

文章编号:1007-5321(2017)01-0099-06

DOI:10.13190/j.jbupt.2017.01.018

逆向归约时间约束 workflow 准确率优化调度

罗智勇^{1,2}, 汪鹏¹, 尤波², 苏洁¹

(1. 哈尔滨理工大学 计算机科学与技术学院, 哈尔滨 150080;

2. 哈尔滨理工大学 机械动力工程学院, 哈尔滨 150080)

摘要: 针对时间约束有向无环图表示的业务流程 workflow 准确率优化问题, 提出了基于截止期的逆向归约优化算法, 摒弃传统算法的单向目标策略, 解决了业务流程完工准确率过低或者完工时间过长的问题. 通过整合传统算法遗留下来的时间碎片, 分析了服务准确率与时间的制约关系, 以逆向归约方式求出优化路径. 模拟数据表明, 逆向归约优化算法能够实现截止期范围内时间与准确率的平衡, 相比于传统算法在相同约束时间下对准确率起到了优化效果. 通过分析业务流程的截止期大小和任务数对算法性能的影响可知, 截止期或任务数的增大提升了算法优化效果.

关键词: 工作流; 时间约束; 逆向迭代; 准确率优化

中图分类号: TP393

文献标志码: A

Optimization Scheduling of Workflow's Accuracy Based on Reverse Reduction under Constraint Time

LUO Zhi-yong^{1,2}, WANG Peng¹, YOU Bo², SU Jie¹

(1. School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China;

2. School of Mechanical Engineering, Harbin University of Science and Technology, Harbin 150080, China)

Abstract: Aiming at optimizing accuracy in business process workflow represented by directed acyclic graph (DAG) in time constraint, the reverse reduction optimization algorithm (RRO) based on deadline was proposed. This algorithm gives up one-way target strategy of traditional algorithm and eliminates the problems of the low accuracy or high completion time of business process. By integrating the time debris left by the traditional algorithm and analyzing the relationship between the accuracy and the time of the service, the optimal path is obtained by reverse reduction. Simulation shows that the reverse reduction optimal algorithm can achieve the balance between completion time and completion accuracy in the range of deadline. Furthermore, the influence of the performance to the algorithm when it applies to different deadlines and different numbers of tasks is analyzed by different simulation data and the increasing of the deadline or the number of tasks improves the algorithm optimization performance.

Key words: workflow; time constraint; reverse iteration; accuracy optimization

以服务为结点映射到网络流从而建立有向无环图(DAG, directed acyclic graph)是供应链技术建模

之一^[1], 在规定的时间内选择合理的服务对优化 workflow 起着关键性作用. 针对混合式 workflow, 以

收稿日期: 2016-10-17

基金项目: 国家自然科学基金青年项目(61403109)

作者简介: 罗智勇(1978—), 男, 副教授, 硕士生导师, E-mail: luozhiyongemail@sina.com.

时间启发函数为基础,通过遗传算法或者动态重排策略可有效地减少成本^[2-3];以关键路径的完工时间与截止期对任务环节扩大活动范围,进而确定时间与成本的制约关系来实现成本的优化^[4-5].

笔者的贡献主要在于提出了一种基于复杂产品工作流在约束时间内控制任务的执行区间,进而优化准确率的逆向归约优化算法,并通过准确率监测点检测反馈,实现准确率的提升^[6].

1 问题描述

1.1 工作流模型的相关定义

工作流建模是将业务流程映射到 DAG 中,是利用业务所有信息进行构图的过程.

定义 1 任务池. 即产品制造过程中所有任务组成的集合,用 P 表示, $P = (p_1, p_2, \dots, p_i, \dots, p_n)$.

定义 2 服务池. 即任务池 P 中任意元素 p_i 对应的服务集合. 用 S_i 表示,则 $S_i = (s_1, s_2, \dots, s_j, \dots, s_m)$.

定义 3 服务池的秩. 即任务 p 对应的服务池中可以完成该任务的服务个数,表示为 r_p .

定义 4 准测点. 业务流程完工时对准确率进行检测,若未达到限定准确率的要求,则反馈循环,重新加工;若达到限定准确率要求,则进入后续环节,进行加工. 准测点用 M 表示,对应的值表示为 ξ . 考虑到成本问题,笔者仅进行 1 次循环.

定义 5 工作流图. 若用 G 表示,可形式化为四元组 $G = \{s_{pk}, E, M, S_p\}$, s_{pk} 为任务 p 用其对应的服务池 S_p 中的第 k 个服务完成, E 为有向边,表示结构流程, S_p 为任务 p 的服务池,且工作流图中存在结点 p_s 和结点 p_e ,分别表示为业务流程的开始和结束.

定义 6 服务属性. 是指工作流图 G 中服务 s_{pk} 完成任务 p 所花费的时间和达到的准确率,表示为 $pt_{pk} = (t_{pk}, a_{pk})$,其中 $p \in P, 0 < k \leq r_p, t_{pk}$ 为花费的时间, a_{pk} 为达到的准确率.

1.2 工作流的约束分析

定义 7 完工时间. 通过对每个任务 p 配置一个服务来实现整个工作流程的完工,其所花费的时间,表示为 T .

定义 8 完工准确率. 通过对每个任务 p 配置一个服务来实现整个工作流程的完工,其所达到的准确率,表示为 A .

定义 9 完工截止期. 指工作流程的完工约束时间,表示为 ψ .

工作流的调度是建立在完工时间低于截止期的前提下才进行优化的,这一约束条件可形式化为

$$\left. \begin{aligned} A &= \prod l_{pk} a_{pk} + A' \geq \xi \\ T &= \sum l_{pk} t_{pk} + T' \leq \psi \\ p &\in P, \quad 0 < k \leq r_p \end{aligned} \right\} \quad (1)$$

其中: A' 为经过反馈后提升的准确率, T' 为多花费的时间.

$$\left. \begin{aligned} \text{s. t. } \sum_{k=1}^{r_p} l_{pk} &= 1 \\ l_{pk} &\in \{0, 1\} \\ p &\in P, \quad 0 < k \leq r_p \end{aligned} \right\} \quad (2)$$

式(1)表示完工时间与准确率相对应的制约关系,即必须在截止期范围内以求得准确率;式(2)中的 l_{pk} 为一个变量,表示当执行某个任务时只能选择其服务池中一个服务来完成.

2 传统单向目标算法

传统单向目标算法策略是基于当前最优解的,即都选择执行时间最短的服务或者执行准确率最大的服务.

2.1 基于时间最小化的算法策略

在每层任务中选择完成时间最小的服务,形成一条工作流路径,求解该方法的准确率为

$$\left. \begin{aligned} T &= \sum \min(t_{pk}) + T' \\ A &= \prod a_{pk} + A', \quad 0 < k < r_p \\ p &\in P \end{aligned} \right\} \quad (3)$$

2.2 基于准确率最大化的算法策略

实现完工准确率最大化,即每个任务由服务集合中准确率最大的服务来完成,完工准确率才能达到最大化,计算该策略的完工时间为

$$\left. \begin{aligned} A &= \prod \max(a_{pk}) + A' \\ T &= \sum t_{pk} + T', \quad 0 < k \leq r_p \\ p &\in P \end{aligned} \right\} \quad (4)$$

3 逆向归约优化算法

3.1 相关定义

定义 10 任务自由度 W_p . 即工作流模型中任务 p 的开始时间组成的集合,表示为 $W_p = [R_p, N_p]$, $p \in P$, 其中 R_p 为任务 p 最早开始的时间, N_p 为任务 p 最晚开始的时间,有

$$\left. \begin{aligned} R_p &= \sum_{i=1}^{p-1} \min(t_{ik}), \quad 0 < k \leq r_p, R_1 = 0 \\ N_p &= \psi - \sum_{i=n}^p \min(t_{ik}), \quad 0 < k \leq r_p \end{aligned} \right\} \quad (5)$$

3.2 算法描述

逆向归约优化算法(RRO,reverse reduction optimization algorithm)通过分析时间与准确率的制约关系,在截止期的基础上实现每个任务环节的分层,建立任务自由度,使每个任务能对服务进行更大范围的选择. 笔者以任务的自由度为基础,求出每个任务 p 的自由度 W_p ,采用由后至前的方式求解每个任务在不同时刻开始时能达到的最大准确率,当遍历到初始任务时,则表示此段工作流程的结束,以此与准测点比较反馈,最终确定约束时间内准确率得到优化的路径.

假设业务流程包含 n 个任务,令 $f(p, t)$ 为第 p 个任务从 t 时刻开始能达到的最大准确率, $t \in [R_p, N_p]$. 最后 1 层任务在对应的自由度区间内,在不同时刻开始时必然能得到 1 个准确率最大值,有

$$\left. \begin{aligned} f(p, t_p) &= \max \{a_{pk}\}, \quad p = n, 0 < k \leq r_p \\ t_p &\in [R_p, N_p], t_p + t_{pk} \leq \psi \end{aligned} \right\} \quad (6)$$

存在任务 p' , 满足 p' 为任务 p 的前驱任务,对整个业务的任务归约后,利用最后 1 层的任务,依次向前求得每个任务在不同时刻开始时的准确率最大值,有

$$\left. \begin{aligned} f(p', t_{p'}) &= \max \{a_{p'k} f(p, t_{p'} + t_{p'k})\} \\ t_{p'} &\in [R_{p'}, N_{p'}] \end{aligned} \right\} \quad (7)$$

通过层层迭代,当遍历到初始任务时,即可得到每完工时间下的最大准确率.

定理 1 一个完整的业务流程若存在 $f(n, t)$ 为截止期 ψ 下的最大准确率,则在截止期 $\psi - t_0$ 下的最大准确率为 $f(n, t + t_0)$.

证明 已知在截止期 ψ 下的最大准确率为 $f(n, t)$,假设 $\psi - t_0$ 下的最大准确率为 $f(n, t + t_1)$,且 $t_1 < t_0$,若 $f(n, t + t_0) = f(n, t + t_1)$,则定理 1 成立;若 $f(n, t + t_0) < f(n, t + t_1)$,则该情况下业务流程的完工时间 ψ' 必定大于 $\psi - t_0$,此时假设与规定矛盾;假设 $\psi - t_0$ 下的最大准确率为 $f(n, t + t_2)$,且 $t_2 > t_0$,则必有 $f(n, t + t_0) \geq f(n, t + t_2)$,则 $f(n, t + t_0)$ 为最大准确率. 综上所述,定理 1 成立.

综上所述,RRO 伪代码如下.

输入:参数 $P; s_{pk}; pt_{pk}; \text{len}$ //len 为集合 R' 中所有任务的数目

输出:最佳路径
By $G;$ //通过生成的 workflow G
For(int $p = 1; p \leq \text{len}; p++$) { List (Formula (5), W_p); } //利用式(5)对 workflow 相应任务取自由度 W_p
 $p = \text{len};$ Figure (Formula (6), a_{pk}); //利用式(6)求得最后 1 层任务不同时刻开始的最大准确率
For($p; p > 0; p--$) { Figure (Formula (7), a_{pk}); } //利用式(7)依次求解前驱的任务在不同时间开始达到的最大准确率
Foreach(A) { If($A < \xi$) { $A = A + A(1 - A)$ } $T = T + T'$ } //对每种完工时间下的结果进行准测点检验,并循环
Search $\max(A_i)$ Where $T \leq \psi;$ //查找截止期内最大准确率的路径,即最佳路径
该算法的时间复杂度为 $O(m^n)$.

4 实验结果与分析

4.1 案例设计

经调研,某机械生产流程分为 6 个环节:冲压环节 p_1 、焊接环节 p_2 、涂装环节 p_3 、部件装配环节 p_4 、总装配环节 p_5 、虚拟的申请环节 p_s 、虚拟的竣工环节 p_e . 其中申请和竣工作为虚拟任务,表示流程的开始和结束,设置准测点 M 对应的 $\xi = 0.90$,若 $A < \xi$,则进行反馈,多花费时间为 1,假设每循环 1 次则多耗费诸多成本,因此当 $A \geq 0.90$ 即完工. 任务对应的服务池为 S_p ,属性如表 1 所示.

表 1 服务时间和准确率

任务集合	任务对应的服务池
p_1	$s_{11}(2, 0.91); s_{12}(4, 0.94); s_{13}(5, 0.98)$
p_2	$s_{21}(3, 0.94); s_{22}(5, 0.97)$
p_3	$s_{31}(2, 0.95); s_{32}(3, 0.96); s_{33}(5, 0.98)$
p_4	$s_{41}(3, 0.93); s_{42}(4, 0.97)$
p_5	$s_{51}(3, 0.92); s_{52}(5, 0.96); s_{53}(6, 0.98)$

通过以上任务池 P 和每个任务 p 对应的服务池 S_p ,对该业务流程建立如图 1 所示的工作流.

4.2 算法分析

基于图 1 所示的工作流,规定本业务流程的完工截止期 $\psi = 17$,分别以 2 种策略完成工作流.

4.2.1 传统单向目标算法策略

1) 基于时间最小化的单向目标算法
基于完工时间最小化,即所有任务都选择执行

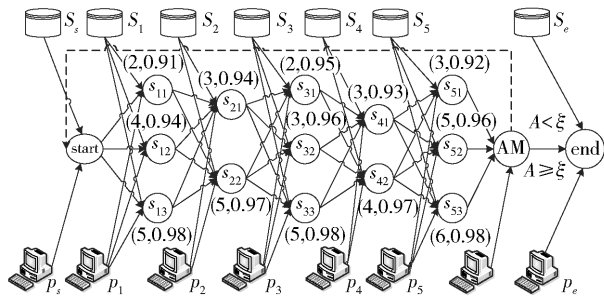


图 1 workflow

时间最小的服务,则调用式(3)求得对应的完工时间为

$$T=t_{11}+t_{21}+t_{31}+t_{41}+t_{51}+t_{61}=13$$

准确率 $A=a_{11}a_{21}a_{31}a_{41}a_{51}=0.695$. 由于 $A<\xi$,则进行反馈,此时

$$A=0.695+(1-0.695)\times0.695=0.907$$

$$T=13+1=14$$

2) 基于准确率最大化的单向目标算法

基于完工准确率最大化,即所有任务都选择准确率最大的服务,则调用式(4)求得对应的 $A_{\max}=a_{13}a_{22}a_{33}a_{42}a_{53}=0.886$,则对应工作流程的完工时间 $T=t_{13}+t_{22}+t_{33}+t_{42}+t_{53}=25$,由于 $A<\xi$,则进行反馈,此时 $A=0.886+(1-0.886)\times0.886=0.987$, $T=25+1=26$.

4.2.2 逆向归约优化算法

调用逆向归约算法可得该业务流程的任务集合 P 、任务对应的服务集合 S_p 和服务属性 pt_{pk} . 冲压环节 p_1 的自由度为 $[0,4]$,焊接环节 p_2 的自由度为 $[2,6]$,涂装环节 p_3 的自由度为 $[5,9]$,部件装配环节 p_4 的自由度为 $[7,11]$,总装配环节 p_5 的自由度为 $[10,14]$,则在对应任务的自由度内,可求出各任务在不同时刻开始达到的最大准确率,如表 2 所示.

由表 2 的结果并结合定理 2 可得,完工时间 $T=\{17,16,15,14,13\}$ 下最大准确率为 $A=\{0.780,0.756,0.733,0.724,0.695\}$. 根据准测点检测结果,需要循环反馈的路径完工所用时间 T' 、完工准确率 A' 如表 3 所示.

由表 3 可知,反馈后准确率最大路径的完工时间为 18,超过完工截止期,因此舍弃该方案,选择准确率 $A=0.940$ 和完工时间 $T=17$ 的优化路径.

表 2 不同任务环节的计算过程

任务	计算过程
	$f(5,10)=\max\{0.92,0.96,0.98\}=0.98$
	$f(5,11)=\max\{0.92,0.96,0.98\}=0.98$
p_5	$f(5,12)=\max\{0.92,0.96\}=0.96$
	$f(5,13)=\max\{0.92\}=0.92$
	$f(5,14)=\max\{0.92\}=0.92$
p_4	$f(4,7)=\max\{0.93f(5,10),0.97f(5,11)\}=0.951$
	$f(4,8)=\max\{0.93f(5,11),0.97f(5,12)\}=0.931$
	$f(4,9)=\max\{0.93f(5,12),0.97f(5,13)\}=0.893$
	$f(4,10)=\max\{0.93f(5,13),0.97f(5,14)\}=0.892$
p_3	$f(4,11)=\max\{0.93f(5,14)\}=0.856$
	$f(3,5)=\max\{0.95f(4,7),0.96f(4,8),0.98f(4,10)\}=0.903$
	$f(3,6)=\max\{0.95f(4,8),0.96f(4,9),0.98f(4,11)\}=0.884$
	$f(3,7)=\max\{0.95f(4,9),0.96f(4,10)\}=0.856$
p_2	$f(3,8)=\max\{0.95f(4,10),0.96f(4,11)\}=0.847$
	$f(3,9)=\max\{0.95f(4,11)\}=0.813$
	$f(2,2)=\max\{0.94f(3,5),0.97f(3,7)\}=0.848$
	$f(2,3)=\max\{0.94f(3,6),0.97f(3,8)\}=0.831$
p_1	$f(2,4)=\max\{0.94f(3,7),0.97f(3,9)\}=0.805$
	$f(2,5)=\max\{0.94f(3,8)\}=0.796$
	$f(2,6)=\max\{0.94f(3,9)\}=0.764$
	$f(1,0)=\max\{0.91f(2,2),0.94f(2,4),0.98f(2,5)\}=0.780$
	$f(1,1)=\max\{0.91f(2,3),0.94f(2,5),0.98f(2,6)\}=0.756$
	$f(1,2)=\max\{0.91f(2,4),0.94f(2,6)\}=0.733$
	$f(1,3)=\max\{0.91f(2,5)\}=0.724$
	$f(1,4)=\max\{0.91f(2,6)\}=0.695$

表 3 各完工时间及准确率

T	A	是否反馈	T'	A'
17	0.780	True	18	0.952
16	0.756	True	17	0.940
15	0.733	True	16	0.929
14	0.724	True	15	0.924
13	0.695	True	14	0.907

4.3 算法比较

针对传统单目标算法和逆向归约优化算法,并结合 4.2 节中各数据,做出如图 2 所示的算法对比结果. 图 2(a)给出了工作流程在不同算法基础上执行到不同阶段时的完工时间和准确率,图 2(b)给出了不同算法对应的完工路径.

从图 2(a)所知,基于准确率最大化算法策略的完工准确率最高,但完工时间超过了截止期,因此被

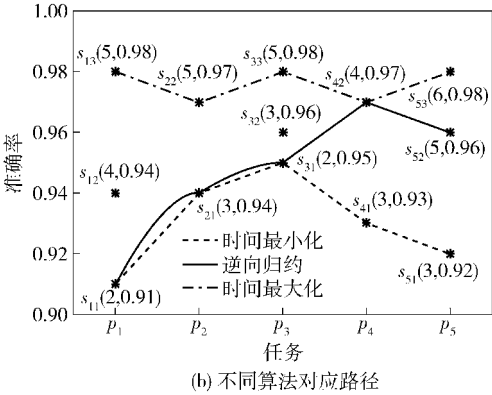
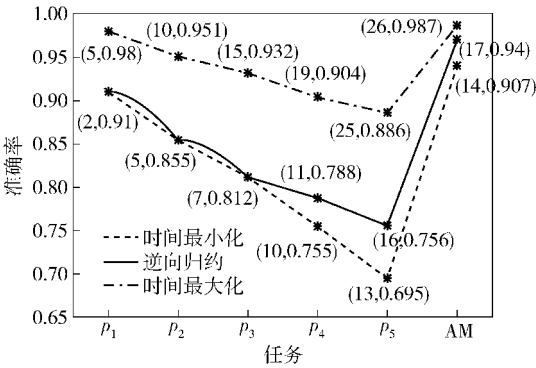


图 2 算法对比

舍弃。基于时间最小化算法策略和逆向归约算法策略的完工时间满足规定条件,其对应的完工准确率分别为 $A_1 = 0.940$ 、 $A_2 = 0.907$,可计算逆向归约算法的提高率 $K = (A_2 - A_1) / A_1 \times 100\% = 3.63\%$ 。因此,逆向归约算法相对于传统单向目标算法起到了优化提升效果,且由图 2(b)得到的对应路径为 $s_{11} \rightarrow s_{21} \rightarrow s_{31} \rightarrow s_{42} \rightarrow s_{52}$ 。

4.4 其他参数对算法性能的影响

4.4.1 截止期 ψ 的大小产生的影响

业务流程随着截止期 ψ 的增大可选择更加适合的服务,完工准确率也可得到提高。随机生成 {10, 15} 个任务数,每个任务服务池的大小取区间 [2, 4],每个服务生成相应的属性;规定准测点 M 的 $\xi = 0.90$,当完工准确率低于该值时进行反馈,增加完工时间为 1,且仅循环 1 次。假设每个业务流程的最小完工时间为 T_{min} ,分别以 T_{min} 增加 10%、15%、20%、25%、30% 作为不同的截止期,图 3 体现不同的截止期对算法性能的影响。

从图 3 可见,当截止期为 T_{min} 时,即基于时间最小化策略,将任务数为 10 的业务流程增加 10%、15%、20%、25%、30% 作为截止期,逆向归约算法的提高率分别为 3.1%、4.4%、7.4%、8.9%、11.0%;

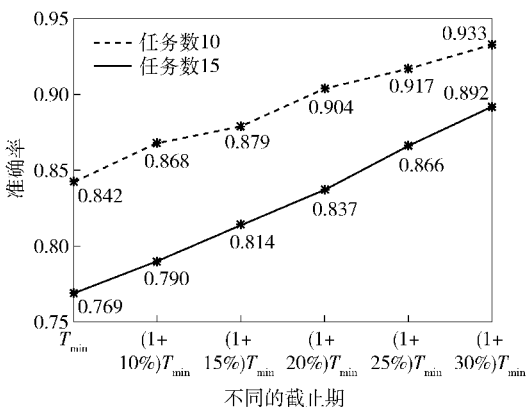


图 3 不同截止期 ψ 对算法性能的影响

任务数为 15 的业务流程,逆向归约算法的提高率分别为 2.7%、5.9%、8.8%、12.6%、16.0%。

4.4.2 任务数对性能的影响

业务流程任务数的变化对总体准确率有影响,同时也影响着逆向归约算法的优化效果。随机生成 {5, 10, 15, 20} 个任务数,每个任务服务池的大小取区间 [2, 4],每个服务并生成相应的属性;规定准测点 M 的 $\xi = 0.90$,当完工准确率低于该值时进行反馈,增加完工时间为 1,且仅循环 1 次。假设每个业务流程的最小完工时间为 T_{min} ,以 T_{min} 增加 20% 作为不同流程的截止期,图 4 所示为不同任务数对算法性能的影响。

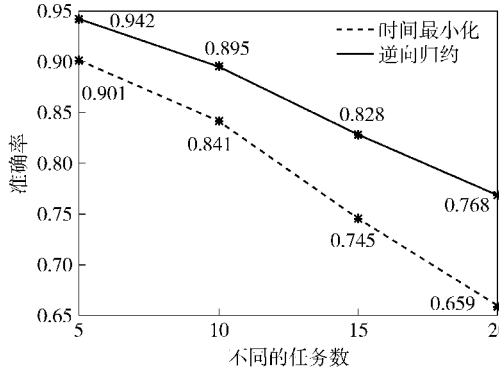


图 4 任务数对算法性能的影响

从图 4 可以得出,随着任务数的增多,逆向归约优化算法的提高率分别为 4.6%、6.4%、11.1%、16.5%。

5 结束语

针对业务流程时间和准确率平衡问题,笔者首先将业务流程的任务环节映射到 DAG 中,生成 workflow 模型。提出了基于 workflow 时间最小化和准确率最

大化的传统单向目标算法,针对该传统算法存在的完工准确率过低或完工时间过长的弊端,提出了在截止期约束下的逆向归约优化算法,解决了传统算法产生时间碎片等问题,能更好地利用时间,从而对每个任务选择更合理的服务. 模拟实验结果表明,逆向归约优化算法相对于传统的单向目标算法,其性能提高了3.63%. 笔者以截止期和任务数作为影响算法性能的参数,利用对应参数可提升算法性能.

参考文献:

- [1] Czarnul P. Modeling, run-time optimization and execution of distributed work-flow applications in the JEE-based BeesyCluster environment[J]. The Journal of Supercomputing, 2013, 63(1): 46-71.
- [2] Nasonov D, Butakov N, Balakhontseva M, et al. Hybrid evolutionary workflow scheduling algorithm for dynamic heterogeneous distributed computational environment[J]. Springer International Publishing, 2014, 299(6): 83-92.
- [3] Chen Wei, Lee Youngchoon, Fekete A, et al. Adaptive multiple-work-flow scheduling with task rearrangement[J]. Journal of Supercomputing, 2015, 71(4): 1297-1317.
- [4] 苑迎春, 李小平, 王茜, 等. 基于逆向分层的网格 workflow 调度算法[J]. 计算机报, 2008, 31(2): 282-290.
Yuan Yinchun, Li Xiaoping, Wang Qian, et al. Bottom level based heuristic for workflow scheduling in grids[J]. Chinese Journal of Computers, 2008, 31(2): 282-290.
- [5] 刘灿灿, 张卫民, 骆志刚. 基于逆向分层的工作流时间-费用优化方法[J]. 国防科技大学学报, 2013, 35(3): 61-66.
Liu Cancan, Zhang Weimin, Luo Zhigang. Time and cost trade-off heuristics for workflow scheduling based on bottom level[J]. Journal of National University of Defense Technology, 2013, 35(3): 61-66.
- [6] Luo Zhiyong, Wang Peng, You Bo, et al. Serial reduction optimization research of complex product workflow's accuracy under the time constraint[J]. Advances in Mechanical Engineering, 2016, 8(10): 1-9.