中国科学: 技术科学 2010 年 第 40 卷 第 5 期: 582 ~ 595
www.scichina.com tech.scichina.com
《中国科学》杂志社
SCIENCE CHINA PRESS

由硅酸盐水泥的观测数据反向萃取水化早期动力学方程

王琳[①], 杨波[②*], 赵秀阳[①②], 陈月辉[②], 常钧[③]

① 山东大学计算机科学与技术学院, 济南 250101;
② 济南大学网络环境智能计算技术山东省重点实验室, 济南 250022;
③ 济南大学材料科学与工程学院, 济南 250022
* E-mail: yangbo@ujn.edu.cn

收稿日期: 2009-12-10; 接受日期: 2010-03-18
国家自然科学基金(批准号: 60873089, 60573065)、山东省自然科学杰出青年基金(批准号: JQ200820)和新世纪优秀人才支持计划(批准号: NCET-10-0863)资助项目

摘要 早期水化反应对于硅酸盐水泥浆体微观结构的形成和强度的发展有着重要影响. 然而由于水泥水化过程中发生了多相多尺寸并且相互关联的复杂的化学和物理变化, 因此使得人工推导水化动力学方程的研究存在很高的难度. 利用基因表达式编程与粒子群算法相结合的进化计算方法从观测到的硅酸盐水泥水化程度时间序列数据中自动地萃取出了水化早期的动力学方程, 并通过GPU进行并行加速来减少运算时间. 研究显示, 根据该动力学方程得到的模拟曲线可以很好的吻合水化早期观测到的实验数据, 而且即使化学组成、颗粒尺寸和养护条件发生改变, 该方程仍然具有良好的泛化能力.

关键词
硅酸盐水泥
水化动力学方程
反向萃取
基因表达式编程
粒子群算法

研究水泥浆体形成与发展过程的规律与特性非常必要. 一方面硬化水泥石的结构与水泥浆的初始结构和发展过程有关, 要得到优良的水泥复合材料, 必须了解和控制水泥浆体的形成过程; 另一方面, 水泥复合材料的各种性能指标, 如抗压强度、流动性、凝结硬化程度、孔隙率和耐久性都直接受到水化反应的影响. 因此水泥水化反应的研究不仅有很高的理论价值, 而且具有重要的实际意义和应用前景.

硅酸盐水泥的水化可以分为5个阶段: 快速反应期、诱导期、加速期、衰退期和稳定期, 尤其以水化早期的反应最为剧烈, 通常在24 h内即进入了衰退期或稳定期. 由于硬化浆体的性能与水化早期浆体结构的形成密切相关, 而水泥浆体的初凝时间与终凝时间则大致分别发生在诱导期和加速期的终止阶段[1], 因此早期水化反应对于水泥浆体微观结构的形成和强度的发展有着重要影响.

水化过程的仿真与建模是一种理解水化反应的化学机理、模拟水化泥浆微观结构的发展以及预测水泥性能的有效工具[2]. 水泥的水化过程可以通过多种方法进行建模, 其中最常用的是构建水化过程的动力学方程[3,4]. 水化程度是指在一定时间内发生水化作用的量和完全水化量的比值, 用α来表示. 水化动力学方程是水化程度α关于时间t的微分方程, 其可以揭示出随着硅酸盐水泥水化过程的进行, 水化程度随时间的发展趋势. 以水化动力学方程为基础, 也可以搭建起水泥浆体微观结构[5]和各种性能指标[6]的

引用格式: Wang L, Yang B, Zhao X Y, et al. Reverse extraction of early-age hydration kinetic equation from observed data of Portland cement. Sci China Tech Sci, 2010, 53: 1540–1553, doi: 10.1007/s11431-010-3228-9

动态模型.

Tomosawa等人[3]针对硅酸盐水泥提出了一种水化动力学的单一方程模型, 该方程通过 4 个速率常数来确定初始不透水层的形成与分解, 活跃的化学反应过程以及温和的扩散控制过程. Krstulovic等人[4]则提出了一种由 3 个方程联立组成动力学模型, 认为水泥材料的水化反应有 3 个基本过程: 结晶成核与晶体生长(NG)、相边界反应(I)和扩散(D), 3 个过程可同时发生, 但是水化过程的整体发展取决于其中最慢的一个反应过程. 阎培渝等人[7]在Krstulovic模型的基础上对上述 3 个基本过程进行了表征, 得到反应速率常数K、反应级数n和表观活化能E_a等动力学参数以及各反应阶段的反应速率与反应度的关系. 然而, 由于硅酸盐水泥水化过程中发生了多相多尺寸并且相互关联的复杂的化学和物理变化, 因此使得人工推导水化动力学方程的研究存在很高的难度. 为了计算的简便, 当前的研究通常忽略水泥的快速反应期而只从诱导期开始模拟[7], 并且假设水泥颗粒为球型进行推导[3], 因此导致建立的模型无法完整精确的表述作为一个连续系统的水泥水化过程.

最近有研究表明, 进化计算方法[8,9]可以自动化的从观测到的实验数据与结果中反向萃取出自然法则与公式. Schmidt等人[10]利用遗传编程算法(GP, genetic programming)[8]搜索来自多种物理系统的移动轨迹数据. 该方法证实了GP算法可以在没有任何物理、动力学与几何学相关知识的情况下发现哈密尔顿函数与拉格朗日函数等几何的或动量守衡的自然法则. Koza等人[11]利用GP算法完成了一条代谢通路的反向工程, 该方法使用E-CELL细胞模拟模型所提供的数据并且考虑了反应类型的信息. Kikuchi等人[12]利用遗传算法(GA, genetic algorithm)[9]重建了基因网络. Qian等人[13]利用GP算法与Kalman滤波器推导出了基因调控网络的非线性微分方程. Chakraborty[14]利用进化计算实现了固态氧化物燃料电池的静态与动态建模. Iba在文献[15]中详细描述了如何由观测到的时间序列数据推导出用常微分方程系统表示的因果关系模式. 该方法表明了遗传编程在生物信息学、化学反应建模以及控制理论等领域的实际应用中都具有了良好的效果.

人工推导水化动力学方程面临的困难与进化计算在自然法则发现方面的进步促使我们探索利用进化计算的方法从观测到的硅酸盐水泥水化程度时间序列数据中自动化的寻找水化早期动力学方程. 选择基因表达式编程算法(GEP, gene expression programming)[16]来搜索水化早期动力学方程的函数形式. GEP是一种用于自动优化计算机程序或函数的新型进化算法, 其在功能上与GP相同, 但是与GP相比, GEP融入了GA线性编码的优点, 从而避免了GP中复杂的树结构操作, 因此其搜索效率明显高于GP[16]. 对于给定动力学方程中的控制系数搜索, Iba[15]使用Levenberg-Marquardt方法来搜索含有超越函数的微分方程的系数组合, 然而这种传统的解析方法容易陷入局部极值, 为了解决这一问题, 选择粒子群优化算法(PSO, particle swarm optimization)[17]来寻找给定动力学方程的全局最优系数组合. 而在求动力学方程的数值解时选择的是经典的四阶龙格库塔法(RK4, fourth-order runge-kuta method).

重建工作的流程如图 1 所示, 首先采用水化热法进行实验, 获取水泥的水化反应放热随时间的变化情况. 并根据热量的释放建立水化程度发展的时间

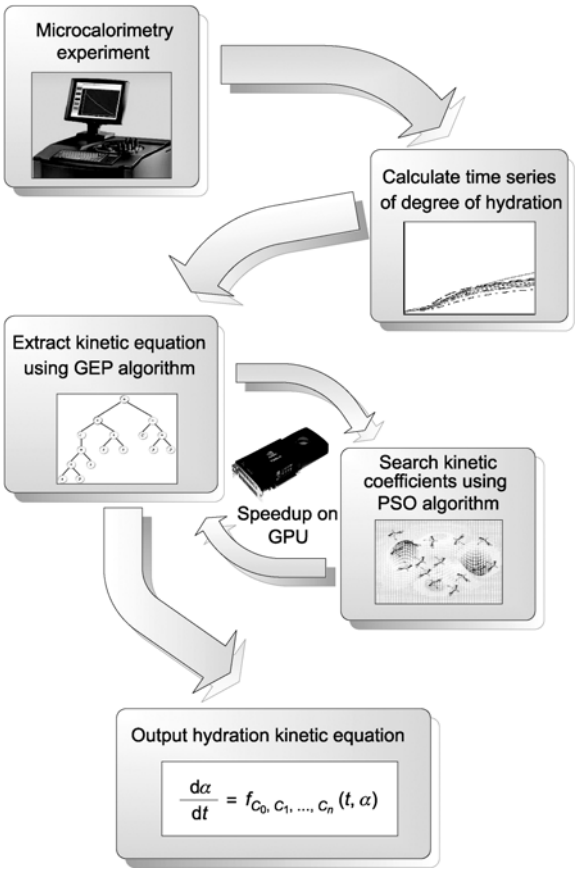

图 1 水化动力学方程重建流程

序列数据. 然后以该时间序列数据作为教师信号, 利用 GEP 算法迭代进化动力学方程的函数形式, 其间对每个生成的方程, 用PSO算法优化其控制系数, 该过程反复进行直到寻找到最优的动力学方程及其控制系数为止. 为了提高进化速度, 使用高性能计算处理器 (GPU, graphics processing unit) 对算法进行加速. 上述过程执行完毕后, 得到硅酸盐水泥水化早期的动力学方程.

1 水化程度时间序列获取实验

实验需要得到硅酸盐水泥的水化程度随时间演化的时间序列数据. 测定水化程度有 2 类方法: 直接法与间接法. 直接法包括岩相分析与X射线分析等, 但是这些方法耗时较长, 无法实时测量水泥浆体的水化程度. 间接法包括化学结合水法、氢氧化钙定量测试法、水化热法等, 考虑到化学结合水法和氢氧化钙定量测试法会对水泥浆体产生破坏性效果, 无法针对同一样本进行连续测定, 实验选择可以对同一水泥浆体样本进行连续观测的水化热法[18]来得到水化程度演化的时间序列.

水泥在加水以后, 会发生一系列的化学与物理变化, 其间释放出热量, 而热量的释放过程又与水化反应的发展有着密切的联系. 根据这一规律, 可以利用下式来确定水化程度α随水化反应进行的变化情况

$$\alpha(t)=\frac{Q(t)}{Q_{\max}}, \tag{1}$$

其中, $\alpha(t)$是在t时刻的水化程度, $Q(t)$是到t时刻为止水化反应所释放出的热量, $Q_{\max}$表示在完全水化后, 水泥样本可以释放出的总热量. 水泥到t时刻为止所释放出的热量$Q(t)$可以利用水化热分析仪直接测量. 而单位质量的硅酸盐水泥理论上可以释放出的总热量$Q_{\max}$(J/g)则可以通过样本的化学组成进行估计[19]

$$\begin{aligned}Q_{\max}=&500p_{C_3S}+260p_{C_2S}+866p_{C_3A}+420p_{C_4AF}\\&+624p_{SO_3}+1186p_{fCaO}+850p_{MgO},\end{aligned} \tag{2}$$

其中, p_i是第i种成分在样本中所占的质量百分比. 通过(1)和(2)式, 可以得到正在水化的硅酸盐水泥的水化程度随时间演化的时间序列.

实验所用硅酸盐水泥样品由山东省水泥质量监督检验站提供, 分别编号为A, B, C, D. 表 1 列出了 4 种样品所含氧化物的质量百分比、比表面积、200 目筛筛余以及根据Bogue公式[20]计算出的水泥的 4 种主要物相的质量百分比和根据公式(2)计算而得的水泥样品的理论放热值$Q_{\max}$.

实验所使用的水化热分析仪是 TAM Air 八通道热导式等温微量量热仪. 取 3 份样品 A、3 份样品 B、3 份样品 C 和 1 份样品 D, 每份各取 2 g 水泥粉末, 按照不同的水灰比和反应温度(表 2)在水化热分析仪中进行水化反应. 水化反应时间选择为 24 h, 一方面, 由于硅酸盐水泥的快速反应期一般在 15 min 内结束, 如果所选择的时间段超过 24 h, 则快速反应期的时间点在时间序列中的比例不足 1%, 在方程进化时无法提供充足的教师信号; 另一方面, 如果所选择的时间段小于 24 h, 则水化过程尚未进入稳定期. 设定采样时间间隔为 1 min, 共得到 10 组水化热数据, 每组数据包含 1440 个时间点, 通过公式(1)可计算出水化程度的时间序列(图 2). 前 8 组样本用于寻找硅酸盐水泥水化早期的动力学方程, 最后 2 组样本用于验证所得到的方程的泛化能力.

表 1 水泥样品的各种参数

Specimen	CaO (%)	SiO_2 (%)	Fe_2O_3 (%)	Al_2O_3 (%)	MgO (%)	SO_3 (%)	K_2O (%)	Na_2O (%)	*f*CaO (%)
A	63.89	22.58	3.03	4.67	2.46	1.75	1.2	0.17	0.64
B	63.74	20.93	3.77	4.45	2.51	1.56	0.93	0.23	1.01
C	64.74	20.72	3.74	4.66	3.34	1.38	0.54	0.19	1.25
D	64.00	21.73	3.50	4.68	2.98	1.75	0.44	0.19	0.64

Specimen	C3S (%)	C2S (%)	C3A (%)	C4AF (%)	Specific surface (cm^2/g)	Sieve residue on 74 μm (%)	$Q_{\max}$ (J/g)
A	45.12	30.79	7.25	9.21	4088	3.56	446.56
B	56.50	17.47	5.42	11.46	3475	6.44	466.05
C	60.34	13.97	6.03	11.37	4397	3.63	489.79
D	51.29	23.70	6.49	10.64	3633	7.94	462.74

表 2　水化热实验的实验参数

Series	Specimen	Water-to-cement ratio (w/c)	Reaction temperature (°C)
0	A	0.35	25
1	A	0.45	30
2	A	0.55	25
3	B	0.35	25
4	B	0.45	30
5	B	0.55	25
6	C	0.35	25
7	C	0.55	30
8	C	0.45	30
9	D	0.45	25

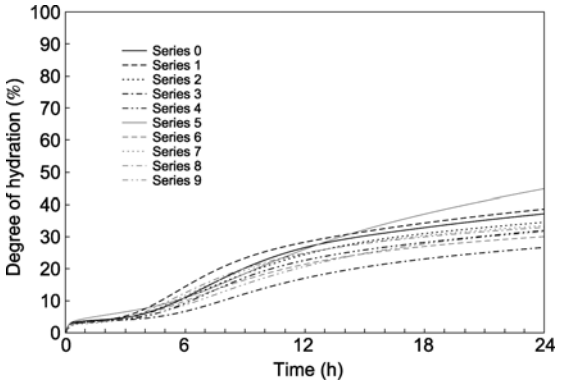

图 2　硅酸盐水泥水化程度在前 24 小时的演化情况

2　自动萃取水化早期动力学方程

以上一节中所获得的水化程度时间序列数据作为教师信号, 可以利用进化计算方法反向重建出硅酸盐水泥水化早期的动力学方程, 本章节对重建方法进行了完整的描述. 首先简介了所使用的 2 种进化计算方法: 用于生成动力学方程函数形式的 GEP 算法与搜索方程系数的 PSO 算法. 之后在 2.3 节中详细描述了动力学方程的萃取方法与过程. 为了提高方程重建过程的运行速度, 在最后一节中重点介绍了计算加速方法.

2.1　GEP算法

GEP算法最早是由Candida Ferreira在 1999 年提出的[16], 是一种用于自动优化计算机程序或函数的新型进化算法. GEP算法模拟遗传生物学中的基因表达过程, 操纵由一组个体所组成的种群, 通过使用一种或多种遗传算子, 按照个体适应值的不同对种群进行不断的选择和再生来实现种群的进化以寻找最优解. GEP算法在个体形式上与遗传算法相似, 其所使用的个体称作染色体, 是由线性的含有多个基因的字符串所组成. 优化过程中, GEP算法通过选择、变异、插串、根插串、基因迁移、基因交叉以及单点和两点交叉等算子来使种群得到进化. 而就算法的功能与目的来说, GEP又与遗传编程相似, 可以寻找并发现解决问题的公式或程序. GEP算法的每个染色体是用K-表达式来表示的表达式树, 该表达式树代表了一个公式或者一段程序, GEP的优化过程即是通过各种算子的操作来实现公式或程序的不断进化的过程. 事实上, GEP算法结合了遗传算法的线性编码和遗传编程算法的表达式进化这两者的优点, 使得其可以通过操作简单的线性字符串来实现复杂的公式或程序进化. 因此, 在解决复杂问题时, GEP算法比传统的遗传编程算法在效率上要高出许多[16].

GEP 算法的染色体是由一个或多个基因所组成的固定长度的线性符号串序列. 符号串中的符号来自于 2 个集合: 函数集 **F** 与元素集 **T**. 其中函数集包含所需要的原子函数, 如+, –, ×, ÷等, 而元素集则包含变量与常数, 如 x, y, C 等(C 代表常数项). 符号串序列的组织形式被称作 K-表达式, 是通过对一棵表达式树按照从上到下、从左到右进行顺序遍历所得到的, 而表达式树是表达式的直观表示方法, 因此通过 K-表达式可以使得线性的染色体与表达式之间建立一一对应的关系. 一个染色体可以由多个基因所组成, 不同的基因之间使用连接函数进行连接. 其中每个基因都包含有头部和尾部, 头部可以从函数集或者元素集中取值, 而尾部则只可以从元素集中取值. 对于一个具体问题, 需要首先选择头部长度 h, 然后通过下式计算出尾部长度 t, 其中 n 是函数集合中函数的最大参数个数:

$$t = h \times (n-1) + 1. \tag{3}$$

图 3 是一个染色体及其所代表的表达式树和表达式的例子. 该染色体包含有 3 个基因, 连接函数设置为＋. 其中每个基因的头部长度设置为 5, 最大函数个数为 2, 因此可以由公式(3)计算出尾部长度为 6, 所以基因的长度为 11, 染色体的总长度为 33. 注意在每个基因中, 仅有前面的部分符号(图中染色体中用黑色表示)已经可以形成一棵完整的表达式树, 后

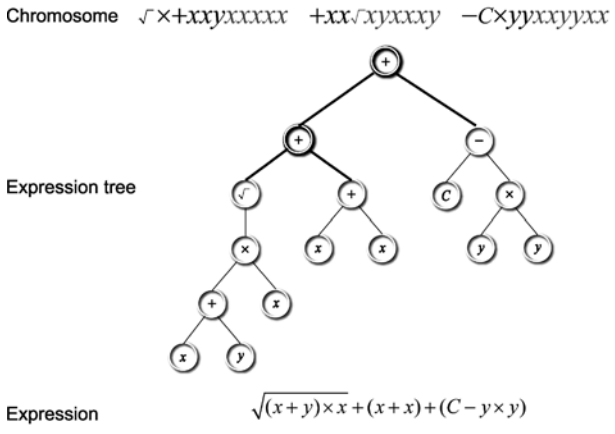

图 3　染色体及其所对应的表达式树和表达式

面的部分(图中染色体中用灰色表示)并未出现在表达式树中, 这一部分被称作无编码区域. 无编码区域的作用是使得在进化过程中, GEP 可以在不增加复杂的限制条件的情况下使用遗传算子, 使得基因和染色体总是可以被解读为符合语法规则的表达式.

2.2　PSO算法

粒子群优化算法是由Kennedy等人[17]受到鸟类觅食过程的启发, 在 1995 年开发的一种全局数值搜索算法. 在PSO中, 每个优化问题的解都是搜索空间中的一个粒子, 所有的粒子形成了一个种群. 每个粒子由一个速度来决定它们飞翔的方向和距离, 然后粒子们就追随当前的最优粒子在解空间中搜索. 当经过一定步数的迭代, 最优粒子的位置向量即是问题的近似最优解.

PSO 算法首先用一群随机粒子初始化种群, 然后通过迭代找到最优解. 在每一次迭代中, 粒子通过跟踪2个极值来更新自身: 粒子本身所找到的最优解, 称做个体极值 ***pbest***; 整个种群目前找到的最优解, 称做全局极值***gbest***. 每次迭代时, 在更新完这2个极值后, 粒子根据如下的公式来更新自己的速度和新的位置

$$\begin{aligned}&\boldsymbol{v}_{k+1}=\varphi_0 r\boldsymbol{v}_k+\varphi_1 r\left(\boldsymbol{pbest}_k-\boldsymbol{x}_k\right)+\varphi_2 r\left(\boldsymbol{gbest}_k-\boldsymbol{x}_k\right),\\&\boldsymbol{x}_{k+1}=\boldsymbol{x}_k+\boldsymbol{v}_{k+1},\end{aligned}\tag{4}$$

其中, k 是迭代次数, φ_0, φ_1, φ_2 表示群体认知系数, r 为按照均匀分布在[0, 1]区间取值的随机数. $\boldsymbol{v}$ 代表粒子速度向量, $\boldsymbol{x}$ 代表粒子位置向量(即问题的一种可能解).

2.3　水化动力学方程的反向萃取方法

从观测到的硅酸盐水泥水化程度时间序列数据反向重建出水泥水化早期的动力学方程, 即期望从水化程度时间序列数据中挖掘出形如下式的一阶常微分方程

$$\frac{\mathrm{d}\alpha}{\mathrm{d}t}=f_{C_0,C_1,\cdots,C_n}(t,\alpha),\tag{5}$$

其中, C_0, $C_1,\cdots$, C_n 是方程的系数, 对于不同品种的硅酸盐水泥, 在不同的养护条件下其系数不尽相同. Krstulovic 模型的系数包括了成核与增长速率数 K_{NG}、相边界反应速率数 K_{I} 和扩散速率数 K_{D} 等; Tomasawa 模型的系数则包含反应界面单位面积的反应速率系数 k_{r}、颗粒的初始半径 r_0 等. 本方法使用 GEP 算法来寻找 f 的函数形式. 在 GEP 优化的过程中, 针对某种形式的 f, 寻找使其适合于时间序列 i 的系数组合 C_{0i}, $C_{1i},\cdots$, C_{ni} 的过程选用粒子群优化算法. 而当函数形式 f 和参数组合 C_{0i}, $C_{1i},\cdots$, C_{ni} 给定的情况下, 求微分方程数值解的方法使用经典的四阶龙格库塔法, 三种算法之间互相嵌套.

为了更有效的寻找最优的函数形式f, 避免寻优过程陷入局部极值导致无法发现最优解. 把一个完整的GEP种群分为多个子种群, 每个子种群单独进化, 迭代到一定代数后(称做一个纪元), 计算所有子种群之间的最优染色体, 并用最优个体按照一定的种群迁移概率替换子种群里的最差个体, 然后算法继续迭代. 使用这种方法是基于古生物学里的断续平衡理论: 物种在一个稳定环境中达到平衡后, 其基因组成不再改变; 而当物种在与他们共同的祖先分开后, 新物种才能开始迅速的进化, 因此几个相互竞争的子种群比所有个体都聚集在一起的大种群在搜索效率上要更高[21]. 此外, 为了增强优化早期染色体之间的竞争性, 在每个子种群初始化时, 先随机生成一定数目的染色体, 并在其中挑选出适应值大于给定数值的染色体加入到初始群体中, 该过程不断迭代, 直到染色体数达到预先规定的群体规模.

为缩小 GEP 算法的搜索范围, 借鉴了 Krstulovic 模型和 Tomasawa 模型中的先验知识, 把其动力学方程分解为一个个独立的元素与函数, 添加入函数集 **F** 与元素集 **T**. 其中函数集 **F** 中包含 13 个函数, 元素集 **T** 中包含 9 个元素, 其中 C 代表系数项:

$$\mathbf{F}=\left\{+,-,\times,\div,-1\times(),\frac{1}{()},\sqrt{\ },\sqrt[3]{\ },()^2,()^3,()^{()},\mathrm{e}^{()},\ln()\right\},$$

$\mathbf{T}=\{C,\alpha,t,1-\alpha,\sqrt[3]{1-\alpha},\sqrt[3]{(1-\alpha)^2},\ln(1-\alpha),1-\sqrt[3]{1-\alpha},\alpha^3\}$, 此外, 由于随着时间的推移, 硅酸盐水泥的水化速率逐渐趋向于 0, 即 $\lim\limits_{t\to\infty}\dfrac{\mathrm{d}\alpha}{\mathrm{d}t}=0.$, 因此定义 $\dfrac{1}{(\ )(\ )\cdots(\ )}$ 为基因之间的连接函数.

GEP 算法和 PSO 算法都是根据个体适应值的不同来对种群进行不断的更新与优化, 因此需要定义合适的适应值函数在进化时作为指导来进行搜索. 首先定义水化动力学方程在相应时间点的数值解与目标时间序列的相对误差

$$\Delta=\left|\frac{\alpha-\hat{\alpha}}{\alpha}\right|, \tag{6}$$

其中, α是在某时刻的时间序列的真实水化程度值, $\hat{\alpha}$ 是根据待评估的动力学方程在给定系数组合 C_0, C_1,…, C_n的条件下计算而得的水化程度值. 因此根据相对误差Δ, 可定义适应值函数为

$$\text{fitness}=\frac{1}{1+\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{m}\Delta_{ij}}, \tag{7}$$

其中, N是时间序列的数目, m是时间序列所含的时间点数目, Δ_{ij}是时间序列 i在第 t_j时刻的相对误差, 适应值越大越好. 由于实验一共包括 8 个时间序列, 每个时间序列包含 1440 个样本点, 因此 N=8, m=1440. 此外, 第 t_0时刻的硅酸盐水泥的水化程度值定义为 0.

GEP 与 PSO 的参数设置对于能否发现最优解有着重要的影响. 由于随着种群规模的扩大和迭代次数的增多, 进化算法发现最优解的可能性逐渐变大. 因此根据这一特性, 首先设置了一个小型种群, 通过迭代有限的步骤来寻找最优参数. 通过反复试验, 在正式运行时采用如下的参数组合. GEP 种群规模设置为 800, 分为 8 个子种群, 算法的最大代数为 1000 代, 以每 10 代为一个纪元, 种群迁移率为 30%. 定义 GEP 基因数为 3, 单点交叉概率为 20%, 两点交叉概率为 50%, 基因迁移概率为 10%, 基因交叉概率为 10%, 变异概率为 5%, 插串率为 10%, 根插串率为 10%. 基因头部长度定义为 20, **F** 中函数的最大参数个数为 2, 因此基因尾部长度为 21, 所以染色体总长度为 123. 初始化种群时要求所有个体的适应值大于 0.0006. 对作为微分方程的系数搜索算法的 PSO 算法, 设置其种群规模为 48, 迭代次数为 100 代, φ_0 设置为 1, φ_1 设置为 1.8, φ_2 设置为 1.8. 对每组待优化参数, 反复执行 4 次 PSO 算法, 取其中最好的结果. 图 4 展示了硅酸盐水泥水化动力学方程反向重建的详细算法流程.

2.4 加速计算

水化动力学方程的反向重建方法需要 GEP, PSO, RK4 三种算法互相嵌套运算, 而且 RK4 算法迭代求微分方程数值解过程的时间复杂度很高, 这就导致了运算时间大大增加. 本方法中, GEP 的种群大小为 800, GEP 算法需要执行 1000 代, 对每个染色体需要检查其对 8 个时间序列的匹配情况, 对每个时间序列求系数时需要执行 4 次 PSO 算法, PSO 算法的种群大小为 48, 每次需要迭代 100 代, 因此总共需要求解 $800\times1000\times8\times4\times48\times100=1.2288\times10^{11}$ 次微分方程, 如果采用常规的串型计算机系统, 该任务耗时非常多.

为了解决这一问题, 使用浪潮倚天桌面超级计算机对其进行加速, 操作系统为 Linux, 编程环境为 C 语言. 该计算机包含两块 NVIDIA 公司生产的 Tesla C1060 高性能计算处理器(由于计算处理器源于传统的图形加速卡, 因此沿用了其简称 GPU), 每块处理器包含 240 个核, 每核频率为 1.3 GHz. 两块计算处理器的单精度浮点运算的峰值速度为 2Tflops. 采用 CUDA 并行计算架构(compute unified device architecture), 将 GPU 视为一个并行数据计算设备, 对所进行的计算进行分配和管理.

为了加速水化动力学方程反向工程的计算过程, 采用主从式同步并行方法, 在 CPU 端进行 GEP 与 PSO 主线算法流程的串行运算(主控端), 而对每一次迭代的适应值评估阶段在 GPU 上实现并行化(从属端). 也就是说, 要在 GPU 上完成当动力学方程的函数形式(GEP 染色体)与系数组合(PSO 粒子)给定的情况下, 使用四阶龙格库塔法计算水化程度并求出与目标时间序列的相对误差, 进而求得适应值的过程. 在每一次迭代时, 当 CPU 端串行完成了所有 PSO 种群的粒子更新工作后, 一次性请求 GPU 计算所有种群的粒子的适应值, 并在 GPU 计算完全部粒子适应值之前处于等待状态. 只有当 GPU 返回结果后, CPU 才进行下一次迭代. 通过这种方法, 每次迭代可以在 GPU 上求解 $800\times8\times4\times48=1.2288\times10^6$ 个微分方程. 每个微分方程的求解与计算相对误差的过程需要占用 GPU 上的一个线程, 不同线程之间进行相互独立的并行运算. 运算过程的时序关系如图 5 所示.

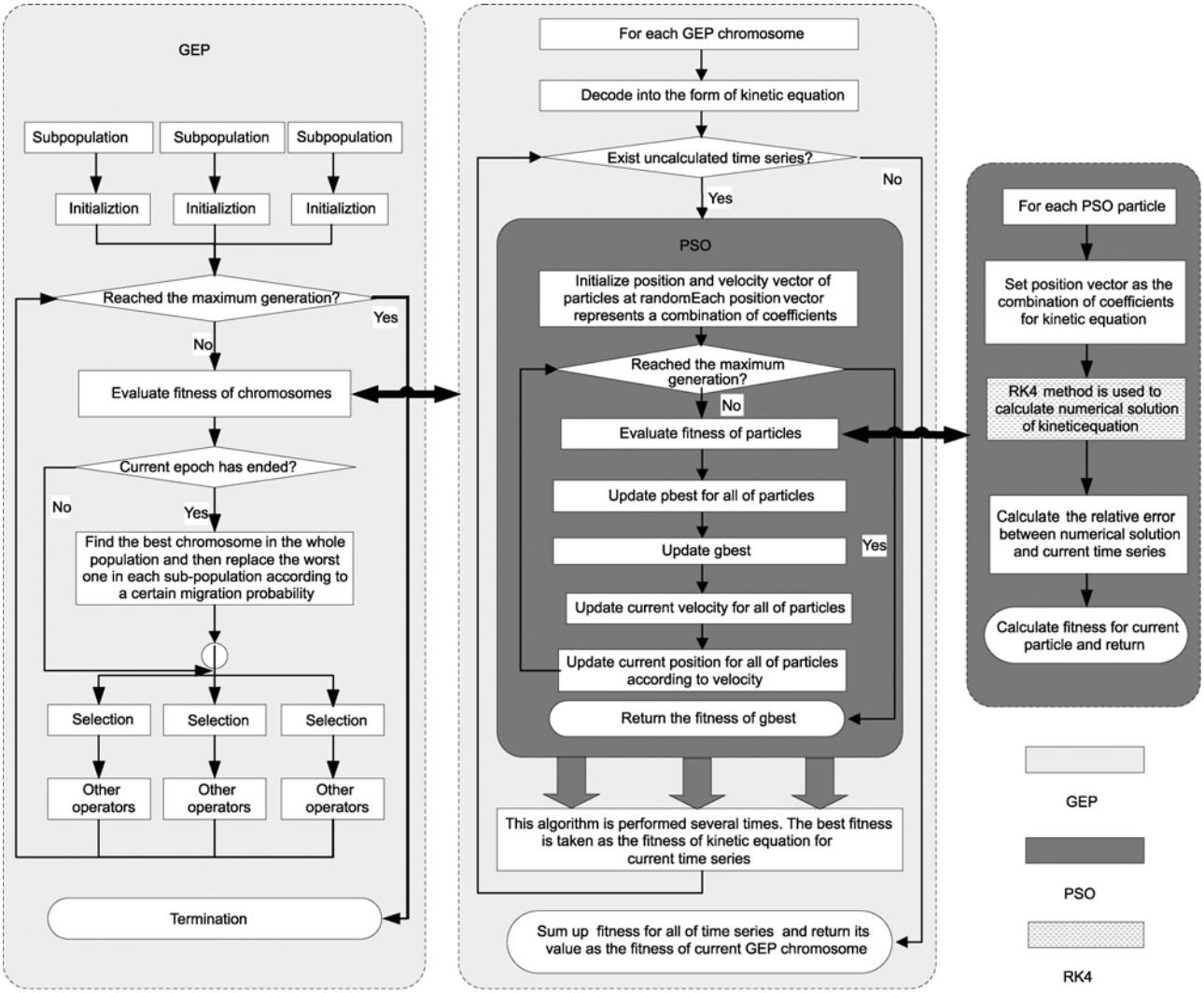

图 4　水化动力学方程反向重建的详细算法流程

此外, 由于存储器访问的时钟周期要远高于寄存器操作[22], 为进一步提高GPU上的线程执行速度, 需要减少线程对存储器的访问次数. 因此首先在CPU上对待评估的GEP染色体所代表的表达式树进行减支, 然后将其重新编码为新的K表达式再发送到GPU上进行运算. 减支过程是: 首先判断函数节点的下层节点是否都为元素节点, 如果不是, 搜索下一个函数节点; 如果是, 将函数节点与下层元素节点进行合并, 形成一个新的终端节点代替原函数节点. 该操作只从元素节点向上合并一层, 且不进行递归操作. 这种做法的好处是: 假设一个函数节点为二目运算符, 当线程使用未减枝的树时, 需要读取该函数节点和 2 个元素节点后才可以进行运算, 共需要读取 3 次内存, 而当线程使用减枝后的树时, 只需要从存储器中读取一次新的终端节点, 即可求出原函数节点的输出值.

3　结果与讨论

硅酸盐水泥早期水化动力学方程的自动反向重建共运行了 1190791.936947 s, 约合 331 h. 其中随机优选个体的初始化过程运行了 92493.968855 s, 约合 26 h; 方程进化过程运行了 1098297.968092 s, 约合 305 h. 为了评估利用 GPU 进行并行加速计算后的近似加速比, 把第 500 代与第 1000 代的 GEP 种群在传统的串型系统上(CPU 为 Intel Core i7 920)各重新运

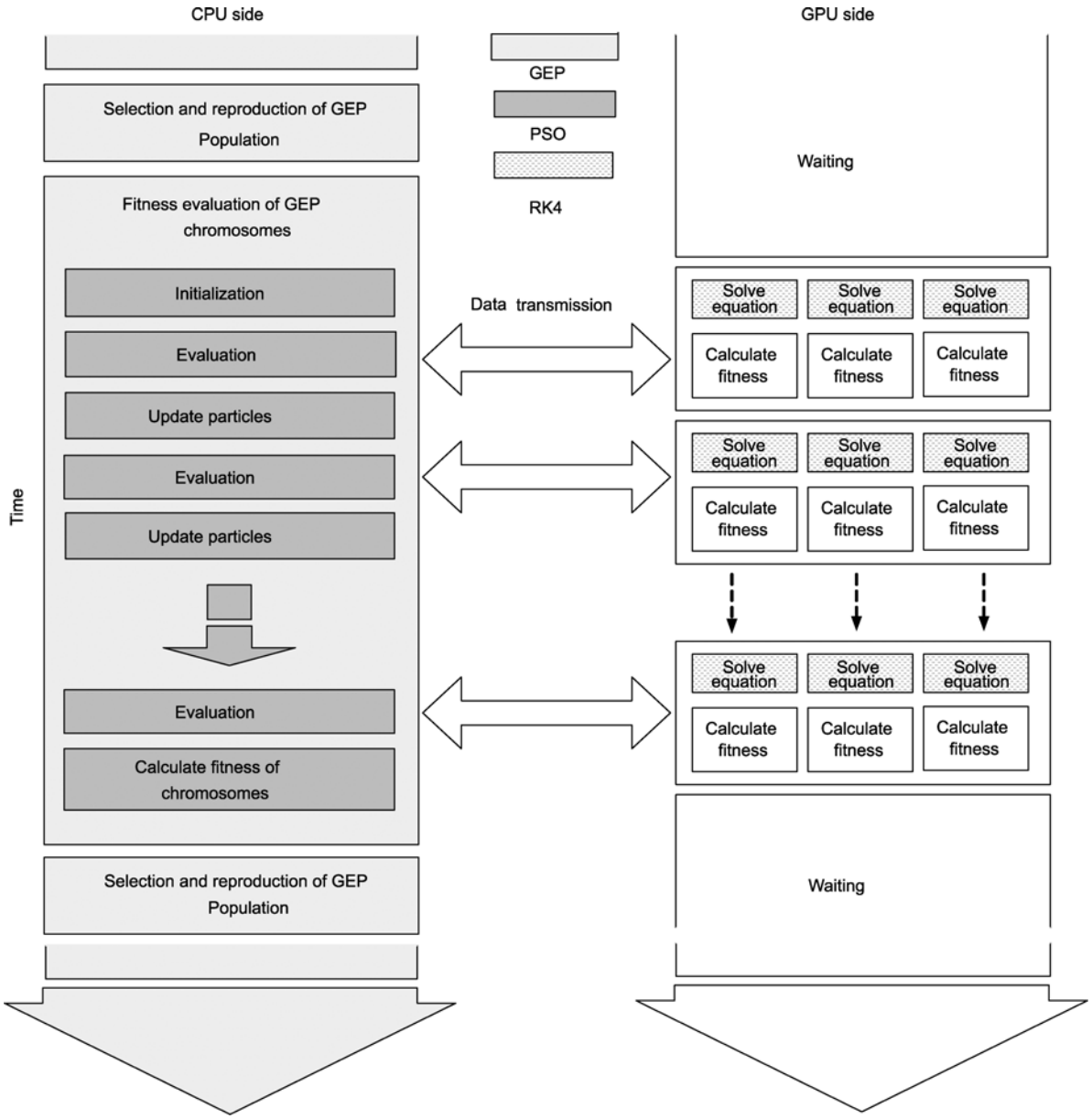

图 5　并行计算时序关系示意图

行一次. 表 3 列出了这两代各自的 GPU 与 CPU 运算时间的对比与加速比, 对其求平均可得在本方法中利用 GPU 的近似加速比约为 180 倍.

表 3　GPU 与 CPU 的运算时间对比

Generation	Execution time (s)		Speedup ratio
	GPU	CPU	
500	1125.743116	208772.264143	185.5
1000	1141.630435	200471.451376	175.6

当方程的进化过程执行完毕后, 对得到的历史最优染色体进行解码, 可得到最优方程为

$$\frac{\mathrm{d}\alpha}{\mathrm{d}t}=\frac{1}{\left(\left(C-\dfrac{1-\sqrt[3]{1-\alpha}}{\mathrm{e}^{\alpha^3+\alpha}}\right)^{Ct}\right)(C)\left(\left(\dfrac{C}{\left(\sqrt{\sqrt[3]{1-\sqrt[3]{1-\alpha}}-t}\right)^2}-C-t\right)^2\right)}, \tag{8}$$

其中, C 代表该位置为系数项, 对其进行进一步化简整理, 并把负号并入系数项, 可得如下动力学方程

$$\frac{\mathrm{d}\alpha}{\mathrm{d}t}=\frac{\left(C_1+\mathrm{e}^{-\alpha(1+\alpha^2)}\left(\sqrt[3]{1-\alpha}-1\right)\right)^{C_0 t}}{C_2\left(C_3+t+\dfrac{C_4}{\left(t-\sqrt[4]{1-\sqrt[3]{1-\alpha}}\right)^2}\right)^2}. \tag{9}$$

即为从数据中自动萃取出的硅酸盐水泥水化早期的动力学方程, 共包含有 5 个系数项 C_0, C_1,…, C_4, 时间 t 的单位是分钟(min), 该方程的适应值约为 0.01, 对 8 组作为教师信号的时间序列中每个时间点的平均相对误差为 0.81%. 动力学方程的进化过程如图 6 所示, 其中显示了历史最优染色体的适应值变化以及每代中整个种群适应值前 10%, 30%和 50%的染色体的平均适应值的变化情况. 可以观察到进化到 300 代以后时, 适应值的变化幅度逐步减小, 进化过程已趋向于收敛, 因此 GEP 算法迭代 1000 代已可以寻找到最优个体.

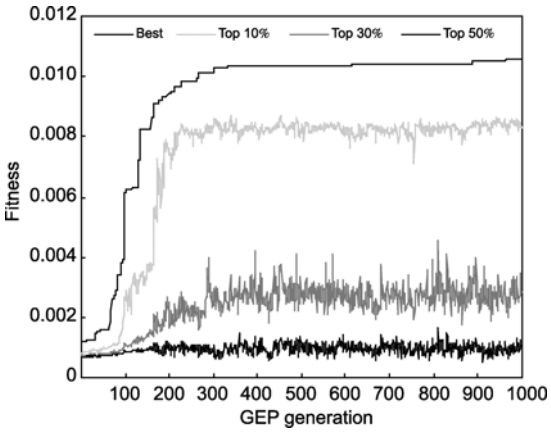

图 6　硅酸盐水泥早期水化动力学方程的进化过程

图 7 至图 16 列出了利用方程(9)得到的水化程度 α与水化速率 $\mathrm{d}\alpha/\mathrm{d}t$ 的模拟曲线同真实的水化实验数据之间的对比情况, 并且给出了平均相对误差及不同模拟曲线所对应的近似最优系数值. 其中序列 9 所使用的样品 D 的化学组成以及颗粒尺寸与其余时间序列都不相同, 而序列 8 尽管使用的样品与序列 6, 7 相同, 但是养护条件不同. 由于这两组时间序列并未用于方程进化的过程当中, 因此可以用来检验方程(9)在实际应用中的泛化能力. 为了得到方程(9)对每组数据的系数组合, 仍然利用 PSO 算法, 迭代次数增大到 1000 代以寻找最优系数, 适应值函数与其余参数不变. 图中列出了方程的系数以及水化程度曲线中所有点的平均相对误差, 可以观察到根据方程(9)生成的模拟曲线与实验数据之间具有很高的吻合度, 平均相对误差基本在 1%左右. 此外, 由该方程生成的曲线不但可以很好的模拟诱导期以后的各个时期, 也成功的模拟了硅酸盐水泥的快速反应期(即水化速率曲线里的第一个峰).

特别的, 序列 8 与序列 9 的匹配情况证实了即使化学组成、颗粒尺寸和养护条件发生改变, 上述动力学方程仍然可以很好的匹配硅酸盐水泥水化早期的实验数据. 因此, 对于不同的硅酸盐水泥样本, 存在一组最优的系数值, 使得方程(9)可以较好的模拟早期水化过程. 进一步的, 为了预测反应过程, 可以利用智能计算方法, 如人工神经网络等, 由样本的相关数据直接计算出动力学方程的系数组合, 如文献[6]. 由于人工神经网络具有逼近任意非线性函数的能力, 以样本的化学组成、粒径分布与养护条件等参量作为输入, 以动力学方程的系数作为输出, 当神经网络通过机器学习算法训练完毕后, 即可以用于完成从样

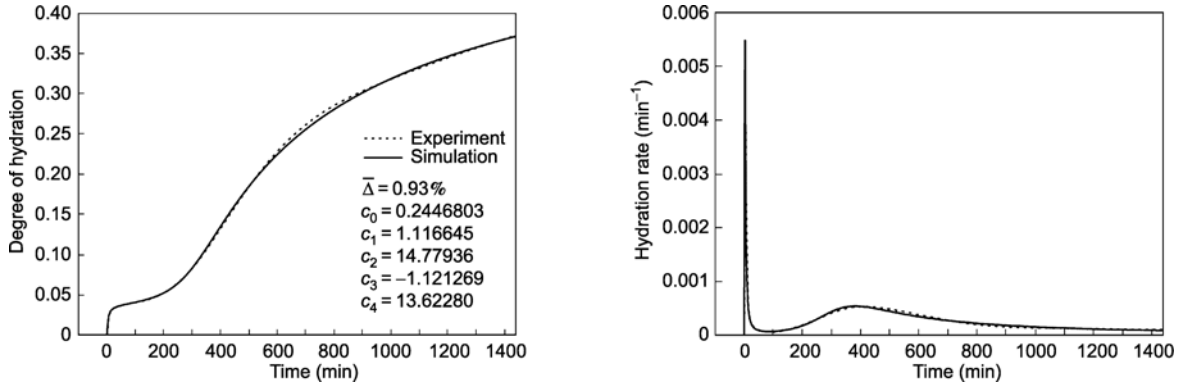

图 7　时间序列 0 的实验数据与动力学方程模拟曲线

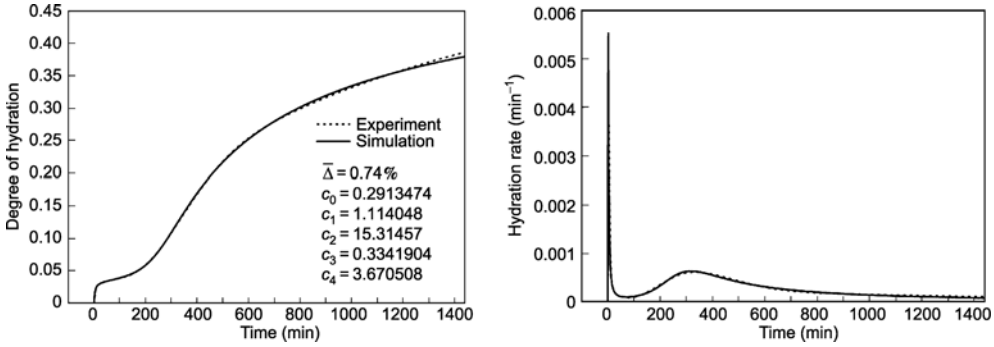

图 8　时间序列 1 的实验数据与动力学方程模拟曲线

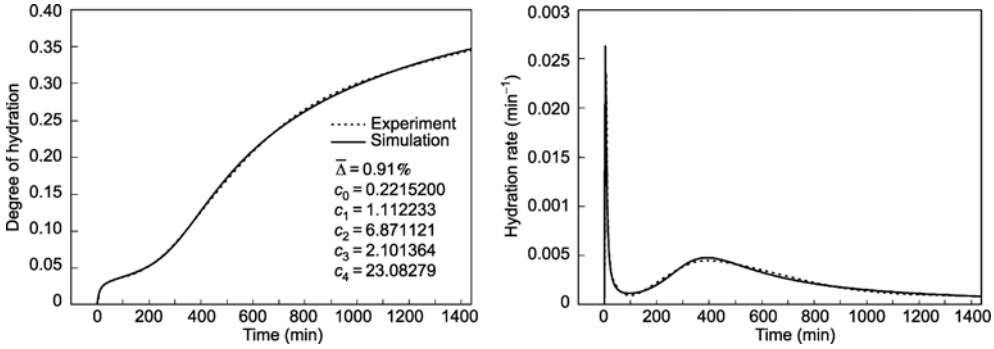

图 9　时间序列 2 的实验数据与动力学方程模拟曲线

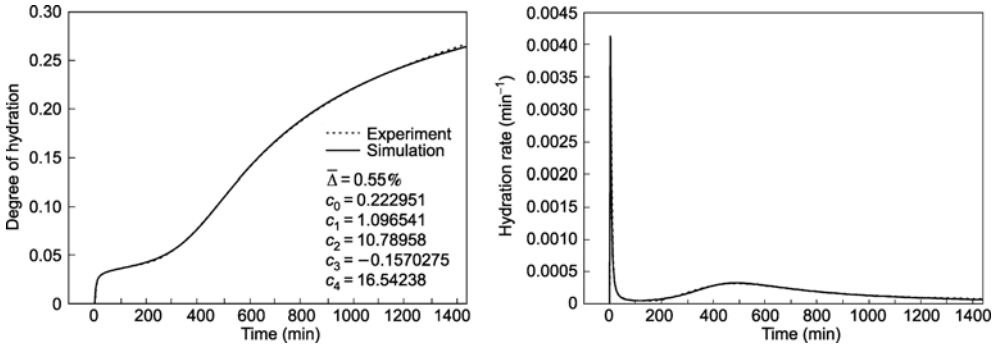

图 10　时间序列 3 的实验数据与动力学方程模拟曲线

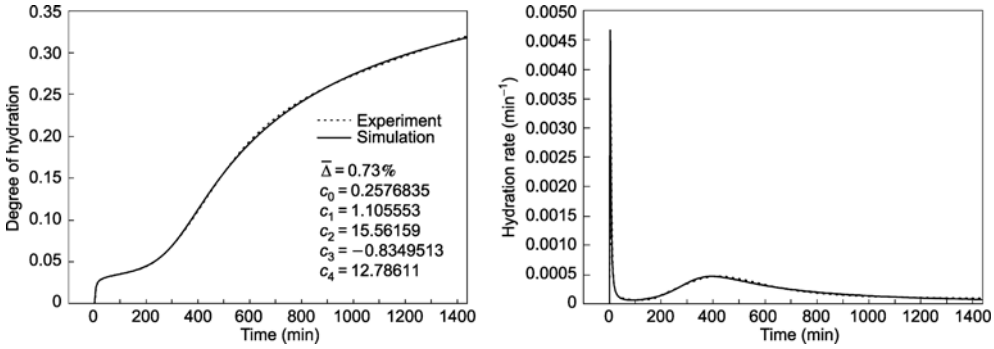

图 11　时间序列 4 的实验数据与动力学方程模拟曲线

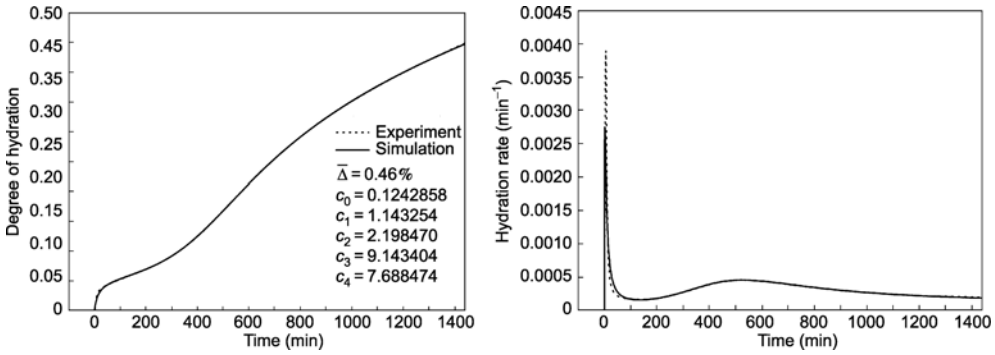

图 12　时间序列 5 的实验数据与动力学方程模拟曲线

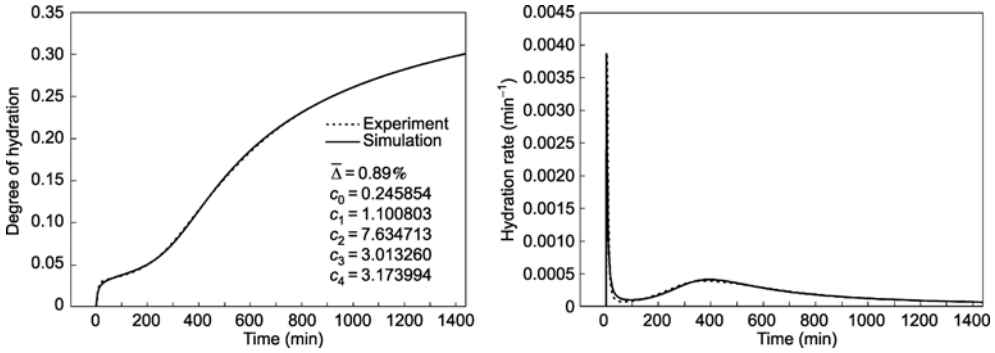

图 13　时间序列 6 的实验数据与动力学方程模拟曲线

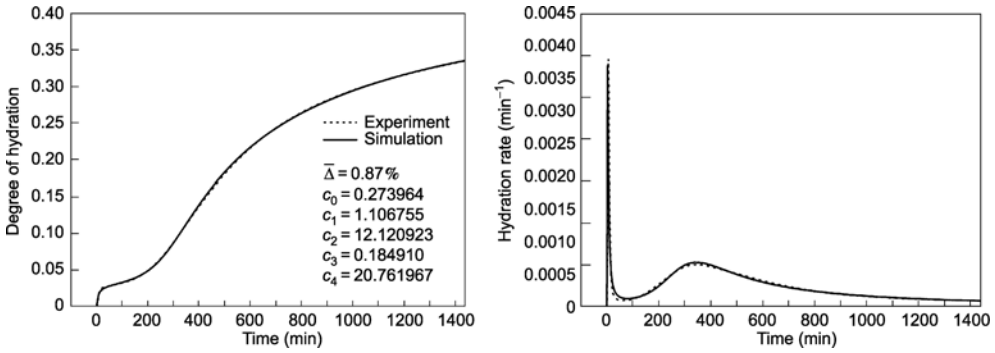

图 14　时间序列 7 的实验数据与动力学方程模拟曲线

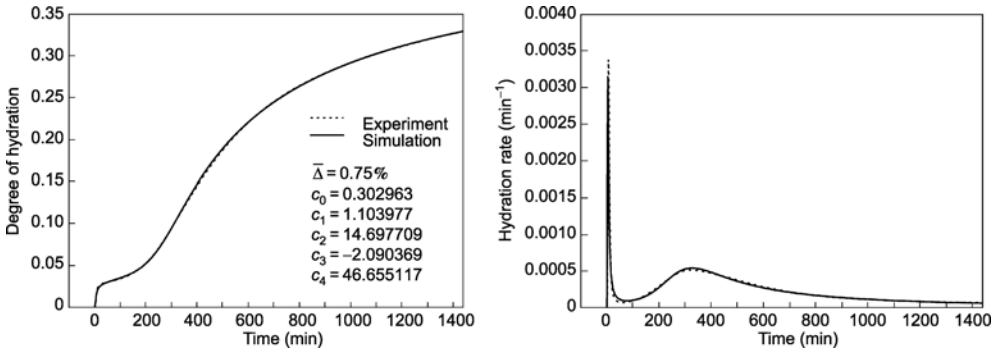

图 15　时间序列 8 的实验数据与动力学方程模拟曲线

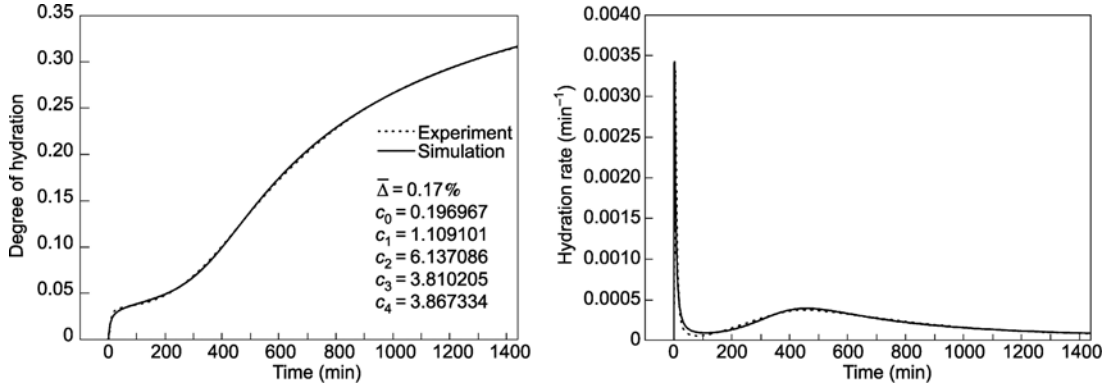

图 16　时间序列 9 的实验数据与动力学方程模拟曲线

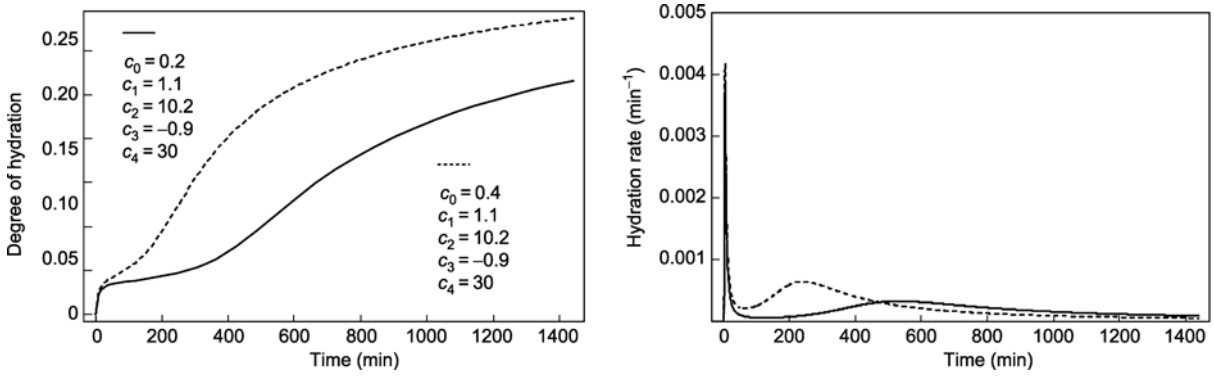

图 17　C_0 对水化过程的影响

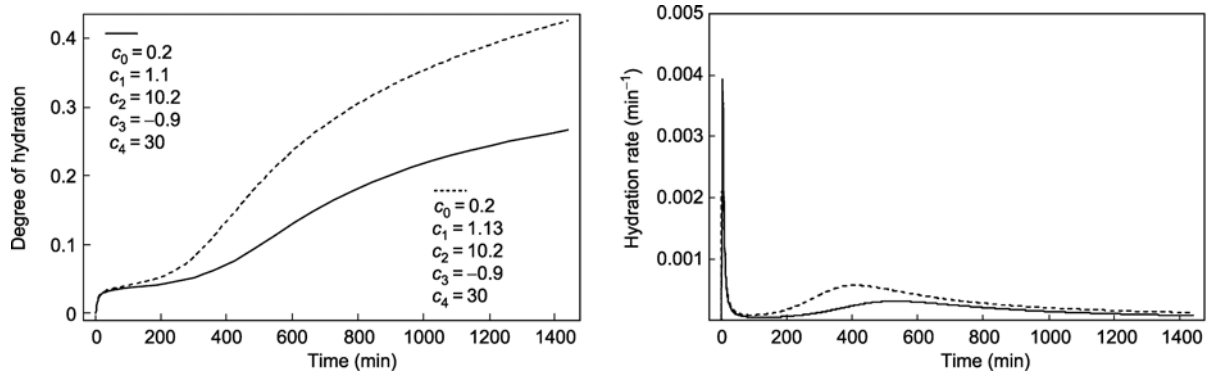

图 18　C_1 对水化过程的影响

本资料到动力学方程系数的相关映射.

虽然使用进化计算方法得到的经验方程具有难以直观理解的特点, 并导致无法得到 5 个系数的具体含义, 但是可以评估各个系数对水化过程的影响. 图 17 至图 21 列出了 5 个系数的变化分别导致由方程得出的水化程度曲线与水化速率曲线的变化情况. 由图中可以看到, 系数 C_0 的主要作用是调节诱导期的长短以及诱导期以后的水化程度, 随着 C_0 的增大, 诱导期和加速期的时间缩短而衰退期和稳定期的时间变长. 系数 C_1 主要调节诱导期以后的水化速率, C_1 较大则诱导期以后的水化速率较快. C_2 的变化可以对水化程度的发展进行整体性的均匀调节, 增大 C_2 可以降低快速反应期、诱导期与加速期的反应速率. 系数 C_3 和 C_4 的变化可以引起快速反应期水化速率的改

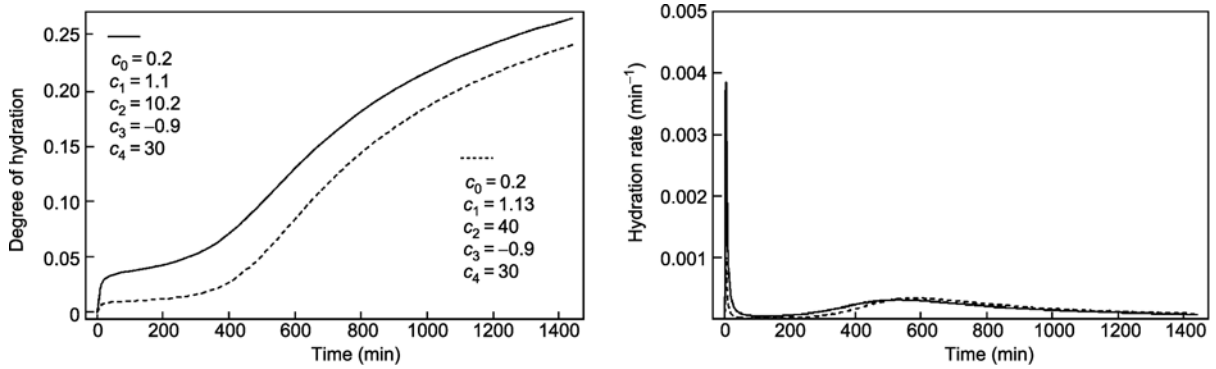

图 19　C_2对水化过程的影响

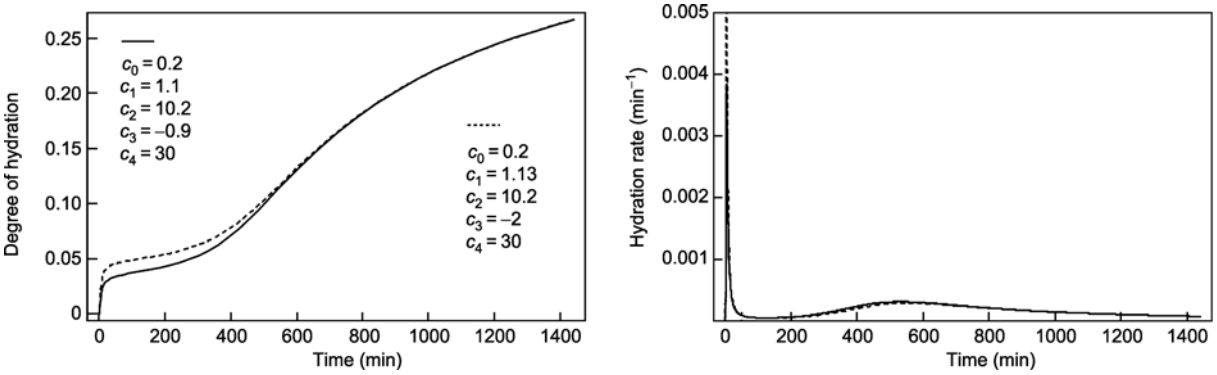

图 20　C_3对水化过程的影响

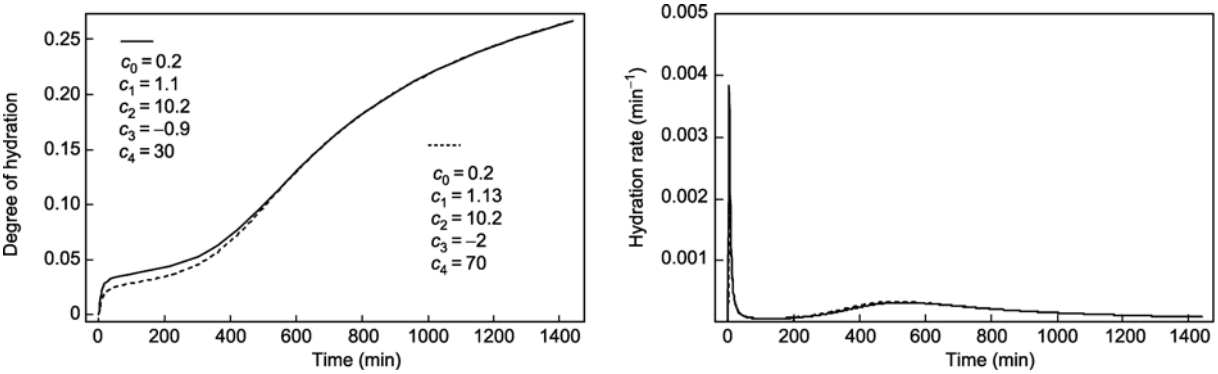

图 21　C_4对水化过程的影响

变, 但是对后续的各个时期影响不大.

4　结论与展望

利用进化计算方法由观测数据自动化的反向萃取硅酸盐水泥水化早期的动力学方程. 采用基因表达式编程算法来搜索动力学方程的函数形式, 采用粒子群算法寻找系数组合, 并利用 GPU 加速计算过程. 通过挖掘多组水化程度时间序列数据, 得到了硅酸盐水泥水化早期的动力学方程. 实践证明, 该方法得到的动力学方程不但可以很好的吻合训练数据, 而且具有良好的泛化能力.

在未来的工作中, 将利用神经网络与模糊系统由样本的化学组成、养护条件等参数计算出水化早期

动力学方程的系数并预测反应过程. 此外, 由于由 Bogue 公式计算而得的化学组成与熟料矿物的实际组成是有差别的, 计划在今后的研究中, 采用 X 射线衍射定量相分析法对矿物组成直接定量.

致谢 作者对山东省水泥质量监督检验站的王华老师, 济南大学材料科学与工程学院的屈雅老师和潭文杰以及济南大学信息科学与工程学院的齐鲁、林金、侯汶杉、李萍萍和姜辉在研究中的帮助表示感谢.

参考文献

1 袁润章. 胶凝材料学. 武汉: 武汉工业大学出版社, 1996
2 Chen W, Brouwers H J H. Mitigating the effects of system resolution on computer simulation of Portland cement hydration. Cem Concr Comp, 2008, 30: 779—787
3 Tomosawa F. Kinetic hydration model of cement. In: Proc Cem Concr, 1974. 53—57
4 Krstulovic R, Dabic P. A conceptual model of the cement hydration process. Cem Concr Res, 2000, 30: 693—698
5 Pignat C, Navi P, Scrivener K. Simulation of cement paste microstructure hydration, pore space characterization and permeability determination. Mater Struct, 2005, 38: 459—466
6 Parka K B, Noguchib T, Plawsky J. Modeling of hydration reactions using neural networks to predict the average properties of cement paste. Cem Concr Res, 2005, 35: 1676—1684
7 阎培渝, 郑峰. 水泥基材料的水化动力学模型. 硅酸盐学报, 2006, 34: 555—559
8 Koza J R. Genetic Programming: On the Programming of Computers by Means of Natural Selection. Cambridge, MA: MIT Press, 1992
9 Holland J H. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. 2nd ed. Michigan: MIT Press, 1992
10 Schmidt M, Lipson H. Distilling free-form natural laws from experimental data. Science, 2009, 324: 81—85
11 Koza J R, Mydlowec W, Lanza G, et al. Automated reverse engineering of metabolic pathways from observed data by means of genetic programming. In: Kitano H, eds. Foundations of Systems Biology. Michigan: MIT Press, 2001
12 Kikuchi S, Tominaga D, Arita M, et al. Dynamic modeling of genetic networks using genetic algorithm and S-system. Bioinf, 2003, 19: 643—650
13 Qian L J, Wang H X, Dougherty E R. Inference of noisy nonlinear differential equation models for gene regulatory networks using genetic programming and Kalman filtering. IEEE Trans Sig Proc, 2008, 56: 3327—3339
14 Chakraborty U K. Static and dynamic modeling of solid oxide fuel cell using genetic programming. Energy, 2009, 34: 740—751
15 Iba H. Inference of differential equation models by genetic programming. Inf Sci, 2008, 178: 4453—4468
16 Ferreira C. Gene expression programming: A new adaptive algorithm for solving problems. Com Syst, 2001, 13: 87—129
17 Kennedy J, Eberhart R C. A new optimizer using particle swarm theory. In: Proc 6th Int Symp Micro Mach Hum Sci, 1995. 39—43
18 王培铭, 丰曙霞, 刘贤萍. 水泥水化程度研究方法及其进展. 建筑材料学报, 2005, 8: 646—652
19 Schindler A K, Folliard K J. Heat of hydration models for cementitious materials. ACI Mater J, 2005, 102: 24—33
20 Bogue R H. The Chemistry of Portland Cement. New York: Reinhold Publishing Corporation, 1955. 245—268
21 陈伦军. 机械优化设计遗传算法. 北京: 机械工业出版社, 2005
22 CUDA programming guide version 2.3.1. NVIDIA, 2009