

云计算环境下基于用户和资源约束的免疫效用均衡任务调度算法^①

吴 洲

(广东松山职业技术学院计算机系, 韶关 512126)

摘 要: 针对云计算中的任务调度问题, 提出了一种免疫均衡效用任务调度算法. 该算法将云计算环境下任务调度问题建模为一个多目标优化模型, 同时兼顾了用户任务的时间跨度和虚拟化资源的负载均衡. 仿真结果表明, 该任务调度算法提高了用户满意度的同时减少了任务的完成时间, 是云平台下一种有效的任务调度策略.

关键词: 云计算环境; 任务调度; 时间跨度; 免疫效用任务调度算法

Novel Immune Effectiveness Balancing Task Scheduling Algorithm Based on User and Resources Constraints in Cloud Computing Environment

WU Zhou

(Department of Computer, Guangdong Songshan Polytechnic College, Shaoguan 512126, China)

Abstract: To deal with problems for task scheduling in cloud computing, this paper proposes a immune effectiveness balancing task scheduling algorithm. The proposed algorithm is modeled as a multi-objective optimization model, taking into account the time span of tasks and load balancing of virtualized resources. The simulation result shows that the proposed algorithm is efficient to improve user satisfaction and reduce the completion time of the task.

Key words: cloud computing environments; task scheduling; time span; immune effectiveness balancing task scheduling algorithm

0 引言

近几年, 随着信息技术在各个领域的广泛应用, 对于大规模且低成本的计算能力的需求越来越迫切. 云计算技术就是在这样的背景下应运而生. 云计算作为我国政府重点发展的战略新兴产业, 在国防军工、科技教育、互联网金融等领域已经产生巨大的商业价值和利润. 云计算是分布式计算、网格计算和效用计算的进一步发展, 是以网络为载体, 以虚拟化技术为基础, 根据用户需求动态配置资源的新兴计算模式^[1]. 云计算作为一种新兴的商业计算模型, 该中心应能对用户任务申请的资源进行可靠、安全、合理分配和管理^[2]. 因此, 云计算环境下对动态任务进行有效调度是保证服务质量的关键.

随着云计算技术的蓬勃发展, Google、亚马逊、微

软、IBM 等大型公司都开发了自己的云计算平台, 对云计算的各种相关技术的研究与实现也呈现了百花争艳的态势. 云计算的任务调度机制作为云计算的关键技术之一还处于研究阶段, 到目前为止还没有形成统一的行业标准, 不少研究者提出了一些改进的调度算法. 针对蚁群算法求解云计算任务调度问题存在收敛速度慢和容易陷入局部最优解的缺陷, 王芳等人提出一种动态自适应蚁群算法的云计算任务调度策略^[3]. 该算法在选择资源节点中引入混沌扰乱, 依据节点信息素浓度自适应调整信息素挥发因子, 由解的优劣性动态更新信息素. 左利云等人提出了结合任务优先等级和最小完成时间的改进型 Min-Min 算法, 该算法根据资源的计算能力和通信能力对资源进行划分, 同时兼顾执行时间最小和负载均衡^[4]. Tayal 等人提出了改

① 基金项目: 广东省教育科学“十二五”规划 2012 年度教育信息技术研究项目(12JXN016)

收稿时间: 2015-01-15; 收到修改稿时间: 2015-03-09

善负载均衡的改进蚁群算法, 该算法考虑了虚拟机的计算能力、网络带宽、任务长度等因素, 以降低任务执行时间, 改善负载均衡^[5]. 董丽丽等人提出一种基于差分进化的多策略变异差分进化任务调度算法, 该算法在标准差分进化算法的基础上加入了基于正态分布的分类和多种变异策略, 变异策略中差异向量个体采用基于个体相似度的轮盘赌选择, 改善了标准差分算法收敛速度慢和易陷入局部最优的缺点^[6]. 李健等人利用粒子群算法来解决大规模图处理任务的调度, 并在文中对粒子群的相关操作进行了重新定义, 文中是以调度长度和资源租赁成本为目标^[7]. 张春艳等人提出了一个保证云服务质量的分组多态蚁群算法, 最终减少平均完成时间^[8]. 针对云计算领域的任务调度问题, 杨镜等人提出了一种基于人工免疫理论的云计算平台动态任务调度算法^[9], 该调度算法能有效提高收敛速度和精度, 快速搜索到合理配置, 提高了集群资源利用率. 在文献[10]中, 申丽君等人提出一种基于改进免疫进化算法的任务调度算法, 利用人工免疫进化原理完成任务调度的全局优化. 通过将粒子群优化算法作为算子嵌入免疫进化算法中, 避免陷入局部最优, 改善收敛效果, 减少任务调度时间开销.

当前虽然各种各样基于不同策略和规则的任务调度算法已经被提出, 但是这些任务调度算法都只是从用户的角度来执行资源的分配, 即将最优的资源分配给用户任务, 这样就容易造成资源的不均衡, 并且只考虑用户任务的时间需求. 针对以上问题, 本文从用户和资源提供者两个方面来考虑任务的调度: 对于用户, 从时间跨度的角度考虑任务调度; 对于资源提供者, 从负载均衡的角度考虑资源分配. 因此, 从时间跨度和负载均衡两方面出发, 提出了一个多目标效用函数, 并设计了一种基于免疫算法的任务调度算法, 从而实现所有资源综合效用的提升.

1 基于用户和资源约束的任务调度模型

云计算任务调度是指在特定的云计算环境中, 根据一定的资源使用规则将用户提交的任务合理分配到虚拟化资源中进行分配和管理.

1.1 资源的属性描述与性能分析

设 $VR = \{r_1, r_2, \dots, r_j, \dots, r_m\}$ 表示云计算系统中的虚拟化资源集合, r_j 表示第 j 个虚拟化资源, $m = |R|$

表示虚拟化资源的规模. r_j 定义为一个四元组 $r_j = \{r_{jd}, r_{jc}, r_{jt}\}$, 其中, r_{jd} 用于标识资源 ID; r_{jc} 表示资源 r_j 单位时间内的计算能力, 如单位时间内资源执行指令数; $r_{jt} = (r_{jtr}, r_{jtb}, r_{jtf})$ 表示资源 r_j 的状态, r_{jtr} 表明资源可接收任务, r_{jtb} 表明资源处于繁忙状态, 负载达到饱和阈值, r_{jtf} 表明资源处于失效状态. 判断资源所处的状态, 可以根据负载均衡因子 LB 来衡量, 当 $LB > \varphi$ 表明资源处于繁忙状态, 暂不接收任务; 当 $0 < LB < \varphi$ 表明资源处于空闲状态, 可接收任务; 当 LB 长期等于 0, 表明资源处于失效状态, 暂不接收任务. 负载均衡因子描述如下:

定义 1. 如果用 La 表示完成的任务量, Lf 表示分配到的任务计算量之和, 则负载均衡因子 LB 为:

$$LB = La / Lf \quad (1)$$

1.2 任务的属性描述与性能分析

设用户提交的任务描述为集合 $T = \{t_1, t_2, \dots, t_i, \dots, t_n\}$, t_i 表示第 i 个任务, $n = |T|$ 表示任务的数量. t_i 定义为一个四元组 $t_i = \{t_{id}, t_{ia}, t_{iw}, t_{ie}\}$, 其中, t_{id} 用于标识任务 ID, t_{ia} 表示任务单位时间内到达的平均数量, t_{iw} 表示任务 t_i 的工作负载(计算量), t_{ie} 表示任务 t_i 对资源的期望, 包括资源的计算性能、存储性能、IO 性能、带宽等等. m 个任务在 n 个虚拟化资源上对应的期望执行时间 ETC 矩阵如下:

$$ETC(n, m) = \begin{bmatrix} ETC_{00} & \cdots & ETC_{0m-1} \\ & \ddots & \\ ETC_{n-10} & \cdots & ETC_{n-1m-1} \end{bmatrix}$$

1.3 任务调度优化模型

云计算任务调度的目标是建立任务与资源之间的映射, 即 $\Omega: T \rightarrow VR$, $\Omega(i, j) = 1$ 表明任务 t_i 分配到资源 r_j 上, 否则 $\Omega(i, j) = 0$. 根据任务计算量及资源计算能力, 任务 t_i 在资源 r_j 上执行所需时间 $Ex(i, j)$ 为:

$$Ex(i, j) = \frac{t_{iw}}{r_{jc}} \quad (2)$$

资源 r_j 上所有任务执行时间总和为:

$$T_j = \sum_{i=1}^m Ex(i, j) \times \Omega(i, j) \quad (3)$$

资源 r_j 的均衡度为:

$$\varepsilon = \sum_{i=1}^m \frac{|Ex(i, j) - ETC(i, j)|}{Ex(i, j)} \times \Omega(i, j) \quad (4)$$

云计算环境下任务调度的时间跨度如下:

$$Ms = \max \{T_j | t_i \in T, r_j \in VR, i=1, 2, \dots, m, j=1, 2, \dots, n\} \quad (5)$$

当 Ms 值越小, 表明任务调度的方案越理想; 当 $\varepsilon \rightarrow 0$, 表明云计算环境下的资源综合效用越高. 因此, 云计算环境下任务调度的目标如下:

$$\begin{cases} Ms = \max \{T_j | t_i \in T, r_j \in VR, i=1, 2, \dots, m, j=1, 2, \dots, n\} \\ \varepsilon = \sum_{i=1}^m \frac{|Ex(i, j) - ETC(i, j)|}{Ex(i, j)} \times \Omega(i, j) \\ \min Ms \\ \min \varepsilon \end{cases} \quad (6)$$

2 免疫调度算法设计

免疫算法(Immune Algorithm, IA)是受生物免疫系统信息处理方式的启发构造的一种仿生智能计算方法, 免疫算法以父代最优个体为基础产生子代群体, 并以最优个体的收敛来代替群体的收敛, 具有自适应性和分布性, 非常适合解决云计算的任务调度问题. 本文在深入研究免疫算法的基础上, 将云计算环境下任务调度问题建模为一个多目标优化模型, 提出了一种免疫均衡效用任务调度 (Immune Effectiveness Balancing Task Scheduling, IEBTS) 算法, 算法兼顾了用户任务的时间跨度和虚拟化资源的负载均衡, 充分提高云计算环境下虚拟化资源的利用率.

2.1 编码方式

本算法将 m 个虚拟化资源分别编为 $0 \sim m-1$, n 个用户提交的任务编号为 $0 \sim n-1$, 并采用实数编码的方式. 通过建立长度为 n 的数组 $X[i]$, 将数组下标 i 表示第 i 个任务, $X[i]$ 表示第 i 个任务在第 $X[i]$ 个虚拟化资源上执行, 这样产生的一组实数字符串就代表一种调度方案. 采用随机的方式初始化抗体种群, 设抗体种群的规模为 S , 则生成 S 个长度为 n 的数组, 数组的每个元素随机地取 $0 \sim m-1$ 的整数.

2.2 亲和度函数

令 x_g^i 表示种群进化到第 g 代时的第 i 个抗体,

则种群中每个抗体的亲和度为 $aff(x_g^i)$ 可以表示为:

$$aff(x_g^i) = e^{\frac{1}{Ms + \varepsilon}} \quad (7)$$

其中, $0 < \varepsilon < 1$, Ms 为任务调度的时间跨度.

通过公式(7)可以看出: 当任务调度的时间跨度越大, 则对应的抗体亲和度值越小.

2.3 算法的操作

IEBTS算法主要包含三大操作: 克隆, 变异, 选择. 其中, 克隆操作采用传统的免疫算法克隆方式, 即将亲和度较大的抗体根据一定的克隆规模进行复制, 这样促使优良的抗体保留到下一代, 有效维持抗体种群的优良性.

令种群亲和度均值为 ν , 种群亲和度的均方差表示为 σ . IEBTS算法的变异操作定义如下:

$$x_g^{i'} = x_g^i + \eta \times \frac{aff(x_g^i)}{\sum aff(x_g^i)} \times N(\nu, \sigma) \quad (8)$$

其中, x_g^i 表示父代抗体的基因, $x_g^{i'}$ 为变异后子代抗体的基因, M 为种群的规模, $N(\nu, \sigma)$ 为均值 ν , 方差为 σ 的随机高斯分布, $\eta(0 < \eta < 1)$ 是控制变异操作的衰减参数, $aff(x_g^i)$ 是抗体 x_g^i 的亲和度值.

IEBTS 算法将抗体群划分为主流抗体群和非主流抗体群. 抗体亲和度值大于平均亲和度值的抗体归纳为主流抗体群, 否则该抗体属于非主流抗体群. IEBTS 算法的选择操作就是将主流抗体群中的抗体选择出来, 直接保留到下一代抗体群.

2.4 IEBTS 算法的执行步骤

输入: 抗体种群的规模 S , 变异概率 P_m , 最大进化代数 G_m , 进化代数 $g = 0$

Begin

- Step 1. 随机产生初始抗体群 $A(g)$;
- Step 2. 计算抗体群 $A(g)$ 中每个抗体的亲和度值;
- Step 3. 对抗体群进行克隆操作;
- Step 4. 将抗体群 $A(g)$ 划分为主流抗体群 $A_l(g)$ 和非主流抗体群 $A_r(g)$;
- Step 5. 对抗体群进行变异操作;
- Step 6. 对抗体群进行选择操作;
- Step 7. $g = g + 1$, 重复 Steps 2-6 直到满足 $g > G_m$.

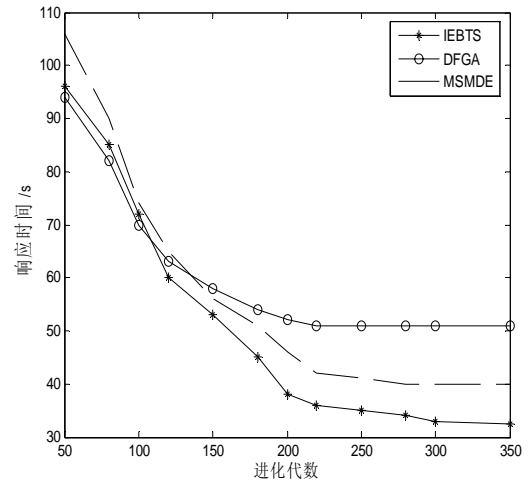
End

输出: 最优的任务调度方案

3 实验测试与结果分析

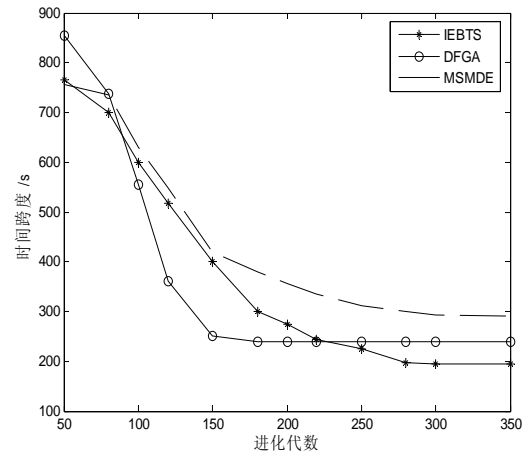
为了检验本文设计的 IEBTS 算法的性能, 我们利用云计算仿真平台 CloudSim 3.0 进行性能测试, 通过 CloudSim 中的 DataCenter、Cloudlet、VM 及一些辅助类模拟云计算的计算和网络资源、创建任务和虚拟机, 构建云计算的局部仿真环境, 并通过在 DataCenter Broker 类中构造不同的任务调度算法, 包括本文设计的 IEBTS 算法, 文献[2]的 DFGA 算法, 文献[6]的 MSMDE 算法. Cloudlet 模拟了云应用服务, 每一个应用服务都会拥有一个预分配的指令长度和其他生命周期内所需要的数据传输开销. VM 是模拟由主机组件托管和管理的虚拟机. VmSheduler 由一个主机组件实现, 模拟为虚拟机分配处理核所用的空间共享和时间共享策略.

在下面的实验过程中, 本文进行了三类对比测试: 响应时间、时间跨度、虚拟化资源均衡度. 其中, 任务个数和虚拟化资源规模分别表示为 m 和 n . 图 1 分别显示的是当 $m = 50, n = 10$ 和 $m = 100, n = 15$ 时, 任务调度的响应时间比较; 图 2 分别显示的是当 $m = 50, n = 10$ 和 $m = 100, n = 15$ 时, 任务调度的时间跨度比较; 图 3 分别显示的是当 $m = 50$ 和 $m = 100$ 时, 10 个虚拟化资源的均衡度比较情况.

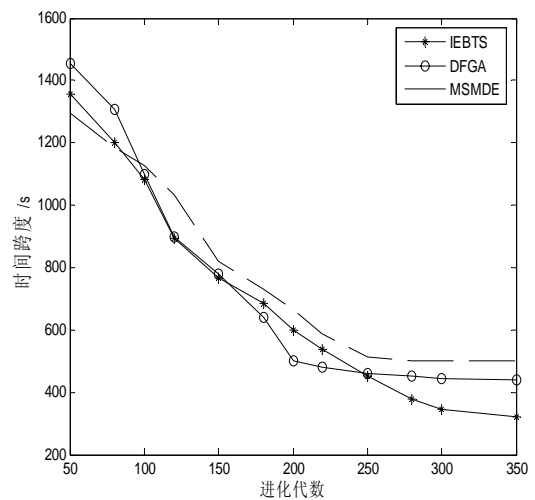


(b) $m = 100, n = 15$

图 1 响应时间比较

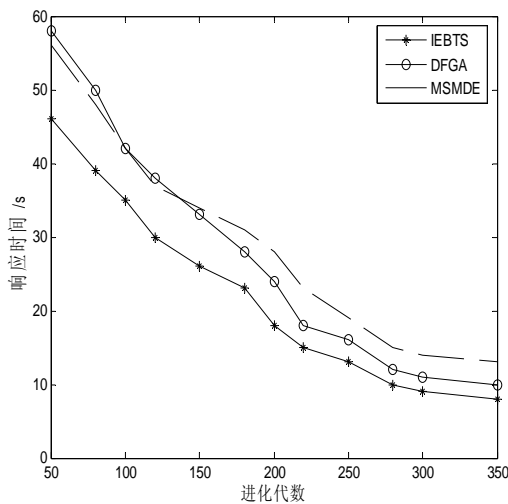


(a) $m = 50, n = 10$



(b) $m = 100, n = 15$

图 2 时间跨度比较



(a) $m = 50, n = 10$

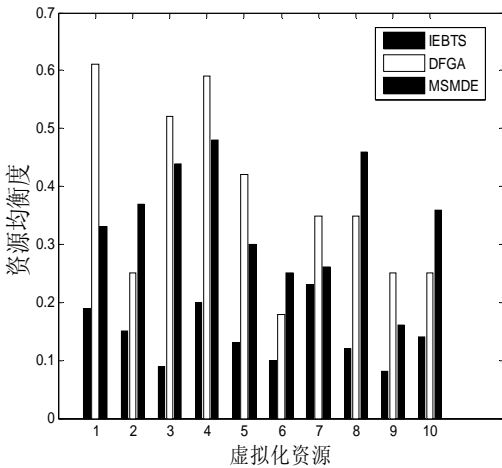
从图 1-2 的实验结果中不难看出：在进化前期，本文设计的 IEBTS 算法优势表现并不明显；在进化后期，随着 IEBTS 算法的变异操作和选择逐渐深入，该算法寻优能力明显增强，并且不容易陷入局部最优。在实验 3 中，IEBTS 算法的资源均衡度最小值接近 0.08，最大值只有 0.23，而 DFGA 算法的最小值是 0.18，MSMDE 算法的最小值是 0.19，在所有情况下 IEBTS 算法的均衡度值 ϵ 都最小，表明 MSMDE 算法能保证云计算环境下的资源综合效用的高效性。

4 结论

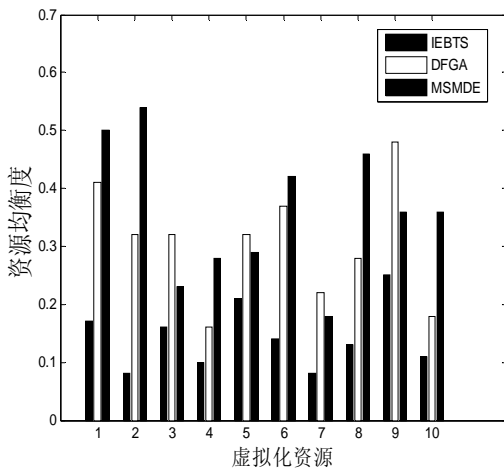
随着云计算技术与应用的迅速发展，云系统规模越来越大，拓扑结构越来越复杂，加上资源的异构性使得如何进行高效云任务调度成为云计算领域中的重要研究课题。本文设计的基于用户和资源约束的 IEBTS 算法能有效提高用户满意度，降低任务执行的时间跨度，是云平台下一种有效的任务调度策略。

参考文献

- 1 许丞,刘洪,谭良.Hadoop 云平台的一种新的任务调度和监控机制.计算机科学,2013,40(1):112-117.
- 2 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.计算机应用,2011,31(1):184-186.
- 3 王芳,李美安,段卫军.基于动态自适应蚁群算法的云计算任务调度.计算机应用,2013,33(11):3160-3162, 3196.
- 4 左利云,左利锋.云计算中基于预先分类的调度优化算法.计算机工程与设计,2012,33(4):1357-1361.
- 5 Tayal S. Tasks scheduling optimization for the cloud computing system. International Journal of Advanced Engineering Sciences and Technologies, 2011, 5(2): 11-15.
- 6 董丽丽,黄贲,介军.云计算中基于差分进化算法的任务调度研究.计算机工程与应用,2014,50(5):90-95.
- 7 李健,黄庆佳,刘一阳,等.云计算环境下基于粒子群优化的大规模图处理任务调度算法.西安交通大学学报,2012,46(12):116-121.
- 8 张春艳,刘清林,孟珂.基于蚁群优化算法的云计算任务分配.计算机应用,2012,32(5):1418-1420.
- 9 杨镜,吴磊,武德安,等.云平台下动态任务调度人工免疫算法.计算机应用,2014,34(2):351-356.
- 10 申丽君,刘丽,陆锐.基于改进免疫进化算法的云计算任务调度.计算工程,2012,38(9):208-210.



(a) $m = 50, n = 10$



(b) $m = 100, n = 10$

图 3 虚拟化资源的均衡度比较