

考虑运输代价的制造 workflow 调度算法

李 聪, 汪 芸, 罗香玉

(东南大学 计算机科学与工程学院 计算机网络和信息集成教育部重点实验室, 江苏 南京 211189)

摘 要: 在云制造环境下, 因制造服务资源所在地域的差异性, 多目标制造 workflow 调度不仅考虑制造服务所需时间、费用, 还需考虑产品运输所需时间、费用, 原有 workflow 调度算法无法有效优化运输代价. 针对此问题, 结合遗传算法全局搜索能力强与粒子群算法收敛速度快的特点, 提出多目标混合遗传粒子群(MOGA-PSO)算法. 仿真结果表明混合算法能够有效降低运输代价, 使得 workflow 调度得到进一步优化, 可适用于云制造环境.

关键词: 制造 workflow; 优化调度; 运输代价; MOGA-PSO 算法

中图分类号: TP393

文献标识码: A

文章编号: 1000-7180(2013)03-0067-04

Manufacturing Workflow Scheduling Algorithm Considering Transportation Cost

LI Cong, WANG Yun, LUO Xiang-yu

(School of Computer Science and Engineering, Southeast University, Key Laboratory of
Computer Network and Information Integration, Ministry of Education, Nanjing 211189, China)

Abstract: In cloud manufacturing environment, for the geographical distribution of resources, the scheduling for manufacturing workflow with multiple goals should consider not only the cost of resources, but also the cost of product transportation. A hybrid algorithm is proposed to solve the problem which took the advantage of global optimization with genetic algorithm and fast convergence with particle swarm algorithm. Compared to the results of normal GA and PSO, simulation results show that the hybrid algorithm is an effective method for manufacturing workflow scheduling.

Key words: manufacturing workflow; optimization scheduling; transportation cost; MOGA-PSO algorithm

1 引言

中国是制造大国, 如何解决多制造企业间的快速互联协作一直是制造业和计算机网络研究中的重要课题. 云制造概念^[1]的提出为该问题提供了良好的解决方案. 为了利用云制造服务模式的资源优势, 大力节约投资成本, 可以将制造任务放在云制造环境下执行.

大型制造任务通常可看作采用有向无环图 DAG(directed acyclic graph)描述的工作流. 工作流调度就是为子任务选择并分配合适的制造服务资源, 任务和服务的映射通常情况多为 NP 难问题^[2].

现有 workflow 调度算法^[3-7]通常不考虑任务间数据传输或产品运输所需代价. 但在云制造环境下, 因制造企业地理位置的分散性, 运输代价有时甚至超过服务本身代价, 所以企业间制造产品的运输代价不能忽略, 必须作为调度约束的重要部分加以考虑.

2 制造 workflow 多目标调度模型

制造 workflow 可采用有向无环图 DAG(directed acyclic graph)来描述. 令 DAG 图记作 $G = \{V, E\}$, 其中 $V = \{1, 2, \dots, n\}$ ($|V| = n$) 表示所有子任务集合; E 是有向边集合表示子任务间的制造依赖关系, 即对 $\forall (i, j) \in E$, i 执行结束后 j 才可执行.

收稿日期: 2012-06-05; 修回日期: 2012-07-13

基金项目: 国家“八六三”高科技计划项目(2011AA040502)

对 $\forall i \in V$, 存在相应的候选服务资源池, 记作 $P(i) = \{s_i^1, s_i^2, \dots, s_i^{m_i}\}$, 其中 s_i^j ($1 \leq j \leq m_i$) 由制造企业提供, m_i 为 $P(i)$ 所有可用服务资源数目. 服务资源 s_i^j 属性有 $s_i^j.c, s_i^j.t$ 分别表示 s_i^j 的执行费用、执行时间. 模型中同时考虑服务资源 s_i^j 所在地域位置属性 $s_i^j.a$, 用以区别企业所在不同区域, 用 $\text{Trc}(s_i^p, s_j^q)$ 和 $\text{Trt}(s_i^p, s_j^q)$ ($i \neq j, 1 \leq p \leq m_i, 1 \leq q \leq m_j$) 分别表示服务资源 s_i^p 与 s_j^q 所在地域之间制造产品的运输费用和运输时间. 工作流的执行时间、费用包含所有子任务所选服务资源的执行时间、费用以及子任务间制造产品运输所需时间、费用.

制造工作流的调度问题就是为各任务分配合适的制造服务资源, 使得完成该制造工作流所需时间、费用达到最优. 故考虑运输代价的多目标制造工作流调度问题可形式化描述为:

$$\min y = F(x) = (T(x), C(x)), \quad x \in S \quad (1)$$

$$\text{s. t.} \quad \sum_{k=1}^{m_i} q_{ik} = 1, \quad \forall i \in V \quad (2)$$

$$f_i \leq f_j - \sum_{k_1=1}^{m_j} s_j^{k_1}. tq_{jk_1} - \text{Trt}(\sum_{k_2=1}^{m_i} s_i^{k_2} q_{ik_2}, \sum_{k_3=1}^{m_j} s_j^{k_3} q_{jk_3}) \quad (3)$$

$$\forall (i, j) \in E, q_{jk_1}, q_{ik_2}, q_{jk_3} \in \{0, 1\}, \\ i \in V, \quad 1 \leq k \leq m_i$$

式(1)为目标函数, S 表示解空间, 对 $\forall x \in S$, $T(x), C(x)$ 表示解 x 执行时间和费用, 其中 $C(x) =$

$$\sum_{i \in V} \sum_{1 \leq k \leq m_i} q_{ik} s_i^k.c + \sum_{(i,j) \in E} \text{Trc}(\sum_{k_2=1}^{m_i} s_i^{k_2} q_{ik_2}, \sum_{k_3=1}^{m_j} s_j^{k_3} q_{jk_3}).$$

式(2)表示为每个任务分配一个且仅分配一个服务资源. f_i 表示任务 i 的完成时间, 式(3)表明任务间满足偏序约束关系.

定义 1 Pareto 支配: 不失为一般性, 仅考虑最小化. 设 $u = (u_1, \dots, u_k)$ 和 $v = (v_1, \dots, v_k)$ 为两个自变量矢量, 那么支配 u Pareto 支配 v 当且仅当 $u_i \leq v_i, i=1, \dots, k$ 并且至少有一项 $u_i < v_i$.

定义 2 Pareto 最优解: 多目标问题的一个矢量解 x 是最优解当且仅当不存在另一个矢量 y , 使得 $f(y)$ Pareto 支配 $f(x)$, 所有的 Pareto 最优解称为 Pareto 最优解集.

定义 3 Pareto 最优前沿集: 对于一个多目标问题的 Pareto 最优解矢量 x , 则 $y = (f_1(x), \dots, f_k(x))$ 为 x 的 Pareto 前沿, 所有的 Pareto 最优前沿称为 Pareto 最优前沿集.

因无时间、费用的具体约束, 考虑运输代价的多

目标制造工作流调度结果为一组时间、费用的 Pareto 最优前沿.

3 混合多目标遗传粒子群算法

3.1 算法框架设计

根据制造工作流的调度问题特点, 结合遗传和粒子群两种算法优点^[8], 得到一个整体性能更优的混合多目标遗传粒子群算法(MOGA-PSO). 该算法前期利用 GA 进行搜索, 得到初步的优化结果, 后期利用 PSO 强化局部搜索, 提高收敛速度. 为防止迭代过程中 Pareto 最优解的丢失, 算法维护保存最优解的外部档案, 优化框架如图 1 所示.

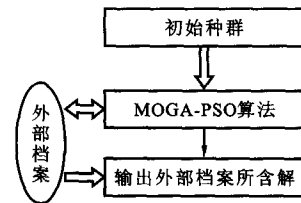


图 1 算法优化框架图

3.2 编码方式

本文采用 n 维向量来表示解向量, n 为 DAG 图中子任务总数, 每一维对应当前子任务所选服务资源序号, 即子任务和资源的映射关系, 第 i 代第 j 个微粒 X_i^j 表示如下:

$$X_i^j = (k_1, k_2, \dots, k_n)$$

$$(1 \leq i \leq g_{\max}, 1 \leq j \leq m, 1 \leq k_p \leq m_p)$$

其中 m 为群体规模, $g_{\max}$ 为最大迭代次数.

3.3 算法参数设计

3.3.1 遗传算法部分参数设计

(1) 选择算子

选择算子采用 Pareto 最优解保存和随机选择的混合策略. 对于大小为 K 的种群, 首先用 Pareto 支配得到此种群中的 Pareto 解, 应用 Pareto 解更新外部档案, 同时这些解被选择到新一代种群中. 经 Pareto 最优解保存后新种群大小小于 K , 则随机选择种群中的解加入新种群, 直到大小为 K 为止.

(2) 交叉算子

采用最基本的单点交叉来实现该算子, 同时采用动态变化的交叉概率 pc :

$$pc = PC_{\min} + (PC_{\max} - PC_{\min}) * (g_{\max} - k) / g_{\max} \quad (4)$$

其中, k 为当前迭代次数, $g_{\max}$ 为最大迭代次数, $PC_{\min} = 0.4, PC_{\max} = 0.8$.

(3) 变异算子

随机选择变异位置,变异位置总数不超过解向量维数的三分之一.动态变化的变异概率:

$$pm = PM_{\max} - (PM_{\max} - PM_{\min}) * (g_{\max} - k) / g_{\max}$$
 (5)

其中, k 为当前迭代次数, $g_{\max}$ 为最大迭代次数, $PM_{\min} = 0.1$, $PM_{\max} = 0.4$.

3.3.2 粒子群算法部分的参数设计

粒子群算法速度更新公式如下:

$$V_{i+1} = w \times V_i + c_1 \times r_1 \times (pbest_i - X_i) + c_2 \times r_2 \times (gbest_i - X_i)$$
 (6)

$$X_{i+1} = X_i + V_{i+1}$$
 (7)

(1)惯性权重 w 、加速常数 c_1 和 c_2

根据经验,取 w 为 1, c_1 和 c_2 为 2.

(2)最大速度 $V_{\max}$ 和最大位置 $X_{\max}$

速度 v 和解向量 x 具有相同的维数,为防止微粒飞出解空间,应设定最大速度 $V_{\max}$ 和最大位置 $X_{\max}$. 设第 i 个任务对应资源池中资源数量为 m_i ,因速度可为负值,所以 v_i 范围 $[-(m_i - 1), (m_i - 1)]$,因位置不可为负,所以 x_i 范围 $[0, m_i - 1]$.

3.4 混合算法切换点的设置

为了最大程度发挥遗传算法全局搜索能力及粒子群算法局部快速收敛能力,寻找两种算法的最佳结合点是混合算法的关键所在. 本文采用动态融合策略. 首先运行遗传算法部分,设置阈值 L ,用于记录迭代过程中外部档案连续没有被更新的次数,若次数达到阈值 L ,则说明子代群体的进化率过低,遗传的优化速度变慢,可切换至粒子群算法继续运行. 在没有超过最大迭代次数 $g_{\max}$ 的前提下,若粒子群算法连续 L 次没有更新外部档案,则算法终止,记录此时迭代次数 t ,则总的迭代次数为 $\min(g_{\max}, t)$.

4 仿真及结果分析

为了评估混合 MOGA-PSO 算法性能,本文定义比较规则如下:相同运行时间内, Np 表示算法所求解个数占全部解个数的百分比, V 表示算法所求解中非劣于其他算法的解占全部非劣解的百分比,以下分别称搜索率、优解率. 求解结果与标准 GA 算法、PSO 算法进行比较.

算法采用 C++ 编程,操作系统 Windows Xp,运行在 Pentium(R) Dual-Core, 2.79GHz, RAM 为 2GB 的 PC 机上. 制造 workflow 对应 DAG 图任务节点数设为 {5, 10, 20}, 图结构随机生成,任务对应资源池中资源数目为 3 到 10 的随机数,资源执行所需费用与时间、异地运输所需费用与时间均满足 ct

$= 2000$. 资源所在地域位置随机生成,地域位置总数设为 10,同一地域之间运输时间费用均为 0.

S 表示每代中所含解个数, T 表示算法运行时间(s 为单位), N 表示 DAG 图中所含任务节点数. 首先给出一次运行中三种算法各自求出的 Pareto 解集,其中 $S=100$, $N=20$, $T=5$,如图 2 所示. 由图 2 可直观看出,MOGA-PSO 算法得到的解较优.

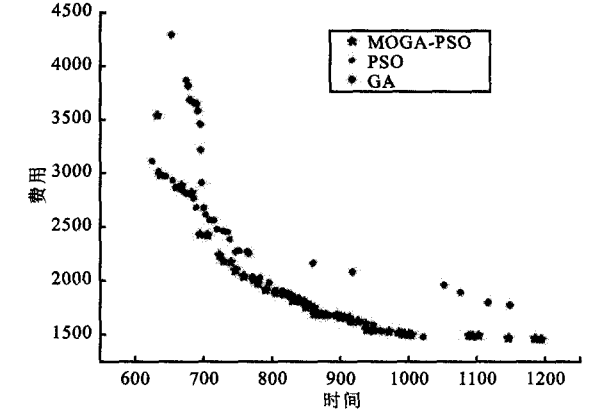


图 2 三种算法得到的 Pareto 解

其次采用控制变量法通过多次运行取平均值来比较算法性能. 表 1 为设置 $S=100$, $T=5$, N 分别取 5、10、20. 表 2 设置 $S=100$, $N=20$, T 分别取 2、5、10. 表 3 设置 $N=20$, $T=5$, S 分别取 50、80、100. 实验重复运行 30 次,结果均保留两位有效数字.

对表 1、表 2、表 3 数据进行分析知:

(1)固定其他变量,变化任务节点数 N . 当 N 较小时,三种算法性能差异不大,但随着 N 的增大,MOGA-PSO 算法的搜索率和优解率显著增加,优于另外两种算法.

(2)固定其他变量,变化运行时间 T . 随着 T 的增大,MOGA-PSO 算法的搜索率和优解率显著增加.

(3)固定其他变量,变化种群大小 S . S 对三种算法性能影响不大,但 MOGA-PSO 表现更优. 由上分析知,相较于 GA 算法和 PSO 算法,混合 MOGA-PSO 算法对于求解带有运输代价的制造 workflow 调度表现出更优的性能.

表 1 任务节点数 N 变化算法性能比较

		GA	PSO	MOGA-PSO
$N=5$	Np	0.34	0.34	0.33
	V	0.36	0.32	0.32
$N=10$	Np	0.28	0.35	0.37
	V	0.18	0.35	0.47
$N=20$	Np	0.23	0.37	0.40
	V	0.01	0.30	0.69

表 2 运行时间 T 变化时算法性能比较

		GA	PSO	MOGA-PSO
T=2	Np	0.27	0.36	0.37
	V	0.03	0.49	0.47
T=5	Np	0.23	0.37	0.40
	V	0.01	0.30	0.69
T=10	Np	0.20	0.39	0.41
	V	0.01	0.39	0.61

表 3 种群大小 S 变化时算法性能比较

		GA	PSO	MOGA-PSO
S=50	Np	0.21	0.37	0.41
	V	0.01	0.31	0.68
S=80	Np	0.22	0.38	0.41
	V	0.02	0.33	0.66
S=100	Np	0.23	0.37	0.40
	V	0.01	0.30	0.69

5 结束语

本文构建了带有产品运输时间、费用代价的无约束多目标制造 workflow 调度问题模型,提出了一种多目标混合遗传粒子群(MOGA-PSO)算法,讨论了算子设计、参数选择和混合算法切换点设置等问题.通过和标准 GA、PSO 算法比较,MOGA-PSO 算法对于解决云制造环境下的 workflow 调度问题具有更强的适应性.后续工作要进一步思考产品运输时间、费用代价对调度算法产生的影响,改进解的编码表示方式以及相关算子,从而进一步提高算法性能.

参考文献:

[1] 李伯虎,张霖,王时龙,等.云制造—面向服务的网络化制造新模式[J].计算机集成制造技术,2010,16(1):1—7.

[2] Ullman J D. NP—complete scheduling problems [J]. Journal of Computer and System Sciences, 1975, 10 (3):384—393.

[3] Yu J, Buyya R, Tham C K. A cost—based scheduling of scientific workflow applications on utility grids[C] //Proc. of the IEEE International Conference on e—Science and Grid Computing, Melbourne, Australia, 2005:140— 147.

[4] 苑迎春,李小平,王茜,等.成本约束的网格 workflow 时间优化方法[J].计算机研究与发展,2009,46(2):194— 201.

[5] 苑迎春,李小平,王茜.基于逆向分层的网格 workflow 任务调度[J].计算机学报,2008,31(2):282—290.

[6] 苑迎春,李小平,王茜.基于串规约的网格 workflow 费用优化方法[J].计算机研究与发展,2008,45(2):264— 253.

[7] Yu J, Kirley M, Buyya R. Multi—objective planning for workflow fxecution on grids[C]// Proc. of the 8th IEEE/ACM International Conference on Grids Computing. Melbourne Australia, 2007:10—17.

[8] Liu J, Li J, Liu K,et al. A hybrid henetic and particle swarm algorithm for service composition [C]// 6th International Conference on Advanced Language Processing and Web Information Technology. Japan, Niigata, 2007:564— 567.

作者简介:

李 聪 女,(1987—),硕士研究生.研究方向为制造 workflow 调度.

汪 芸 女,(1967—),教授,博士生导师.研究方向为分布式计算、传感器网络和容错理论.

罗香玉 女,(1984—),博士研究生.研究方向为分布式存储.