

# 云计算基于遗传粒子群算法的多目标任务调度

刘春燕<sup>1</sup>, 杨巍巍<sup>2</sup>

(1. 武汉理工大学华夏学院 信息工程系, 湖北 武汉 430070;

2. 中国五环工程有限公司 设备室, 湖北 武汉 430070)

**摘 要:**合理地任务调度是云计算长期以来存在的挑战。云任务的调度过程具有动态性的特点, 仅从单一方面来优化调度策略已不能满足用户需求。针对上述问题, 从任务完成时间、任务完成成本、资源利用率三个方面出发, 提出一种基于遗传与粒子群算法融合的多目标任务调度算法。在遗传算法的变异操作中引入粒子群算法, 既可以发挥遗传算法全局搜索能力强的优势, 又可以利用粒子群算法的反馈特性改善变异操作提高收敛速度。通过 CloudSim 平台进行云环境仿真实验, 将此算法与遗传算法(GA)和粒子群算法(PSO)进行比较。实验结果表明, 在相同的条件设置下, 该算法在用户满意度和资源利用率方面都优于遗传算法和粒子群算法, 是一种云计算环境下有效的任务调度算法。

**关键词:**云计算; 任务调度; 多目标; 遗传算法; 粒子群算法

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1673-629X(2017)02-0056-04

doi:10.3969/j.issn.1673-629X.2017.02.013

## A Multi-objective Task Scheduling Based on Genetic and Particle Swarm Optimization Algorithm for Cloud Computing

LIU Chun-yan<sup>1</sup>, YANG Wei-wei<sup>2</sup>

(1. Department of Information Engineering, Huaxia College of Wuhan University of Technology, Wuhan 430070, China;

2. Department of Equipment, Wuhuan Engineering Co., Ltd., Wuhan 430070, China)

**Abstract:** How to schedule tasks reasonably remains a long-standing challenge in cloud computing. The process of the cloud task scheduling has the characteristics of dynamic, so to optimize the scheduling strategy only from a single aspect cannot meet the needs of users. To solve the above problem, from three aspects of task completion time, task completion cost and resource utilization, a multi-objective task scheduling algorithm based on genetic algorithm and particle swarm optimization algorithm is proposed. Particle swarm optimization algorithm is introduced into mutation operation of genetic algorithm which can not only give play to advantage of quick global searching speed for genetic algorithm, but also apply particle swarm optimization algorithm's feedback characteristic to improve mutation operation and convergence rate. CloudSim is adopted to simulate the cloud environment, and the GA and PSO is compared. The simulation results show that under the same conditions, the combined algorithm outperforms other two algorithms on task completion time, task completion cost and resource utilization. It is an efficient task scheduling algorithm in the cloud computing environment.

**Key words:** cloud computing; task scheduling; multi-objective; genetic algorithm; particle swarm optimization algorithm

## 0 引言

近年来,云计算以其独特的方式成为互联网发展的热门话题。云计算主要是利用虚拟化技术,通过网络将庞大的计算处理任务分拆成多个较小的子任务,再交由多部服务器所组成的庞大系统,经搜寻、计算之后将处理结果回传给用户<sup>[1-2]</sup>。

在云计算中,任务调度策略直接影响到用户的任

务执行效率以及云环境下资源的使用效率。当前许多研究者提出将智能算法引入到云任务调度中,收到了一定的成效。其中,李建锋等设计了一种基于双适应度遗传算法的任务调度算法,不但把总任务的完成时间作为调度策略的优化目标,同时还兼顾了任务的平均完成时间,最大限度地提高了云计算的效率<sup>[3]</sup>;封良良等在综合考虑了总任务完成时间和总任务完成成本

收稿日期:2016-04-13

修回日期:2016-08-10

网络出版时间:2017-01-10

基金项目:湖北省教育科研计划指导性项目(B2015373)

作者简介:刘春燕(1983-),女,硕士,讲师,研究方向为云计算、智能优化算法。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170110.1028.066.html>

两个因素的前提下,提出了基于改进粒子群的任务调度算法<sup>[4]</sup>;楼涛等提出了基于混合蚁群遗传算法的Hadoop集群作业调度,一定程度上解决了传统遗传算法的“早熟”问题,加快了收敛速度<sup>[5]</sup>;朱宗斌等提出了改进GA的云计算任务调度算法,综合考虑了任务调用的时间和成本<sup>[6]</sup>。以上各种任务调度算法绝大多数都存在目标单一、收敛早熟等问题。

针对上述问题,文中设计了一种基于遗传粒子群算法的多目标任务调度算法(GA-PSO),从时间、成本、资源利用率三方面均衡考虑,以期达到最小的完成时间兼顾成本最少,同时最大限度保证云环境整体的资源使用效率的目的。该算法的关键是将粒子间信息传递的特性引入遗传算法影响变异因子,避免变异的盲目性。在PSO粒子的作用下,有效解决了传统GA的不足。

## 1 云计算的任务调度分析

云环境下物理资源的服务能力相差甚远,随着云计算的快速发展,云用户对计算资源的需求逐日递增,随之增加的必然是虚拟机数量。当虚拟机数量远大于云环境中的物理资源时,虚拟机需按照一定的调度顺序逐个进行处理,怎样合理分配云计算虚拟机任务,并且极大地满足用户需求是资源分配亟待解决的新问题。虚拟化技术将云环境下差异巨大的物理资源统一整合,在云用户和云计算资源之间搭建共享虚拟资源池。首先将云用户的任务拆分成若干子任务分配给合适的虚拟机,再由虚拟机匹配物理资源,完成虚拟机到云计算资源的映射,以此保证整个用户需求的服务质量最优。结合云计算的结构特点,虚拟机资源调度采用两级模式,实现以用户为中心的虚拟机任务调度策略。第一级调度是划分用户任务找到合适的虚拟资源,第二级调度是将虚拟资源池中的虚拟机匹配给物理资源。两级调度是一个不可分割的整体,在资源分配策略的指导下,根据用户对任务的需求,将任务合理地分配到合适的虚拟机上运行,并结合当前物理资源的负载状况和计算能力,在物理资源中寻找合适的资源分配给虚拟机。通过两级调度,用户的需求得到了很好的满足,而且对物理资源进行了整体统筹,不会造成资源浪费。

文中的研究重点为第二级调度,即通过调度算法将虚拟机上的任务分配到物理机上执行,保证分配的最优性。

## 2 基于遗传粒子群算法的多目标调度

### 2.1 问题描述

云环境下任务调度问题实质是多目标组合优化问

题。目标是建设合理数量的虚拟机,将多个虚拟机分配到物理资源节点上,使得云用户任务的完成时间和成本耗费最低,并兼顾被分配物理资源的服务能力,达到最大的资源利用率。例如,云用户提交任务的预期完成时间为 $p(i)$ ,预期完成成本为 $w(i)$ ,云环境下的物理资源 $A, B, C$ ,它们的服务能力等级分别为高,中,低。当云用户任务被拆分成子任务分配到5个虚拟机上执行时,虚拟机与物理机的配置会存在多种不同的情况,第一种映射: $A \{1, 2, 3, 4, 5\}$ ,第二种映射: $B \{1, 5\}, C \{2, 3, 4\}$ 。第一种映射任务执行时间 $\text{time} < p(i)$ ,执行成本 $\text{cost} \gg w(j)$ ;第二种映射 $\text{time} \approx p(i), \text{cost} < w(j)$ 。很显然,第二种分配方案更能满足用户需求,且最大化了资源利用。

### 2.2 设计思想

针对云计算环境下的任务调度问题,参照Map/Reduce模型<sup>[7-9]</sup>。为了能得到总任务完成效率最优的任务调度结果,充分发挥遗传算法和粒子群算法的优点,将遗传算法与粒子群算法进行有效融合,构造一种遗传粒子群算法。

遗传粒子群算法(GA-PSO)的主要思想是在任务调度的前期阶段,利用遗传算法群体性全局搜索能力强的优势,通过选择、交叉、变异等遗传操作找到任务分配的较优解,在变异操作中引入粒子群算法,更有效地产生接近目标的新个体。

### 2.3 遗传编码

遗传算法是由John Holland教授于1975年首先提出的一类仿生型优化算法,具有并行搜索、群体寻优的特点,但对系统中的反馈信息利用不够,当求解到一定范围时出现大量的冗余迭代,求精确解的效率低下,容易出现局部最优解<sup>[10-14]</sup>。

为解决云环境下的任务调度问题,首先需要将调度方案编码成染色体。文中采用间接编码方式。即对每个任务占用的资源进行编码,染色体的长度等于总的子任务数,每个基因位的位置编号代表子任务的编号,基因位的数值表示所占用的资源编号。

### 2.4 目标函数与适应值函数

目标函数定义为:

$$F(X) = \text{Time} + \text{Cost} + \sum_{i=1}^m s(i) \quad (1)$$

$$\text{Time} = \sum_{j=1}^n \left( \text{avg} \left( \sum_{i=1}^m \text{Time}_i \right) \right)_j \quad (2)$$

$$\text{Cost} = \sum_{j=1}^n \left( \text{avg} \left( \sum_{i=1}^m \text{Cost}_i \right) \right)_j \quad (3)$$

$$s(i) = \sqrt{(r_{\text{time}})^2 + (r_{\text{cost}})^2} \quad (4)$$

其中, $m$ 表示物理资源数量; $n$ 表示总的任务数; $\text{Time}_i$ 表示第 $i$ 个任务通过虚拟机调度在第 $m$ 个物理

资源上执行的时间;  $Cost_i$  表示第  $i$  个任务所使用物理资源上的成本费用;  $s(i)$  表示物理机  $r_i$  的服务能力;  $r_{time}$  表示物理机  $r_i$  的计算能力;  $r_{cost}$  表示物理机  $r_i$  的耗能。

### 2.5 遗传操作

遗传算法的遗传操作主要包括选择、交叉和变异,通过这些操作不断产生新个体,从而搜索出最优解。

#### (1) 选择操作。

根据适应度函数值在种群中按照式(5)计算出每个个体的选择概率。

$$P(i) = \frac{f(i)}{\sum_{j=1}^{SCALE} f(j)} \quad (5)$$

#### (2) 交叉操作。

文中采用自适应交叉方法。首先在个体间按较大交叉概率交换个体间的某些位,为了避免早熟现象的发生,在算法后期,交叉概率相应减小,这样有利于优良新个体的产生,加快了算法的收敛速度。

#### (3) 变异操作。

变异操作可促进基因交流,有利于丰富种群的多样性。但传统的变异是随机的,与历史状态无关,也没考虑到变异算子的优良性,容易造成算法迭代过程中的不稳定性,影响收敛。文中在该阶段引入粒子群算法,利用粒子群体间的信息素,即当前最优解和历史最优解重构变异算子。将遗传算法中得到的历史最优  $f(i)$  所对应的  $x$  替代式(6)中的  $p_{gd}^k$ ,将第  $i$  位上的历史最优  $f(i)$  对应的  $x$  替代式(6)中的  $p_{id}^k$ 。

$$v_{id}^{k+1} = wv_{id}^k + c_1\xi(p_{id}^k - x_{id}^k) + c_2\eta(p_{gd}^k - x_{id}^k) \quad (6)$$

$$x_{id}^{i+1} = x_{id}^k + v_{id}^{k+1} \quad (7)$$

这样的变异操作具备了感应信息的能力,能够根据历史最佳值和个体最佳值影响变异的方向和幅度,提高个体对进化环境的适应能力。

## 3 实验结果及分析

### 3.1 实验仿真环境及参数设置

为了验证 GA-PSO 算法的可行性和有效性,主要从任务调度性能和资源利用率对其进行仿真。实验利用 CloudSim 模拟器构建了云环境任务调度模拟平台,并在相同仿真平台下对文中提出的 GA-PSO 算法、双适应度 GA 算法<sup>[3]</sup>和改进 PSO 算法<sup>[4]</sup>进行任务调度对比分析测试。编写继承 CloudSim 基类的方法。具体实现如下:Host 类可以获取各物理资源的重要信息,VmAllocationPolicy 抽象类用于实现虚拟机的分配,CloudSim 只实现了简单的虚拟机分配策略,通过自定义 VmAllocationPolicyADpso 类实现多目标虚拟机分配,方法 adpdoallocate( $p, v$ ) 是该类的核心,重写该方

法实现云计算虚拟机分配算法。

CloudSim 仿真程序模拟出 30 台物理机,分别部署 30,50,100 个虚拟机,任务数分别设置为 100,200,300,400,500 时进行测试。基于遗传算法与粒子群算法各自的搜索和求解优势,结合反复的实验测试,文中为遗传粒子群算法设置的参数见表 1。

表 1 参数设置

算法	参数	取值
GA	种群规模/个	100
GA	交叉率	0.6
GA	变异率	0.1
PSO	种群规模/个	100
PSO	学习因子 $c_1$	2
PSO	学习因子 $c_2$	2
PSO	惯性因子	0.9

### 3.2 性能分析

根据以上参数进行仿真实验,每组实验仿真 10 次求平均值,得到三种算法资源利用率对比图,如图 1 所示。

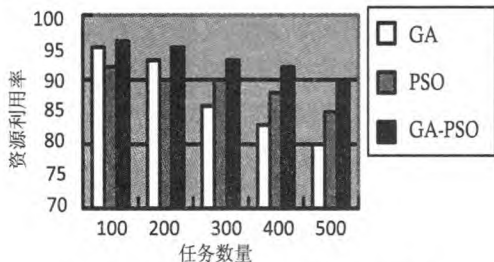


图 1 三种算法资源利用率对比

从图 1 可以看出,当任务数量小于 200 时,云环境中的网络资源充裕,三种算法的资源利用率均达到 90% 以上,GA-PSO 略优于 PSO 和 GA,但差距不明显。随着任务数量的不断增加,基于云资源处理能力的差异,三种算法的资源利用率均有所下降,但 GA-PSO 在任务分配时考虑了云资源的差异性,资源利用率明显优于 GA 和 PSO,性能改进的效果明显。

图 2 和图 3 分别比较了三种算法的任务完成时间和任务耗费成本。

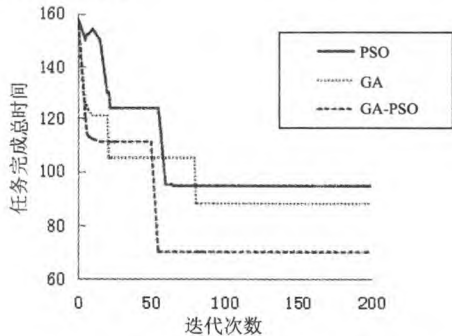


图 2 三种算法任务调度总时间对比

从图2可以看出,GA-PSO算法的任务调度总时间最少,且收敛最快。但在搜索前期优势并不明显,随着迭代次数的增加,在遗传算法中加入PSO的信息素影响变异操作,随着信息素的积累,搜索最优解的速度迅速得到提升。

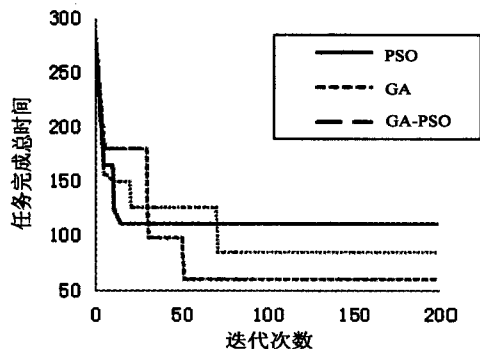


图3 三种算法任务调度时间总费用对比

从图3可以看出,GA-PSO的总费用最少,其次是GA,最高的是PSO。因为双适应度GA以任务完成时间和任务使用费用为双目标;而GA-PSO也同时考虑了时间和费用两方面,并且注重了资源利用率,使得虚拟机在分配任务时会优先选择成本和处理能力较均衡的物理机执行。

#### 4 结束语

文中对云环境下的任务调度问题进行了研究。针对云任务调度中普遍存在的目标单一问题,提出多目标云任务调度,综合考虑任务完成时间、任务完成费用和云资源利用率三个因素,针对遗传算法前期搜索能力强,但迭代过程反馈信息利用不够引起的最优解质量不高的问题,提出了改进方法。在遗传算法的变异操作中引入粒子群算法,利用粒子群体间的反馈信息影响变异因子,避免盲目变异产生新个体,从而加快了算法的收敛速度。CloudSim模拟仿真实验证明,GA-PSO在云任务调度的总体性能上优于GA和PSO。

#### 参考文献:

[1] Abadi D J. Data management in the cloud: limitations and op-

portunities [J]. IEEE Data Engineering Bulletin, 2010, 32 (1): 3-12.

[2] Lee Y C, Zomaya A Y. Energy efficient utilization of resources in cloud computing systems [J]. Journal of Supercomputing, 2012, 60(2): 268-280.

[3] 李建锋, 彭 舰. 云计算环境下基于改进遗传算法的任务调度算法 [J]. 计算机应用, 2011, 31(1): 184-186.

[4] 封良良, 张 陶, 贾振红, 等. 云计算环境下基于改进粒子群的任务调度算法 [J]. 计算机工程, 2013, 39(5): 183-186.

[5] 楼 涛, 杜文才, 钟杰卓. 基于混合蚁群算法的Hadoop集群作业调度 [J]. 海南大学学报: 自然科学版, 2015, 33(4): 340-346.

[6] 朱宗斌, 杜中军. 基于改进GA的云计算任务调度算法 [J]. 计算机工程与应用, 2013, 49(5): 77-80.

[7] Bhandarkar M. MapReduce programming with apache Hadoop, parallel & distributed processing [C] // Proceedings of IEEE international symposium on digital object identifier. [s. l.]: IEEE, 2010.

[8] 尚宏佳, 周 萍, 杨 青, 等. 融合多核和MapReduce的连接聚集查询优化 [J]. 计算机研究与发展, 2015, 52(S): 9-18.

[9] Prabhavathi P, Nithyanandan L. Network selection in wireless heterogeneous networks [C] // Proceedings of 2013 international conference on communications and signal processing. Melmarunathur: [s. n.], 2013: 357-361.

[10] 刘卫宁, 靳洪兵, 刘 波. 基于改进量子遗传算法的云计算资源调度 [J]. 计算机应用, 2013, 33(8): 2151-2153.

[11] 王小平, 曹立明. 遗传算法-理论、应用与软件实现 [M]. 西安: 西安交通大学出版社, 2002.

[12] 倪 霖, 段 超, 贾春兰. 差分进化混合粒子群算法求解项目调度问题 [J]. 计算机应用研究, 2011, 28(4): 1286-1289.

[13] 贾建芳, 杨瑞峰, 王 莉. 混合遗传粒子群优化算法的研究 [J]. 自动化仪表, 2013, 34(9): 1-3.

[14] 靳其兵, 张 建, 权 玲, 等. 基于混合PSO-SQP算法同时实现多变量的结构和参数辨识 [J]. 控制与决策, 2011, 26(9): 1373-1376.