

基于粒子群算法的工作流服务主体优选方法

韩 猛

(聊城职业技术学院, 聊城 山东 252000)

摘 要: 研究工作流服务主体优选问题, 在工作流系统中, 工作机负载能力有差异性, 而且整个系统负载具有动态性, 传统算法难以获得最优工作流服务主体优选方案, 导致系统资源利用率较低. 为了提高系统资源利用率, 系统负载保持均衡, 提出一种粒子群算法的工作流服务主体优选方法. 首先对工作流服务主体优选问题建立相应数学模型, 然后采用粒子群算法对其进行求解, 即工作流服务主体最优选择方案, 最后进行仿真测试. 测试结果表明, 相对于传统方法, 粒子群算法可以针对不同类型的任务分配不同的工作机, 实现系统多种资源的负载均衡, 提高系统资源的利用率.

关键词: 工作流; 负载均衡; 多目标优化; 粒子群算法

中图分类号: TP391

文献标识码: A

文章编号: 1000-7180(2012)11-0189-04

Preferential Choice Scheme of Workflow Service Provider Based on Particle Swarm Optimization

HAN Meng

(Liaocheng Vocational and Technical College, Liaocheng 252000, China)

Abstract: Research on workflow service subject selection issues, in workflow management system, work load ability difference, and the system load is dynamic, the traditional algorithm is difficult to obtain the optimal workflow service subject optimization scheme, leading to low utilization rate of system resources. In order to keep the system responsible for balance, raise the utilization rate of resources system, a method based on particle swarm algorithm optimization method of workflow service subject. The workflow services body preferably built corresponding mathematical model, and then the particle swarm optimization algorithm to solve the model, workflow service main body the optimal solution, finally carries on the simulation test. The test results show that, compared with the traditional method, particle swarm optimization algorithm can be used for different types of task allocation in different working machine, improve the utilization rate of system resources, to achieve a variety of resource load balancing system.

Key words: workflow; load balancing; multi-objective optimization; particle swarm optimization

1 引言

工作流技术是一种新兴的计算机应用技术, 广泛应用于办公自动化、电子邮件、目录管理等多个领域. 迁移工作流很好地解决了传统工作流存在的各种难题, 改善了传统工作流系统的灵活性, 成为工作流研究中的热点问题^[1].

在工作流系统中, 停靠站服务器为迁移实例分

配工作机, 不同的服务主体优选方法, 工作机上负载不同, 整个系统的资源利用率不同^[2]. 当前主要迁移实例分配算法有: 轮转调度算法、最小连接度算法和目标地址散列调度和源地址散列调度, 这些算法实际都是一些网络负责算法, 因此都有各自的不足^[3-5]. 由于每个迁移实例的迁移域同, 轮转调度算法没法实施. 最小连接度算法忽略任务大小, 不能有效辨别负载差异. 目标和源地址散列调度容易引起

负载分布不均^[6]. 同时这些算法均无能准确反映工作机负载的动态变化和区分不同迁移实例占用的系统资源差异, 因此不能很好保持多种资源的负载动态均衡^[6]. 随后有学者提出了模糊综合评判算法和遗传算法的工作流服务主体优选机制^[7-8]. 模糊综合评判算法具有较强的主观性, 操作十分复杂, 不能保证系统负载均衡, 导致系统资源利用率较低. 遗传算法是一种人工智能算法, 相对于其它算法, 负载均衡性能更好, 工作流系统的资源利用率更高. 但是其存在局部最优、收敛速度慢等缺陷, 不能很好的适应系统负载均衡性动态性.

为解决当前工作流服务主体优选算法存在的缺陷, 提出一种粒子群算法的服务主体优选机制. 该算法借鉴鸟群觅食行为, 模拟鸟群的飞行过程对迁移域中的工作机进行优选, 并采用 Pareto 方法在最优解集中选取最能反映系统工作状态负载分配方案, 使系统资源得到充分的应用, 最后通过仿真实验检验其有效性.

2 工作流服务主体优选问题与数学模型

2.1 服务主体优选问题

迁移工作流管理系统是在传统工作流系统中引入移动计算技术, 将工作流系统分为工作流引擎、工作位置和迁移实例三部分, 提高了工作流系统动态适应性^[9].

停靠站服务器在进行工作的时候, 需要确定工作状态过程中最适合操作的工作机, 以保证工作区域负载均衡, 进而保证资源被有效利用. 服务主体优选机制需要满足以下目标:

- (1) 保证在区域中工作机的资源占有率最小.
- (2) 迁移域中总资源的平均占有率最小.

2.2 数学模型

假设欲迁移 mi 的迁移域中有 m 台工作机, 有 $n-1$ 个 mi 正在运行, 第 i 台工作机的资源占有率为 G_i , 迁移域中总的资源占有率为 F , n 表示迁移实例个数, 这样根据服务主体优选问题可得如下目标函数:

$$\min G_i = \sum_{j=1}^n S_{ij} C_{mj} / A_i \quad (1)$$

$$\min F_i = \sum_{i=1}^m G_i, \quad i = 1, 2, \dots, m, j = 1, 2, \dots, n \quad (2)$$

约束条件为:

$$\begin{cases} \sum_{i=1}^m S_{ij} = 1 \\ S_{ij} = \{0, 1\} \end{cases} \quad (3)$$

其中, A_i 为第 i 台工作机的总资源, S_{ij} 表示为第 j 个迁移实例在工作机 S_i 上的运行状态.

一个迁移实例在同一时刻只能在一台工作机上运行, 根据式(1)和式(2)可知, 不仅要考虑迁移域中各个工作机的资源占有率, 同时也要考虑系统总资源占有率, 是一个典型的多目标优化问题. 对于多目标优化问题, 如果通过利用传统算法进行求解, 求解速度慢, 耗时长, 且难以获得全局最优解, 无法适应工作流系统的负载动态变化特点. 而粒子群算法是一种群智能算法, 具有参数设置简单、并行和收敛速度快等特点, 因此本文采用粒子群优化算法对服务主体优选的数学模型进行求解.

3 粒子群算法的服务主体优选方法

3.1 粒子群算法

粒子群算法是 Kennedy 和 Eberhart 受到鸟群捕食行为启发提出的一种随机搜索方法, 被用于解决优化问题^[10].

粒子在飞行过程中, 通过跟踪个体最优位置 Pbest 和全局最优位置 Gbest 对自己速度和位置不断更新, 具体更新方式如下:

$$v_{k+1} = v_k + c_1 * \text{rand}() * (Pbest_k - x_k) + c_2 * \text{rand}() * (Gbest_k - x_k) \quad (4)$$

$$x_{k+1} = x_k + v_{k+1} \quad (5)$$

式中, Pbest_k 表示为粒子 k 当前最优位置, Gbest 表示种群中的最优位置, rand() 表示 0 和 1 之间的随机数, c_1 和 c_2 表示为学习因子.

粒子的速度被限制在一个最大速度范围 $v_{\max}$ 内, $v_{\max}$ 约束条件为:

$$v_k = \begin{cases} v_{\max} & (v_k > v_{\max}) \\ v_k & (-v_{\max} < v_k < v_{\max}) \\ -v_{\max} & (v_k < -v_{\max}) \end{cases} \quad (6)$$

3.2 粒子群算法的服务主体优选设计

3.2.1 迁移实例的编码

服务主体优选设计优化目标是通过停靠站服务器, 快速找到迁移实例合适的工作机, 因此首先对每一个粒子进行编码, 从可以代表服务主体优选的数学模型的解, 该解可以明确指出每个迁移实例的工作位置.

在工作机工作中, 可以把其运行的状态设为 1, 非公作为 0, 极其在工作中的状态可以显示为一个

长度为 $m \times n$ 的 0、1 串. 假设 S_i 表示工作机集合, S_{ij} 表示第 j 个迁移实例在工作机 S_i 上的运行状态. 如果其在第 j 台工作机上运行, 则 $S_{ij} = 1$, 否则 $S_{ij} = 0$, 粒子编码方案如图 1 所示.

S_{11}	S_{12}	S_{13}	$\cdots$	S_{1n}	S_{21}	S_{22}	S_{23}	$\cdots$	S_{2n}	$\cdots$	S_{m1}	S_{m2}	S_{m3}	$\cdots$	S_{mn}
0	1	0	$\cdots$	0	1	0	0	$\cdots$	1	$\cdots$	0	0	1	$\cdots$	1

图 1 迁移实例的编码

3.2.2 初始化粒子群

构建粒子群需要对其初始化, 方法有两种, 一种是采用随机分配的方法完成, 这种方法不需要先验知识. 如果拥有一定的先验知识, 可以运用另一种方法, 在满足条件的基础上, 选取随机样本. 保证运算实例的迁移域不一致, 因此采用第二种方法产生初始粒子群.

对于迁移域与待分配工作机的 mi 的迁移域交集为 1 的已运行迁移实例, 粒子群算法的优化结果可能导致其脱离自己的迁移域, 提出编码固定点方式保证相应迁移实例的服务主体不发生变化.

对于迁移实例 k , 如果它的迁移域 $\text{domain}(mi)_k$ 与待分配工作机的 mi 的迁移域 $\text{domain}(mi)$ 的交集为 1, 那么表示在整个粒子群优化过程中, 其位置不需要改变, 其位置的编码都为 1. 在进行数据初始化中, 先要完成固定点的编码工作, 对每一个粒子运用随机编码的方式进行初始化, 进行迭代计算过程, 完成整个粒子群的初始化工作.

3.2.3 适应度函数设计

根据建立的优选模型, 可以知道这个问题是一个多元优化问题, 运用权值计算对单一的源目标进行求解为:

$$\min f(x) = \sum_{i=1}^m \omega_i G_i + \omega_{m+1} F$$

目标函数可直接转换为适应度函数. 取适应度函数为:

$$fit(f(x)) = \begin{cases} C_{\max} - f(x), & f(x) < C_{\max} \\ 0, & \text{otherwise} \end{cases}$$

式中, $C_{\max}$ 表示 $f(x)$ 的最大估计值.

3.2.4 粒子群位置和速度的更新

在粒子群的飞行过程中, 全局最优解引导其它粒子运动的方向和位置变化, 因此按适应度值将粒子群分为两种类型: 精英粒子和普通粒子, 精英粒子在个体优化过程, 无需要进行位置和速度更新, 只引导其它粒子向该粒子聚集, 而普通粒子采用式(4)和(5)的方式更新自己的位置和速度. 对普通粒子进行

更新后, 对其合法性进行判断, 如果是非法解, 就需要对其位置和速度重新进行更新.

3.2.5 算法结束条件

当算法找到最优粒子后, 则粒子群算法优化过程结束.

4 仿真实验

4.1 仿真环境

为了检测基于粒子群算法的工作流服务主体优选方法的性能, 对其进行仿真实验. 工作流系统由 5 台工作机组成, 它们配置具体如表 1 所示.

表 1 工作流系统的工作机配置

	CPU 主频	内存
工作机 1	2.0G	512M
工作机 2	1.8G	1G
工作机 3	2.16G	512M
工作机 4	2.0G	2G
工作机 5	2.4G	1G

4.2 对比算法和评价指标

为了使算法的结果具有说服力, 采用遗传算法、模糊综合评判算法作为对比算法. 系统资源通常可以划分为输入/输出资源和 CPU 资源两类, 系统性能评判标准常采用输入/输出负载和 CPU 使用率, 因此本文采用 CPU 使用率和内存使用率对整个系统性能进行验证.

4.3 实验结果与分析

4.3.1 不同种群规模下的算法执行效率

由于不同种群规模条件下, 粒子群算法找到最优工作流服务主体优选方案的时间, 分别取种群规模为 10, 20, 30, 40, 50, 找到最优方案算法的执行时间如表 2 所示.

表 2 不同种群规模下的算法执行时间

种群规模	算法执行时间(s)
10	11.015
20	11.158
30	9.974
40	11.617
50	11.715

从表 2 可知, 不同种群规模下, 粒子群算法执行效率差异很大, 当种群规模选择 30 时, 算法的执行效率最高, 因此本文的粒子群算法的种群规模为 30.

4.3.2 系统 CPU 负载均衡对比

遗传算法、模糊综合评判算法和本文粒子群算法的 CPU 负载的变异系数曲线如图 2 所示。从图 2 的对比结果可知,基于粒子群算法和遗传算法的 CPU 资源负载平衡度要明显优于模糊综合评判算法,主要是由于粒子群算法和遗传算法采用启发式优化策略,动态考虑了各工作机之间的负载处理能力差异。同时粒子群算法的优化结果更优于遗传算法,对比结果表明,采用粒子群算法可以针对不同类型的任务,对迁移任务进行最优分配,实现多种资源的负载均衡,大幅度提高了系统资源的利用率。

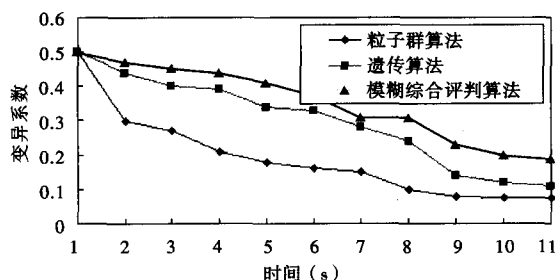


图 2 系统 CPU 负载的变异系数曲线

4.3.3 系统输入/输出负载均衡对比

遗传算法、模糊综合评判算法和本文粒子群算法的输入/输出负载的变异系数曲线如图 3 所示。从图 3 可知,相对于模糊综合评判算法,粒子群算法和遗传算法的输入/输出负载更加均衡,同样粒子群算法的优化方案要优于遗传算法,对比结果再次证表明,采用粒子群算法可以很好实现系统负载均衡,进一步提高了资源的利用率。

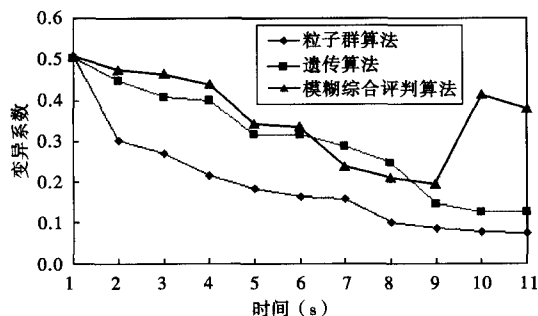


图 3 系统输入/输出负载的变异系数曲线

5 结束语

在工作流系统中,停靠站服务器为每一个工

作机分配最适合的任务,因此需要对服务主体选择进行优化,提高资源利用率。针对工作机负载差异性、动态性,提出一种基于粒子群算法的工作流服务主体优选方法。仿真实验结果表明,该算法充分利用了粒子群优化算法群智能搜索、并行处理等优点,使迁移实例的工作位置达到最优,最大限度保证了系统负载平衡,从而有效提高了系统资源的利用率。

参考文献:

- [1] 牛军钏,胡保月,赵大哲,等. 基于 Internet 支持群组决策的工作流管理系统的设计实现[J]. 小型微型计算机系统, 1999, 20(21): 102-105.
- [2] 李峰,郭玉钊,林宗楷. 工作流管理系统中协同建模技术研究[J]. 计算机辅助设计与图形学学报, 2000, 12(11): 810-812.
- [3] 杨公平,曾广周,卢朝霞. 迁移工作流系统中停靠站服务器的设计与实现[J]. 计算机工程与应用, 2004, 40(29): 111-112.
- [4] 曾广周,党研. 基于移动计算范型的迁移工作流研究[J]. 计算机学报, 2003, 26(10): 1343-1349.
- [5] 杨公平,曾广周,卢朝霞. 迁移工作流系统中的服务主体优选机制研究[J]. 计算机工程与应用, 2004, 40(30): 18-19.
- [6] 吴刚,吴泉源,王怀民. 一种移动智能体的工作流管理模型[J]. 计算机辅助设计与图形学学报, 2001, 13(6): 527-531.
- [7] Kennedy J R. Mendes, Neighborhood topologies in fully informed and best-of-neighborhood particle swarms[J]. Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2006, 36(4): 515-519.
- [8] 刘衍民,赵庆祯,邵增珍. 一种自适应多样性保持的多目标粒子群算法[J]. 济南大学学报:自然科学版, 2011, (3): 296-300.
- [9] 程秋云. 迁移工作流系统中基于遗传算法的服务主体优选机制[D]. 济南:山东大学, 2007.
- [10] 郭娟,杨为民. 基于微粒群算法的二维最大熵图像分割方法[J]. 计算机仿真, 2005, 22(11): 94-97.

作者简介:

韩 猛 男, (1976—), 硕士, 实验师. 研究方向为数据挖掘、实验室建设、Java 应用。