

基于改进免疫遗传算法的网格任务调度

张京军,刘文娟,刘光远

(河北工程大学 信息与电气工程学院,河北 邯郸 056038)

摘要:为改进网格计算中任务调度的低效问题,采用十进制的实数编码规则产生初始抗体群,由免疫遗传算法经过克隆和变异算子生成资源集中的蚁群信息素,进而利用蚁群算法的并行性展开全局搜索,通过 CloudSim 仿真平台进行模拟,与粒子群算法及蚁群遗传算法进行对比,结果表明,改进的免疫遗传算法能够大幅提高网格计算任务调度效率,有效地解决网格任务调度问题。

关键词:网格计算;免疫遗传算法;任务调度;蚁群算法

中图分类号:TP311

文献标识码:A

Task scheduling in grid computing based on improved immune genetic algorithm

ZHANG Jing-jun, LIU Wen-juan, LIU Guang-yuan

(School of Information and Electrical Engineering, Hebei University of Engineering, Hebei Handan 056038, China)

Abstract: For improving the inefficient of the current grid computing, the decimal real number encoding rules was chosen to generate the initial antibody group. Firstly, the immune genetic algorithm was used to generate the initial pheromone distribution in the collection, then the parallelism of the ant colony algorithm was used for global search, finally the particle swarm optimization (PSO) and ant colony genetic algorithm were compared by using CloudSim as a simulation platform to simulate. The results indicate that the improved immune genetic algorithm can provide efficient task scheduling strategy and it can solve the problem more effectively.

Key words: grid computing; immune genetic algorithm; task scheduling; colony optimization;

网格任务调度已被证明是一个典型的 NP-hard 问题^[1],近几年,启发式任务调度算法引起了国内外学者的广泛关注。文献[2]提出一种基于粒子群网格任务调度算法,利用多约束线性规划对任务调度进行建模,通过迭代获得各计算节点的最佳任务调度数;文献[3]基于蚁群遗传算法进行网格任务调度,使用基本遗传算法生成初始解种群,结合蚂蚁算法求出全局最优解;文献[4]应用分层遗传算法实现对网格任务调度策略的优化设计,使用柯布-道格拉斯效益模型来降低计算的复杂度,最大化系统的总效用。然而,上述算法在解决任务调度时通常存在一定的自适应性,难以避免局部最小值的问题,无法有效地满足网格

任务调度的要求。人工免疫系统是一种软计算方法^[5-6],利用免疫算法的多机制求解多目标函数的自适应特性,在很大程度上避免了“早熟现象”。基于此,本文结合免疫进化理论对网格任务调度策略进行优化,通过优化免疫算子等进化参数来实现最优跨度。

1 任务调度问题的定义

网格任务调度的实质是将 n 个相互独立的任务分配到 m 个异构可用资源上,使得任务总体完成时间最少。本文以简单的网格任务调度为例,只考虑各个子任务之间的约束关系,子任务的计算时间和资源分配策略,不存在通信和同步要求。

在算法执行之前,任务信息、资源信息以及任务在相应资源上的执行时间都是已知的。假定 G 表示一个网格系统,由 $G = \{T, P, E\}$ 这样的三元组进行定义:

定义1 任务集合 $T = \{t_1, t_2, \dots, t_n\}$, 资源节点集合 $P = \{P_1, P_2, \dots, P_m\}$;

定义2 任务的估计执行时间用一个 $n \times m$ 矩阵 $E(t_i, p_j)$ 表示, T 对 P 的所有资源分配所形成的矩阵 E 表示为 $E_{n \times m} = E(t_i, p_j); 1 \leq i \leq n, 1 \leq j \leq m$ 。

用 $begin(P_j)$ 表示任务 t_i 被派送到计算资源 P_j 执行的开始时间, $finish(t_i, P_j)$ 表示任务 t_i 在计算资源 P_j 上的完成时间,则有:

$$finish(t_i, P_j) = begin(P_j) + E(t_i, P_j) \quad (1)$$

设 $C_A P(j)$ 表示在调度策略 A 下,计算资源 P_j 完成分配到该资源上的最后一个子任务所需的时间。若 $C(A)$ 表示在调度策略 A 下任务完成所花费的总时间,则

$$C(A) = \max(C_A P(j), \forall 1 < i < m) \quad (2)$$

网格任务调度要解决的问题就是找到一个调度策略 A ,使得 $C(A)$ 最小。

2 任务调度算法的优化

2.1 免疫遗传算法

本文中,把网格任务调度优化问题抽象为抗原,对应的每一个抗体就是一种可行的调度方案。首先,根据网格任务调度问题的相关参数 G 构造算法的第一个染色体,并据此生产初始群体,而抗原对应于调度问题的目标函数,抗体对应于调度问题的候选解,算法的核心在于通过记录克隆变异后所形成的最优个体信息来代替整个群体的进化,以减少网格计算任务调度总时间。

免疫遗传算法的具体实现过程如下:

步骤1:随机生成群体规模为 M 的初始父代群体 $x_i^{[7-8]}$;

步骤2:根据先验知识抽取疫苗,构造优化目标函数;选择当前代中的最优解,确定初始最优个体 $x_{i,best}^0$ 及 M 个初始群体分量 i 分量的标准差取值 σ_i^0 ;

步骤3:判断当前种群中是否已包含最佳个体,若是则算法结束,否则进行以下步骤;

步骤4:父代种群选取 i 个分量进行复制,在解空间内生成种群规模仍为 M 的新的子代群体,

复制操作按照公式(3)进行。

$$x_i^{t+1} = x_{i,best}^t + \sigma_i^t N(0, 1) \quad (3)$$

式中: x_i^{t+1} - 子代个体第 i 个分量; $x_{i,best}^t$ - 父代个体第 i 个分量; $\sigma_i^{t+1} = \sigma_i^t \exp\left(\frac{At}{T}\right)$ - 父代群体第 i 个分量的标准差,分母 T - 总进化代数,分子 A - 标准差动态调整系数, t - 当前进化代数; $N(0, 1)$ - 服从标准正态分布的随机数;

步骤5:计算由公式(3)进化生成的 M 个子代群体中每个个体的适应值,确定最优个体 $x_{i,best}^{t+1}$, 比较与父代最优个体的适应值,若满足 $F(x_{i,best}^{t+1}) \geq F(x_{i,best}^t)$, 则选择 $x_{i,best}^{t+1}$ 为最优个体,否则,最优个体仍保留父代的 $x_{i,best}^t$;

步骤6:迭代的执行步骤3-5,直到满足终止条件,输出最后一代的最优解。

2.2 改进免疫遗传算法实现过程

蚁群算法(Ant Colony Optimization)其原理是通过蚂蚁群体之间的正反馈信息传递机制而达到寻优目的。将蚁群算法作为一个算子加入到免疫遗传算法中,利用蚁群算法的并行性以及求解高效性等特征进行全局搜索,避免了单纯的免疫遗传算法系统中反馈信息没有利用而导致的冗余迭代。算法调度流程如图1所示。

算法的具体过程描述如下:

步骤1:确定参数,初始化种群规模 N , 变异概率 P_m 和最大迭代次数 G_{max} 。

步骤2:初始化抗体种群,在解空间中随机产生 N 个抗体构成初始种群 $P(0) = \{a_1, a_2, \dots, a_n\}$ 。

其中,每个初始抗体对应于一个满足网格任务调度约束关系的可行调度方案,每个任务选择相应的资源节点映射,选择的原则是对该任务有较快执行速度的机器有较大的机会被选中。由于二进制编码会产生问题所不能接受的等位基因,本文采用十进制的实数编码,抗体为 m 位,每个基因位取值范围为 $1 \sim n$,表示为:

$$A_b = \{b_1, b_2, \dots, b_m\}; 1 \leq b_j \leq n \quad (4)$$

步骤3:对第 k 代种群, $0 \leq k_j \leq G_{max}$, 计算抗体群中每个抗体的亲和度 $f_i, i = 1, 2, \dots, N$, 并按亲和度值升序进行排列。

步骤4:对第 k 代种群 P_k 的前 z 个抗体实施克隆变异得到新抗体群 A_c , 再次计算亲和度并排序,选出 A_c 中最好的抗体作为蚁群算法的初始信

息素,设计的亲和度函数为:

$$f(i) = 1 / \sum_{j=1}^m total(b_j^p) \quad (5)$$

式中: b_j^p - 每个抗体的适应度值。

步骤5:运用蚁群算子对步骤4中所得到的抗体再进行处理,得到一组优化的新抗体,蚁群算子的路径信息素更新规则如下:

$$\tau_{ij}(t+1) = \rho \cdot \pi_{ij}(t) + \sum \Delta \tau_{ij}^k(t) \quad (6)$$

式中: $\Delta \tau_{ij}^k(t)$ - 蚂蚁 k 在路径 (i, j) 上留下的单位长度轨迹信息素数量; ρ - 信息素的持久性 ($0 \leq \rho \leq 1$); $\pi_{ij}(t)$ - 计算资源的信息素浓度。

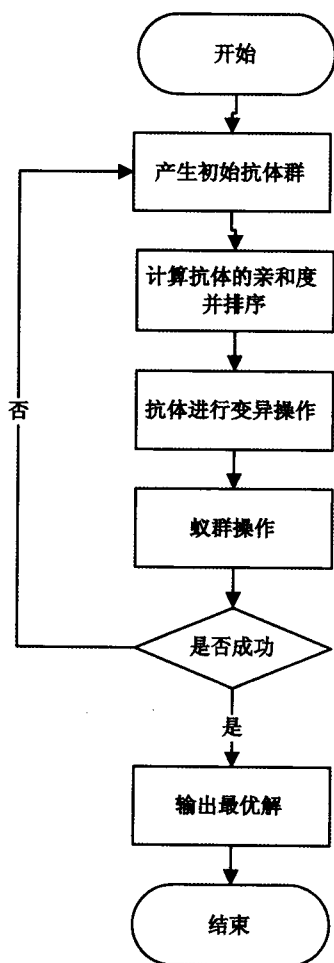


图1 算法调度流程

Fig.1 The algorithm scheduling process

步骤6:对比蚁群算子得到的优化新抗体与初始抗体种群及克隆变异后的抗体,记录亲和度大的那组染色体。

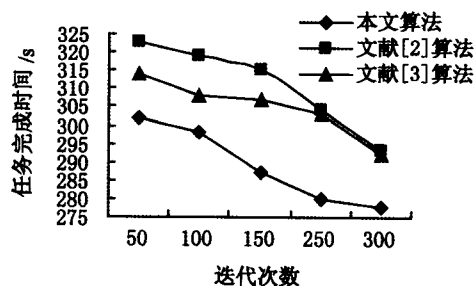
步骤7:当 $k \geq G_{\max}$ 时,进化过程结束,输出最优解;否则,置 $k \leftarrow k + 1$ 返回步骤3。

3 实例测试

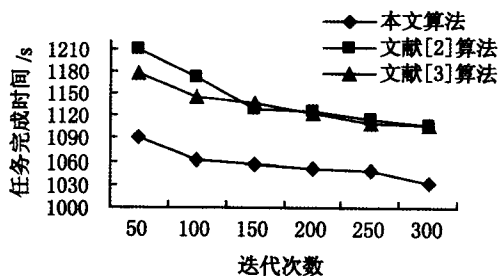
本文采用网格计算仿真平台 CloudSim 进行模

拟,实验过程中,分别采用粒子群任务调度算法^[2],蚁群遗传算法^[3],改进的加入蚁群算子的免疫遗传算法完成不同任务规模的任务调度。网格实验环境中,算法相关参数设置如下:相应的杂交概率设置为 $P_c = 0.7$,变异概率 $P_m = 0.15$,终止条件设为进化300代。

调度实例1:资源数 $m = 3$,任务数 $n = 8$, $t = \{100, 90, 90, 42, 41, 30, 20, 7\}$;实例2:资源数 $m = 5$,任务数 $n = 40$, $t = \{42, 72, 15, 27, 89, 88, 55, 61, 83, 72, 63, 4, 16, 91, 43, 71, 26, 31, 90, 83, 77, 73, 6, 38, 13, 27, 72, 11, 10, 4, 96, 17, 62, 62, 24, 84, 12, 32, 46, 69\}$ 。



(a) 任务完成时间图 ($m=3, n=8$)



(b) 任务完成时间图 ($m=5, n=40$)

图2 不同算法下任务执行总时间和收敛性比较

Fig.2 The comparison of task execution total time and convergence under the different algorithms

由图2不同任务数下算法运行的最终结果可以看出,当任务规模较小时,使用粒子群算法或蚁群遗传算法做任务调度,算法的最优解为290 s,整体执行时间较长,其稳定性也相对较差,而采用改进的免疫混合算法,最优解可以达到275 s,算法收敛效果较好;当任务规模较大时,采用改进的免疫遗传算法,当算法迭代到第120代左右时,已经达到了稳定的收敛状态,同时最优解明显高于其他两类算法,表明该算法在进行任务调度时能够有效地降低总的任务执行时间,而且收敛效果也比粒子群算法或单纯的蚁群遗传混合算法好,全局相比之下,改进的免疫算法效果更理想,能够

更好的优化网格任务调度。

4 结论

基于生物免疫机理的改进遗传算法,能够保证获得解的全局最优性,有效地解决网格任务调度问题。

参考文献:

- [1] 王浩,李飞. 基于 QoS 约束的网格任务调度算法[J]. 四川理工学院学报:自然科学版, 2013, 26(1): 25-28.
- [2] 刘瑶,陈志刚,曾志文. 可信禁忌粒子群优化网格任务调度算法[J]. 计算机工程与应用, 2009, 45(12): 86-89.
- [3] 王刚,钟志水,黄永青. 基于蚁群遗传算法的网格资源调度研究[J]. 计算机仿真, 2009, 26(4): 240-243.

- [4] 刘海迪,杨裔,马生峰. 基于分层遗传算法的网格任务调度策略[J]. 计算机研究与发展, 2008, 45(1): 35-39.
- [5] 何丽丽,张京军. 改进遗传算法在桁架拓扑优化中的应用[J]. 河北工程大学学报:自然科学版, 2009, 26(3): 19-21.
- [6] CASTROL L N, TIMMIS J. Artificial immune systems as a novel soft computing paradigm [J]. Soft Computing 2003,7(8): 536-544.
- [7] 胡广青,姜波,吴胡. 基于灰色关联-遗传神经网络的煤与瓦斯突出预测模型[J]. 中国煤炭地质, 2011, 23(9): 27-31.
- [8] 窦贤明,杨永国,徐伟伟,等. 基于遗传算法和 BP 神经网络的矿井涌水量预测[J]. 中国煤炭地质, 2009, 21(10): 69-71.

(责任编辑 马立)

(上接第 79 页)

根据以上数据分析,数值模拟结果稍微低于现场观测值,说明预测区煤层顶板抗压能力较实验室测试值偏小,现场实测钻孔不足以控制工作面,基于此模拟预测值与现场实测值存在一定偏差,偏差率约 14%。RFPA^{2D}数值模拟上覆岩层冒落和裂隙带总高度结果基本接近实际观测结果,勘查区煤层顶板冒裂带在没有现场实测条件下,采用理论计算基础上加 14% 可以作为煤层冒裂带高度值。

4 结论

1) RFPA^{2D}数值模在处理综放矿造成的覆岩破坏这类问题时快速、简捷,对不同煤岩层赋存状态均可灵活加以运用。另外,通过与采空工作面实测和经验公式估算对比、检验,考虑与实测值的偏差率,可以作为其他采区冒裂带高度值,说明这种方法的预测结果是可靠的,因此,此法为综放开采条件下两带高度预测提供了一种有效的工具。

2) 煤层开采冒裂带发育高度与上覆松散层的

岩性力学性质和煤层开采厚度相关,顶板强塑性岩层分布越厚,煤层开采厚度越薄,将会对导水裂隙带高度起到限制作用,对安全高效开采有利。

参考文献:

- [1] 缪协兴,刘卫群,陈占清. 采动岩体渗流理论[M]. 北京:科学出版社,2004.
- [2] 朱万成,唐春安,杨天鸿,等. 岩石破裂过程分析 RFPA^{2D}系统的细观单元本构关系及验证[J]. 岩石力学与工程学报,2003,22(1):24-24.
- [3] 中国煤炭地质总局水文地质工程地质环境地质勘查院. 任家庄煤矿 21 采区水文地质补充勘探报告[R]. 2012,6.
- [4] 桂和荣,廖多荪. 综放开采最大导水裂隙带高度的应力法预测[J]. 煤炭学报,1997,22(4):375-379.
- [5] MTT_865-2000,导水裂缝带高度的钻孔冲洗液漏失量观测方法[S].
- [6] 武强. 煤矿防治水规定释义[M]. 北京:中国矿业大学出版社,2009.

(责任编辑 刘存英)