基于混沌遗传算法的网格工作流调度应用*

苏 翔，刘洋君，杨 健，闫园园
（江苏科技大学 经济管理学院，江苏 镇江 212003）

摘 要：动态网格环境中，多 QoS（服务质量）约束下的工作流调度问题是决定其任务执行成功与否及效率高低的关键。现有的网格工作流调度算法难以满足实际应用中的不同需求，同时算法欠优化，难以提供多种策略，由此提出了一种基于期限与预算两个 QoS 约束的改进型混沌遗传算法。首先，为避免算法出现收敛停滞将混沌机制引入遗传算法并对变异概率进行自适应处理。其次，提出时间和预算的线性结合概念，将目标函数转换为适应值函数。最终基于工作流调度中的平衡结构和非平衡结构测试了算法的有效性。

关键词：网格工作流；工作流调度；截止期限约束；预算约束；混沌遗传算法
中图分类号：TP393 **文献标志码**：A **文章编号**：1001-3695(2013)09-2645-03
doi:10.3969/j.issn.1001-3695.2013.09.022

Application of grid workflow scheduling based on chaos-genetic algorithm

SU Xiang, LIU Yang-jun, YANG Jian, YAN Yuan-yuan
(*School of Economics & Management*, *Jiangsu University of Science & Technology*, *Zhenjiang Jiangsu* 212003, *China*)

Abstract: In dynamic grid environment, workflow scheduling problem with multi-QoS (quality of service) constraints is the key to decide the task execution success and efficiency. The existing grid workflow scheduling algorithm was difficult to meet the different needs of practical application, and the algorithm was lack of optimization, it was difficult to provide a variety of strategy, so this paper introduced an improved chaos-genetic algorithm based on the deadline and budget QoS constraints. In order to avoid algorithm appear convergence stagnation, it used the chaos mechanism to combine with genetic algorithms and dynamically adjusted the probability of mutation. Then it put forward the concept of linear combination of time and budget, so that the objective function was transformed into a fitness function. Finally the results of applying chaos-genetic scheduling algorithm show greater performances on both balanced and unbalanced workflows.

Key words: grid workfolw; workflow scheduling; deadline constraints; budget constraints; chaos-genetic algorithms

随着信息时代的到来，网络和计算机得到广泛的应用，高速的通信资源由于其廉价性，使得地理位置各异的资源被利用来处理一些复杂的问题，如科学计算、生产制造等问题。对这些问题的深入研究导致了网格环境下如何高效利用资源问题的出现。而网格环境下，许多相互依赖的任务形成了不同的工作流，如何安排这些工作流成为最大的挑战。

网格工作流调度问题不同于一般的任务调度，在调度时不仅要考虑为任务选择一个最佳资源，还要考虑各个任务之间的时序、因果及用户的 QoS 需求等一系列的约束条件[1]，以及协调各个任务的执行来达到最终的目标。该问题由于其复杂性与冲突性，成为国内外学者研究的热点问题。如李玺等人[2]提出基于截止时间满意度的网格工作流调度算法，苑迎春等人[3]提出基于串归约的网格工作流费用优化方法，单冬红等人[4]提出基于遗传算法负载均衡的网格工作流技术研究，它们都能较快地完成工作流的调度，但这些算法只能在一个时间 QoS 约束下进行，具有很强的局限性。张伟哲等人[5]使用多目标演化算法进行求解，但该算法较为复杂且没有必要给出 Pareto 的最优解集。墨尔本大学的 Yu 等人[6]提出基于时间和预算约束下的遗传算法解决工作流调度问题，对本文具有较大借鉴意义，但该算法由于遗传算法自身的局限性，效率不高。

纵观以上学者的研究，在多 QoS 约束下，如何高效完成工作流调度任务依然是一个研究难点。实际生产调度中，各个管理实体对网格 QoS 的要求各异，有时甚至相互冲突，考虑过多的 QoS 需求不仅增加生产负担，而且达不到预期目标。预算及截止期限是用户评价工作流调度 QoS 最重要的两个指标。鉴于此，本文引入了一种高效性和鲁棒性强的全局寻优算法——遗传算法来解决基于截止期限及预算的 QoS 约束工作流调度问题。但传统遗传算法（traditional genetic algorithm, TGA）本身具有局限性，容易产生收敛停滞和陷入局部极值点的问题[7,8]，为此本文引入一种高效性的自适应算法——混沌遗传算法（chaos-genetic algorithm, CGA）。它是在遗传算法的基础上引入了混沌系统，并根据遗传算法特点动态调整其变异算子，以提高算法的准确性及效率。

1 工作流调度问题描述

从资源调度的角度出发，网格结构可以利用工作流被抽象成一个比较简单的结构模型。这种抽象是为了从繁琐的网格结构中得到一种更有利于调度研究的模型。一个工作流的应用程序可以利用有向无环图（directed acyclic graph, DAG）建

收稿日期：2012-12-01；**修回日期**：2013-01-25 **基金项目**：江苏省自然科学基金资助项目（BK2009728）

作者简介：苏翔（1965-），男，江苏镇江人，教授，硕导，博士，主要研究方向为管理信息系统、企业信息化；刘洋君（1987-），男，福建福州人，硕士研究生，主要研究方向为信息管理与信息系统（becauseboy@gmail.com）；杨健（1967-），女，江苏镇江人，主要研究方向为计算机应用；闫园园（1987-），女，河南焦作人，硕士研究生，主要研究方向为信息管理与信息系统.

模。已知条件如下：

a)定义一组有限的任务集 $T_i(i=1,2,\cdots,n)$ 和一组定向连接 (T_i,T_j) 的弧的表单，其中任务 T_j 是 T_i 的子任务，子任务在所有其父任务完成后才能开始执行。在用户指定的流程执行中，定义 B 为预算约束，D 为截止期限约束。

b)定义 m 为可用服务器的总数和一组能够执行任务 T_i 的服务集 $s_i^j(i=1,2,\cdots,n;j=1,2,\cdots,m_i,m_i<m)$，每个任务只能为其指定一个可行的服务器来执行，且每个服务器执行任务的处理时间及费用不同。

c)定义服务 s_i^j 执行任务 T_i 的处理时间为 T_i^j，服务费用为 C_i^j，工作流调度问题即是将每一个任务 T_i 映射到合适的服务 s_i^j 上，并根据用户提出的截止期限及预算优化执行时间及成本。

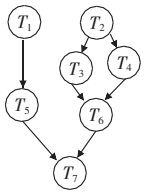
2 混沌遗传算法设计

传统遗传算法是基于生物进化原理的普适性全局优化算法，它依据适应度函数，对种群重复施加遗传操作，最终使得初始种群中的个体逐代得以优化并逼近最优解。本文针对遗传算法易陷入早熟的缺点，将混沌系统嵌入到遗传算法中。与原有的随机搜索相比，混沌系统的遍历性能够在全局范围内进行无重复搜索，减少搜索的盲目性及随机性[9,10]，从而提高搜索的效率。以下是改进的混沌遗传算法的具体实现过程。

2.1 编码

对于一个工作流调度问题，用传统遗传算法基于工序的编码方法无法得到问题的解。网格工作流调度问题不仅需要确定任务的执行顺序，还需要为每一个任务选择一个合适的服务器。

图 1 中的染色体表示为工作流调度的一个可行解，它由两部分组成，第一部分为任务的编码，第二部分为服务分配的编码，它的长度等于任务总数。一个网格工作流调度方案可以看做是一个服务的排列[11]，服务排列中每一个任务的分配包括四个元素（任务编码、服务编码、开始时间、结束时间），前两个元素识别任务被分配到哪个服务器上。由于加入时间参数后算法将会变得异常复杂，于是在本文中忽略时间的因素。



$S_1 \rightarrow T_1\text{-}T_3$
$S_2 \rightarrow T_2\text{-}T_4\text{-}T_7$
$S_3 \rightarrow T_5\text{-}T_6$
$S_4 \rightarrow T_1\text{-}T_3\text{-}T_7$
$S_5 \rightarrow T_2\text{-}T_3\text{-}T_7$
$S_6 \rightarrow T_4\text{-}T_6$

任务编码	T_1	T_2	T_3	T_4	T_5	T_6	T_7
服务器分配编码	1	2	1	2	3	3	2

图 1 网格工作流调度问题的染色体

2.2 种群初始化及适应度函数

对于初始种群，本文利用混沌系统来取代原有的随机生成方式。将待优化变量利用混沌机制映射到混沌空间，成为混沌变量。进化过程的混沌变量可以利用下面的方程式进行定义：

$$CS_i^{(k+1)}=4CS_i^{(k)}(1-CS_i^{(k)}) \quad i=1,2,\cdots,m \tag{1}$$

其中：CS_i 表示第 i 个混沌变量，k 和 $k+1$ 表示迭代的次数，CS_i 的取值区间是(0,1)。混沌优化机制的工作步骤如下：

a)将(0,1)区间划分为 m 个等分区间（m 表示能够执行指定任务的服务器总数）。

b)每一个初始随机种群中的基因值将会被映射到取值范围在区间(0,1)上的新的值 CS_i 上。

c)$S_i^{(1)}$ 的值是通过线性映射操作机制确定：

$$\frac{s_i^{(1)}}{m_i}=CS_i^{(1)} \tag{2}$$

其中：$S_i^{(1)}$ 的值为当代给任务 T_i 分配的服务器，m_i 表示能够执行任务 T_i 的资源总数。

d)下一次迭代的混沌变量 $CS_i^{(2)}$，它是通过应用式(1)以及先前一步生成的 $CS_i^{(1)}$ 的值生成的。

e)混沌变量 $CS_i^{(2)}$ 是用来生成 $s_i^{(2)}$，它利用方程

$$s_i^{(2)}=[CS_i^{(2)}\times m_i] \tag{3}$$

由此，通过式(1)～(3)中定义的混沌机制继续生成每一个染色体中的 $s_i^{(k)}$ 的值。在这个阶段，生成的每个染色体的适应度值也能够被估算出。遗传算法在搜索进化过程中，一般是以适应度函数作为依据，利用种群中每一个个体的适应度函数值进行搜索进化，因此适应度函数的建立尤为重要。本文的适应度函数是基于成本预算和截止期限两个 QoS 约束的，使得用户在决策时更具有选择性。

对于成本约束调度，成本—适应度函数是用来约束结果的成本，于是定义对于个体 I 的成本—适应值函数为

$$F_{\text{cost}}(I)=\frac{C(I)}{B} \tag{4}$$

其中：$C(I)$ 表示执行个体 I 所需实际成本总和，B 表示该工作流的预算成本。

对于成本约束调度，时间—适应度函数是为了约束个体执行的时间值满足截止期限的约束。于是定义对于个体 I 的时间—适应值函数为

$$F_{\text{time}}(I)=\frac{t(I)}{D} \tag{5}$$

其中：$t(I)$ 表示执行个体 I 所需时间的总和，D 表示该工作流的截止期限。最终的适应值函数是结合式(4)(5)的，它表示为

$$F(I)=\begin{cases} F_{\text{cost}}(I)+F_{\text{time}}(I) & F_{\text{cost}}(I)>1 \\ & \text{或 } F_{\text{time}}(I)>1 \\ \dfrac{C(I)}{\text{maxcost}}\times\dfrac{t(I)}{\text{maxtime}} & \text{其他} \end{cases} \tag{6}$$

maxcost、maxtime 分别表示当前种群个体的最大执行成本及最大完工时间。

2.3 遗传操作

遗传操作是对选择的个体进行交叉和变异操作，产生继承优良特性的下一代，经过不断的迭代，最终会收敛到最适应环境的个体。

2.3.1 选择算子

选择也就是从当前种群中选择适应值高的染色体以生成交配池的过程。本文采用轮盘赌的方法来实现，该方法使得经过混沌优化的个体中适应度值较高的个体具有较高概率被选中，使得上一代个体的优良特性得以被遗传。

2.3.2 交叉算子

对轮盘赌选出的个体采用一定的交叉概率交换其部分基因，产生新的基因组合，它使得个体有希望交流其优秀的基因，并产生比父代更优秀的解的结构。本文中采用两点交叉的方法，对选定的父个体随机选择两个交叉点，将位于交叉点之间的所有任务对应的服务器进行交换。交叉操作如图 2 所示。

2.3.3 改进的变异算子

变异操作一定程度上能避免种群有效基因的缺失，从而增加种群的多样性。本文针对遗传算法局部搜索能力差的特点，

对变异概率进行自适应处理[12]:在进化的前期为保持种群多样性,变异概率较大;进化后期为避免破坏已存在的优秀群体减小变异概率,从而提高算法局部微调功能。变异操作如图 3 所示。定义变异概率 P_m 为

$$P_m(\tau)=1-a\times r^{[1-(\frac{\tau}{k})]^b} \tag{7}$$

其中:k 为最大进化代数,τ 表示当前进化代数,b 取值为 2,$r\in[0,1]$。从式(7)可看出,对于较小的 r 值,在进化初期,变异概率 $P_m\approx1$,此时变异空间较大;而在进化后期 $P_m\approx0$,种群在小范围空间内进行局部搜索,将选中的个体进行变异操作。

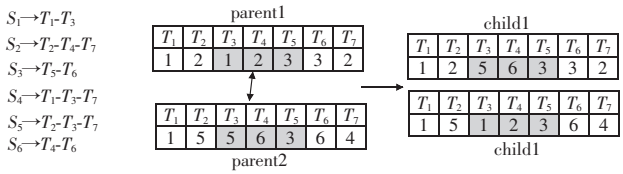


图 2　交叉操作

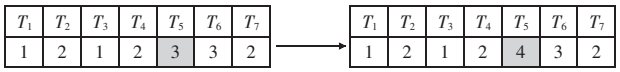


图 3　变异操作

3　实验与分析

根据工作计划,工作流调度问题可以分为平衡结构(图 4)和非平衡结构(图 5)两类。为证明算法的有效性,本文将分别在两种结构下与 TGA 进行比较。

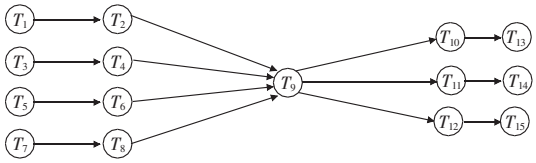


图 4　平衡工作流结构

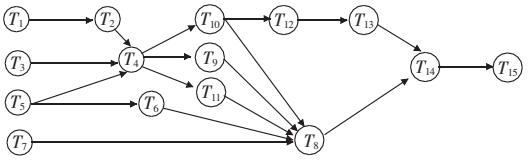


图 5　非平衡工作流结构

本文使用了 Yu 的数据集[6]进行测试,主要参数设置为:种群规模为 10,迭代次数为 100,交叉概率为 0.5,变异概率见 2.3.3 节,数据集连续运行 5 次取最优结果。

图 6、7 给出了两种结构下算法运行的实验结果,其中分别设定 $D=220(H)$ 和 $D=240(H)$,垂直轴表示的是总成本与用户预算的比值。分析结果表明,当预算较低时(如 $B=3000$ 时),实际执行成本与用户预算的比值大于 1,假设失败。但随着用户预算的增加(如 $B=5000$ 时),两种算法的结果均满足用户要求,而本文引入的混沌遗传算法(CGA)在每一级预算下所得的结果都优于 TGA。

图 8、9 中,设定 $B=5000$,由于在平衡结构下完成工作流任务的时间较短,所以在非平衡结构下设置截止时间的范围从 $D=180(H)$ 到 $D=300(H)$,而在平衡结构下设置 $D=160(H)$ 到 $D=280(H)$ 之间。实验结果表明 TGA 在每一级截止期限下完成任务的时间均多于 CGA。

以上所有的假设中,在一些状态下传统遗传算法的表现与 CGA 差距不大(如图 5 中,将预算约束设置为 7000 时),是由于 TGA 没有陷入局部最优解的困境中,而在剩余的假设状态中尤其是在复杂的非平衡结构下,CGA 的表现均明显优于 TGA。

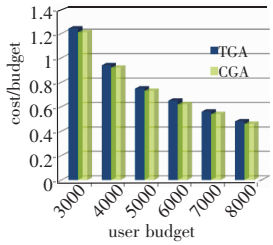


图 6　平衡结构执行成本对比

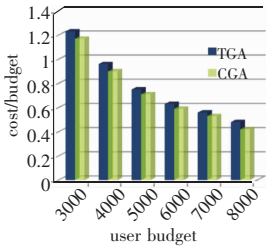


图 7　非平衡结构执行成本对比

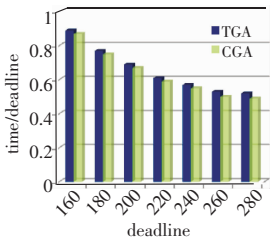


图 8　平衡结构执行时间对比

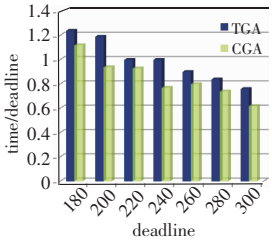


图 9　非平衡结构执行时间对比

4　结束语

针对网格工作流调度的特点,本文介绍了一种基于成本及截止期限约束的混沌遗传算法,它在满足用户 QoS 需求的同时将混沌机制引入遗传算法中,并改进了变异算子,使得最终能快速搜索到全局最优解。实验证明这种新算法解决了网格工作流调度效率低的问题,并有效地克服了遗传算法的早熟收敛缺陷,取得了较为满意的结果,具有很高的实用价值。

参考文献:

[1] 李金忠,夏洁武,曾劲涛,等. 网格工作流调度算法研究综述[J]. 计算机应用研究,2009,26(8):2816-2820.

[2] 李玺,胡志刚,胡周君,等. 基于截止时间满意度的网格工作流调度算法[J]. 计算机研究与发展,2011,48(5):877-884.

[3] 苑迎春,李小平,王茜,等. 基于串归约的网格工作流费用优化方法[J]. 计算机研究与发展,2008,45(2):246-253.

[4] 单冬红,杨照峰. 基于遗传算法负载均衡的网格工作流技术研究[J]. 计算机数字与工程,2011(10):81-84.

[5] 张伟哲,胡铭曾,张宏莉,等. 多 QoS 约束网格作业调度问题的多目标演化算法[J]. 计算机研究与发展,2006,43(11):1855-1862.

[6] YU Jia, BUYYA R. Scheduling scientific workflow applications with deadline and budget constraints using genetic algorithms[J]. Scientific Programming,2006,14(3-4):217-230.

[7] 苏翔,窦培华. 一种基于带熵的遗传算法在车间调度中的应用[C]//第十届中国管理科学学术年会. 2008:142-146.

[8] 缪朝炜,杨凤,徐东升,等. 越库转运问题的自适应遗传算法研究[J]. 管理科学学报, 2011,14(6):24-34.

[9] 谭跃,谭冠,叶勇,等. 具有混沌局部搜索策略的双种群遗传算法[J]. 计算机应用研究,2011,28(2):469-471.

[10] 王建华,李南,罗建强,等. 价格敏感型供应链网络规划模型及其混合遗传算法[J]. 管理工程学报,2011,25(3):167-171.

[11] 程相,孙继银,李琳琳,等. 军事网格工作流调度算法研究[J]. 计算机工程,2012,38(10):45-47.

[12] 胡彬,杨景曙,王粒宾. 改进的混沌遗传算法及其应用[J]. 计算机工程,2010,36(5):170-172.