式和宿主基因换位的环境下, 自然选择不仅要考虑到物理环境和其他物种的环境, 还要考虑染色体自身的微环境. 剪切和粘贴的转座子仅以半自治方式存在染色体内, 他们主要负责自身复制, 不能以相同的染色体作为宿主.

至于复制和粘贴操作, 转座子不仅可以发生跳跃, 也可为基因组携带交换信息, 使 DNA 片段在染色体中的插入位置两两交换信息. 这些染色体的 DNA 片段能够通过重组操作与不相关的基因交换信息. 因此, 复制和粘贴操作最终有利于染色体的显性性状, 此外, 重组概率的增加为基因探索提供了一个更有针对性的战略, 而不是像随机突变那样在广阔的空间里徘徊.

2.4 跳跃基因环境选择

众所周知, 能发生跳跃行为的基因在很大程度上取决于它生存的环境, 特别是在有压力的环境之下. 当基因感觉到压力时, 就会发生跳跃. JM-GEP 就是使多个压力诱导跳跃基因发生跳跃行为. 在进化过程中, 环境的选择在进化计算中扮演着一个重要的角色. 一个强大的选择压力可能导致过早收敛, 然而较弱的选择压力可能得不到最优解. JM-GEP 能够施加适当的选择压力, 使算法在找到全局最优解的同时保持种群多样性. 因为新的遗传算子可以执行水平转变, 尤其是当多个压力诱导时. 因此, 它创造了更多的机会来实现更好的收敛性和多样性, 以避免过早收敛.

3 JM-GEP 算法

提出 JM-GEP 算法是为避免算法出现早熟和陷入局部收敛的缺陷.

(1) 跳跃算子的改进由上可知, 跳跃算子中跳跃率、跳跃方式的选择率、跳跃位置、插入位置的选择都是随机的, 大大降低了跳跃基因的跳跃概率同时也破坏了种群的多样性. 因此本文提出动态变化的跳跃概率, 使得改进的跳跃算子具有自适应性. 首先按适应度将染色体从高到低排序, 然后每 20 个为一组, 每组染色体的跳跃基因的跳跃概率函数定义为如下式 3.1 所示:

$$p_i = \alpha \left(1 - \frac{i}{N/20} \right), \quad i = 1, 2, \dots, N/20, \quad (3.1)$$

式中 α 为概率调节常数, 且 $\alpha \in [0, 0.15]$.

(2) 适应度函数用于评价重金属的形态随时间变化的两组数据的吻合程度, 统计学中, 更

多是采用判定系数 (Coefficient of Determination), 记作 (R -Squared), 计算公式为

$$R^2 = \frac{\sum_{i=1}^n (f'_i - f_i)^2}{\sum_{i=1}^n (f_i - f_{\text{ave}})^2}, \quad (3.2)$$

其中, f_i 为观测数据, f'_i 为拟合数据, f_{ave} 为观测数据的均值.

3.1 JM-GEP 算法形式化描述

跳跃算子的描述:

第一步 随机生成 N 个染色体种群 $\{C_1, C_2, \dots, C_N\}$, 评价每个染色体的适应度函数, 然后按适应度大小从高到低进行排序, 每 20 个为一组, 计算每组的适应度函数.

第二步 选择每组中适应度大的个体进行变异操作.

第三步 设跳跃率 $P_{j,t} = 0.1$, 跳跃方式的选择率为 $JmPnStprand$. 如果随机选择的个体跳跃率 $p_1 < P_{j,t}$, 则发生跳跃行为; 在此基础上, 如果 $JmPnStprand > P_{j,t}$, 发生剪切和粘贴操作; 否则, 发生复制和粘贴的操作.

这里跳跃基因的选择位置及长度和要插入的位置都是随机产生的.

JM-GEP 算法的形式化描述:

第一步 初始化种群, 产生数目一定的构成染色体的基因组;

第二步 计算每代的适应度函数值, 并判断是否达到最优解, 若是, 输出最优个体和最优解的值, 否则, 转第三步;

第三步 按轮盘赌法或锦标赛法选择个体进行遗传操作;

第四步 进行跳跃操作, 并计算每一代的跳跃率, 生成新个体;

第五步 按照一定的概率进行变异操作, 并生成新个体;

第六步 按照一定的概率进行转座、交叉、重组操作, 生成新个体, 返回第二步.

由上新算法的描述看出, 每一代中都加入了跳跃算子, 并计算每一代的跳跃率, 使算法有更高更强的搜索能力.

3.2 Shannon 信息熵分析 JM-GEP 种群多样性

Shannon 熵是 1948 年, 香农提出的概念, 用于解决随机变量的“不确定性”或信息的量化度量问题. 由 1.3 节可知跳跃基因的选择位置及长度和要插入的位置都是随机产生的, 所以种群多样性的测度可以用 Shannon 熵这样一个表征基因均匀度的函数来表示^[6].

针对一条基因序列, 定义各符号 (函数符和终点符) 发生跳跃的频率为 p_i .

$$p_i = \frac{1}{F+T} \quad (i = 1, 2, \dots, F+T), \quad (3.3)$$

其中 F 表示函数符集的元素个数, T 表示终结符集的元素个数, 根据最大熵原理, 此时序列熵为

$$H = - \sum_{i=1}^{F+T} p_i \ln p_i = - \frac{1}{F+T} \ln \frac{1}{F+T} = \ln(F+T) \quad (3.4)$$

达到最大值, 且不随时间变化.

在本文中, 重金属的形态随时间变化的基因序列长度 L 是一定值. 在时刻 t , 各函数符号和终结符号在该基因序列中出现跳跃的次数为 $f_{i,t}$ ($i = 1, 2, \dots, n$) ($n = F + T$), 序列熵记为 H_t , 则在 $t = 0$ 时刻, 该基因序列的序列熵为

$$H_0 = - \sum_{i=1}^n \frac{f_{i,0}}{L} \ln \frac{f_{i,0}}{L}. \quad (3.5)$$

记跳跃率为 p_{jm} , 则在 $t + 1$ 时刻, 第 i 个符号在序列中出现跳跃的次数 $f_{i,t+1}$ 为

$$f_{i,t+1} = n f_{i,t} - \sum_{j \neq i} f_{j,t} p_{jm}. \quad (3.6)$$

因为 $f_{i,t} + \sum_{j \neq i} f_{j,t} = \sum_{i=1}^n f_{i,t} = L, t \in [0, +\infty]$, 并记 $\Delta f_{i,t} = f_{i,t+1} - f_{i,t}$, 则上式 (3.6) 可变为

$$\Delta f_{i,t} = n f_{i,t} - L. \quad (3.7)$$

由于每代的进化时间长度相对于整个进化时间很小, 将时间 t 连续化, 则式 (3.7) 可以化为如下微分方程

$$\frac{df_{i,t}}{dt} = n f_{i,t} - L. \quad (3.8)$$

解微分方程 3.8 得到

$$H_t = - \sum_{i=1}^n \frac{\frac{L}{n} + (f_{i,0} - \frac{L}{n}) e^{-nt}}{L} \ln \frac{\frac{L}{n} + (f_{i,0} - \frac{L}{n}) e^{-nt}}{L}, \quad (3.9)$$

$$\frac{dH_t}{dt} = \sum \frac{n f_{i,t} - L}{L} \ln \frac{L}{e f_{i,t}} = \ln \prod_{i=1}^n \left(\frac{L}{e f_{i,t}} \right)^{\frac{n f_{i,t} - L}{L}}. \quad (3.10)$$

由上可得, 当 $\begin{cases} L - n f_{i,t} < 0, \\ \frac{L}{e f_{i,t}} > 1 \end{cases}$ 或 $\begin{cases} L - n f_{i,t} > 0, \\ \frac{L}{e f_{i,t}} \leq 1 \end{cases}$ (本文中 ($n > e$) 恒成立), 即 $f_{i,t} < \frac{L}{n}$ 或 $\frac{L}{e} < f_{i,t} < L$ 时, 基因序列熵 H_t 值随进化代数增加而增大. 同样, 当 $\frac{L}{e} < f_{i,t} < \frac{L}{e}$ 时, H_t 随代数增加而减小, 呈下降趋势. 因此可分析得出在进化前期, 序列熵 H_t 是关于 t 的减函数 (从式 3.9 中可以看出, 它也是关于 $p_{jm} t$ (p_{jm} 为跳跃率) 的减函数, 加入跳跃算子之后 JM-GEP 过程中各代种群多样性均优于相对应时刻 GEP 中的多样性, 从而在整个进化过程中, JM-GEP 较 GEP 过程能够更快地趋向种群多样性最大值. 亦即加快了达到种群多样性最大值的速度, 使算法具有更高、更快的搜索能力, 同时保持了种群的多样性.

4 基于 GEP 和 JM-GEP 的重金属形态获取预测模型

为了详细说明基因表达式编程在重金属形态获取预测建模中的实际应用过程, 下文将针对 Cr 的含量随时间变化的数据进行演化建模, 从而验证该方法的可行性和有效性. 引用的比较实验数据均来自本校实验数据.

4.1 Cr 的含量随时间变化演化建模

如表 1 是重金属 Cr 随时间变化的时间序列 (只取前 16 个数据). 其中, t_i 表示 Cr 变化的周数, 这里平均每四周测一次, T_i 是指重金属 Cr 的含量, 即重金属下一次随时间变化的含量. 表 1 是针对 T_i 建立 GEP 及 JM-GEP 模型. 如下表 2 是进化计算用到的参数列表.

表 1: Cr 的含量随时间变化序列

x	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
t_i	0	4	8	18	24	32	36	44	48	52	56	60	64	68	72	76
T_i	0.01	0.05	0.08	0.07	0.09	0.05	0.03	0.01	0.09	0.08	0.06	0.11	0.11	0.05	0.03	0.01

表 2: GEP & JM-GEP 参数列表

参数	取值	参数	取值
种群大小	60	函数符集	{+ - */^}
基因个数	5	终结符集	{x, 0, 1, ..., 9}
头长	6	选择算子	轮盘赌
最大进化代数	1000	转座率	0.1
变异率	0.044	重组率	0.3
跳跃率		JM-GEP 采用公式 (3.1)	
适应度函数	GEP 采用公式 (2.1) ($R = 100$), JM-GEP 采用公式 (3.2)		
终止条件	最大进化代数		

在 VC6.0 和 MATLAB7.2 混合编程环境下对表 1 中的时间序列进行进化计算编程, 经过 1000 代的演化运算后, 得到适应性较好的模型结构表达式如下式 (4.1), (4.2) 所示:

$$Y_{GEP}(x) = -6.780092 + \frac{x}{1.313761} \cdot 10^{0.402417} + 2x^2 \cdot 10^{0.126865}, \quad (4.1)$$

$$Y_{JM-GEP}(x) = 1.1425 + 2x + 3x^2 + \frac{0.962584}{x} - \frac{x}{0.160436} - 10^{0.646596}. \quad (4.2)$$

4.1.1 模型仿真

图 4 和图 5 给出了模型 (4.1), (4.2) 的重金属的形态随时间变化的模型仿真图. 其中图 4 和图 5 对应的 X 轴表示表 1 中的样本 (x) 即 Cr 随时间变化的序号, Y 轴表示 T_i (即重金属随时间变化的累计数据). 由上图 4 和图 5 看出 GEP 和 JM-GEP 模型均较好的拟合了重金属的形态变化数据. 其中图 4 和图 5 对比, GEP 运行到第 350 代才找到最优解, 其适应值为 1547.791070, 耗时 10 秒. 而图 5 中的 JM-GEP 只运行到了第 100 代就得到了最优解, 最大适应值为 4747.393139, 仅耗时 4 秒. 这充分说明 JM-GEP 模型具有较高预测效率, 且拟合度更好 (适应度值表征了预测值与实际值的误差, 误差越小, 适应值越大). 可见, 采用新算法演化得到的重金属形态预测模型计算结果较 GEP 经典模型理想.

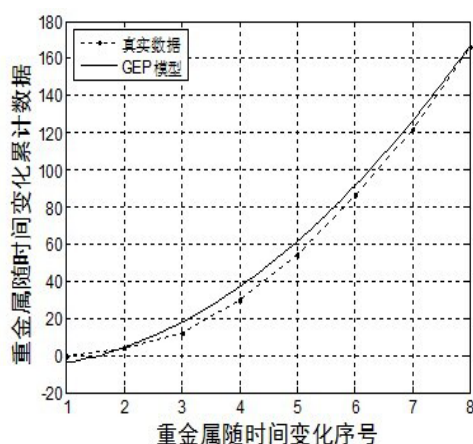


图 4: 模型 (4.1) 仿真结果

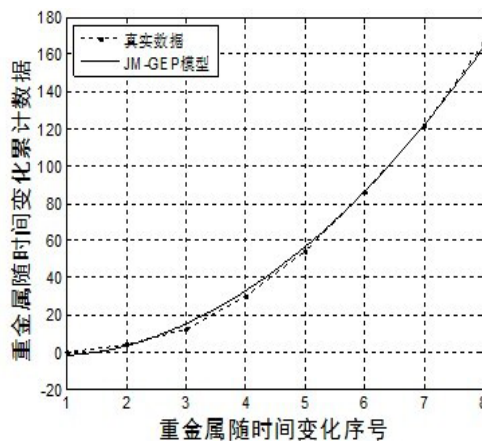


图 5: 模型 (4.2) 仿真结果

4.1.2 与其他模型进行对比评估

如表 3 是 GEP 和 JM-GEP 针对表 1 中的数据和其他模型进行对比评估的结果. 由表 3 可看出 GEP 和 JM-GEP 都较好的模拟了重金属的形态随时间变化的数据, 但 JM-GEP 较 GEP 有更好的模拟度.

表 3: 针对表 1 中的数据对比评估

评估模型	t_i	下一变化值	评估模型	t_i	下一变化值
GEP 模型	63.4839	227.4839	G-O 模型	70.2176	234.684690
JM-GEP 模型	45.720357	211.720357	Moranda 模型	68.3674	242.400123
指数模型	64.5823	251.879013	S-W 模型	71.547330	249.569820
J-M 模型	66.9940	267.505001			

4.1.3 模型拟合优良度

拟合优良度可以采用 X 平方测试或 Kolmogorov-Smimov 测试法来评估. 本文选用平方误差法来计算公式如下所示:

$$\text{error} = \sum_{i=1}^m (f_i - y_i)^2, \quad (4.3)$$

式中, f_i —新模型预测值, y_i —重金属 Cr 的含量随时间变化的序列真实值, i —数据序号; m —数据总个数. 采用拟合优良度公式 (4.3) 分别计算得出 GEP 模型 (3.10) 与 JM-GEP 模型 (4.1) 的预测值与真实值之间的误差平方和, 分别与 GP 模型、BP 模型相对比, 详细结果如表 4 所示.

表 4: 模型拟合优良度比较

评价指标	JM-GEP 模型	GEP 模型	GP 模型	BP 模型
拟合优良度	3.469e+003	4.008e+003	2.687e+003	2.4397e+003

由此可知, 采用 GEP 模型对重金属 Cr 的含量随时间变化数据进行预测, 其逼近程度良好, 经改进后的 JM-GEP 模型误差更小, 精度更高. 两模型的拟合度均优于 GP 模型, 以及训练度好的 BP 模型.

5 重金属 Cr 的含量在未来时间点的预测

由上式 (4.1), (4.2) 可得, 重金属 Cr 在未来几个时间点的含量值如下表 5 所示:

表 5: Cr 的含量预测

时间点	JM-GEP 模型	GEP 模型	GP 模型
第 17 点	799.9975	791.8059	790.88
第 18 点	895.6683	999.3339	796.89
第 19 点	996.6962	999.3339	887.02
第 20 点	1.1031e+002	1.1121e+002	0.9031e+002
第 21 点	1.2148e+002	1.2309e+002	1.0148e+002
第 22 点	1.3319e+002	1.3556e+002	1.1319e+002
第 23 点	1.4544e+001	1.4864e+001	1.2544e+001
第 24 点	1.5822e+001	1.6232e+001	1.3822e+001
第 25 点	1.7154e+001	1.7659e+001	1.6154e+001

表 5 得出 Cr 分别在第 17 点到 25 点的含量, 通过和经典模型和 GP 模型相比较, JM-GEP 模型的预测能力优于其他算法, 能更好的体现预测能力.

5.1 重金属 Cr 的含量预测模型

由上式 (4.1), (4.2) 可得, 重金属 Cr 在未来几个时间点的含量值预测模型如下图 6, 7 所示. 由表 5 和图 6, 7 中模型变化曲线对比可看出, JM-GEP 曲线较 GEP 和 GP 模型能更好

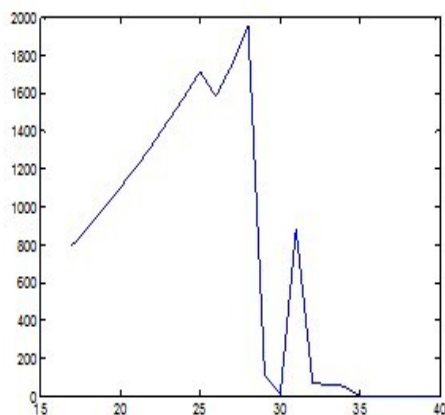


图 6: GEP 预测 Cr 含量变化曲线

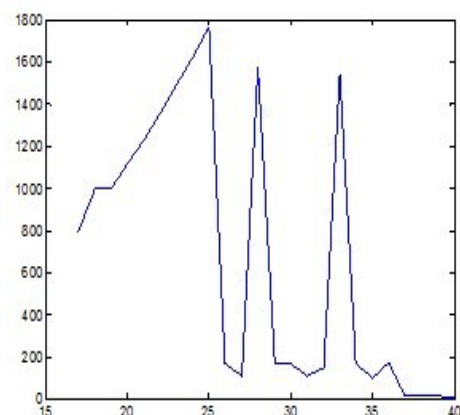


图 7: JM-GEP 预测 Cr 含量变化曲线

的反应 Cr 的含量在未来时间点的变化趋势.

6 结语

用 GEP 进行预测, 不需要知道各因素间的因果关系, 只需要提供足够的实验或者实验数据, 无须知道目标函数, 就可以达到准确预测的目的. 本节重点在于采用跳跃基因的 GEP 算法 (JM-GEP) 对重金属形态随时间变化的时间序列进行分析, 得到预测精度和拟合度更好的数学模型, 通过实例计算和分析可知, 此方法优于传统的 GEP 方法且在求解速度上也快于传统的 GEP 算法.

参 考 文 献

- [1] Quan G, Yan J. Binding constants of lead by humic and fulvic acids studied by anodic stripping square wave voltammetry[J]. Russian J. Electr., 2010, 46(1): 90–94.
- [2] Christl I, Metzger A, Heidmann I, et al. Effect of humic and fulvic acid concentrations and ionic strength on copper and lead binding[J]. Envir. Sci. Tech., 2005, 39 (14): 5319–5326.
- [3] 章骅, 何晶晶, 吕凡等. 重金属在环境中的化学形态分析研究进展 [J]. 环境化学, 2011, 30(1): 130–137.
- [4] Anirhomayoun Saffarzadeh, Takayuki Shimaoka. Impacts of natural weathering on the reformation/neiformation processes in landfilled MSW bottom ash: A geoenvironmental perspective[J]. Waste Mgt., 2011, 31(12): 2440–2454.
- [5] 浦黄忠, 甄子洋, 王道波, 刘媛媛. 用于多峰函数优化的改进跳跃基因的遗传算法 [J]. 南京航空航天大学学报, 2007, 39(6): 829–832.
- [6] 肖静. 基因表达式编程在软件可靠性建模中的应用研究 [D]. 河北: 河北工程大学硕士学位论文, 2012.
- [7] 张爱华, 曹晓刚, 钟守楠. 求多峰函数全部全局最优解的改进遗传算法 [J]. 数学杂志, 2009, 29(1): 56–60.

AN HEAVY METALS MORPHOLOGY ACQUIRED METHOD BASED ON IMPROVED GEP

LI Jun-xia, LIU Fu-jun, ZHANG Zhi-pan

(School of Information and Electric Engineering, Hebei University of Engineering,
Handan 056038, China)

Abstract: In this paper, we use the gene expression programming application in heavy metal form questions. By using an improved gene expression programming and Shannon entropy, we demonstrate the efficiency of the algorithm, and get heavy metal form prediction model and obtain the result of the heavy metal forms of knowledge from the model. The new model also can be widely used in other time series prediction problems.

Keywords: gene expression programming; JM-GEP; heavy metals prediction model; Shannon information entropy; knowledge acquisition

2010 MR Subject Classification: 68Q05; 68Q30