

文斌,邵明安.黄土高原小流域水循环 GIS 设计与实现[J].农业工程学报,2005,21(10)

周明耀,冯小忠,夏继红.基于组件式 GIS 的管道输水灌溉系统规划设计软件研制[J].农业工程学报,2006,22(4)

苏中滨,孟繁疆,康丽,等.基于 Agent 技术虚拟植物模型的研究与探索[J].农业工程学报,2005,21(8)

MultiGen-Paradigm Inc. Vega Prime Desktop Tutor Version 2.0[M]. MultiGen-Paradigm Inc., 2005

#### 作者简介

雷军,副研究员,研究领域:城市与区域发展.联系方式:新疆乌鲁木齐市北京南路 818 号中国科学院新疆生态与地理研究所,邮编:830011.电话:0991-7885356,传真:0991-7885380. E-mail: leijun@ms.xjb.ac.cn.

## 数字油田云计算负载均衡的多智能体遗传算法解决方案初探

程国建 刘丽景 刘天时

(西安石油大学计算机学院 西安 710065)

**摘 要** 将云计算引入数字油田是新技术发展的必然趋势,它有助于油气公司快速存取、组织及可视化勘探、开发及生产的相关数据,从而进行有效决策,降低企业运营成本。由于数字油田云计算系统具有大规模、异构、分布、自治的特点,其负载均衡更加复杂和难以有效实现。本文采用多智能体遗传算法(Multi-agent Genetic Algorithm, MAGA),通过参数选择训练建模方式来解决数字油田云计算中的负载均衡问题。设计了相关的 MAGA 算法流程,并进行了实例测试,试验结果表明在数字油田云系统中采用 MAGA 构建的负载均衡模型,能够有效避免传统负载均衡算法中由于过度依赖某一个参数指标而造成的“单点失效”问题。另外,该算法能够实时的根据任务的变化而动态更新,故它较传统方法更能适应不断变化的云系统,这对复杂的油气企业云计算平台的建立是一个有效的尝试。

**关键词** 云计算 数字油田云 负载均衡 多智能体遗传算法

### 1 引言

云计算指的是一种超大规模网络计算、虚拟化架构、资源可扩展、相对廉价的服务交付和使用模式,用户可以通过网络按各自所需获得服务。油田信息的处理和存储经历了一个从集中到分散再到集中的过程,云计算就是信息存储和处理的工业化过程,是信息服务业的基础设施和服务平台。将云计算引入数字油田是新技术发展的必然趋势。严格讲,数字油田云计算即是将计算分布在由大量计算机构成的资源池上,各种应用系统根据需要获取计算力、存储空间和各种软件服务。而数字油田云计算系统的性能由于资源竞争和任务不确定性等因素而高度动态变化,因此资源调度和任务管理是数字油田云计算系统中的一个关键问题。现有的算法如静态的轮询法、随机法由于容易造成单点失效,显然不能适用于如此大规模的数字油田云系统中的负载均衡。动态的算法包括网格负载均衡中常用的 UT 策略、Spring 策略也仅仅只是考虑到了单个参数的影响,并且它们的执行效率会随着数字油田云规模的不断增大而下降。

本文的目的就是利用 MAGA (Multi-agent Genetic Algorithm) 来解决大规模数字油田云计算中的负载均衡问题。通过设定多种不同的仿真环境,在多种试验环境中用 MAGA 算法进行参数选择训练,从而得到的相对较优的、较通用的相关参数值,进而得到数字油田云负载均衡策略。通过试验,将 MAGA 生成的负载策略与 UT 策略、Spring 策略应用于同一仿真环境中,通过比较三种策略运行中的最大 CPU 使用率、平均 CPU 使用率、最大内存使用率、平均内存使用率、任务执行成功率来分析它们的优缺点。试验表明,对于数字油田云这种大规模的动态云系统, MAGA 生成的负载策略所达到的 CPU 最大使用率和最大内存使用率更低,更能保证整个系统的稳定性和健壮性。

## 2 云计算中负载均衡模型的建立

由于向数字油田云系统中所提交的任务类型是多种多样的,对于每种不同类型的任务,云系统中最适合的处理机的选择方式也是不同的。我们可以通过建立如下模型来进行负载均衡处理:

$$\begin{cases} S = \frac{\sum_{i=1}^n \min\{t(C_{1i}kx_{ij} + C_{2i}x_{cpi} + C_{3i}x_{mi} + C_{4i}x_{mri} + C_{5i}x_{di} + C_{6i}x_{dij})\}}{n} \\ C_{1i} + C_{2i} + C_{3i} + C_{4i} + C_{5i} + C_{6i} = 1 \quad (i = 1, 2, \dots, n; i \text{ 代表第 } i \text{ 种任务}) \end{cases} \quad (1)$$

其中:

$j$ ——表示当前所选择的处理机,  $j$  的大小为当前集群系统中处理机的个数;

$S$ ——系统处理任务平均所用时间;

$k$ ——执行第  $i$  种任务时,当前机器的  $cpu$  个数。

$\min\{t(C_{1i}kx_{ij} + C_{2i}x_{cpi} + C_{3i}x_{mi} + C_{4i}x_{mri} + C_{5i}x_{di} + C_{6i}x_{dij})\}$ ——当前集群系统中执行第  $i$  种任务所需的最小时间,其中  $x_{ij}$  为执行任务时,当前选择的机器的  $cpu$  速率;  $x_{cpi}$  为执行任务时,当前选择的机器的  $cpu$  利用率;  $x_{mi}$  为执行任务时,当前选择的机器的内存大小;  $x_{mri}$  为执行任务时,当前选择的机器的内存剩余大小;  $x_{di}$  为执行任务时,当前选择的机器的硬盘剩余大小(以 G 为单位);  $x_{dij}$  为从负载均衡器发送任务到当前选择的机器上,所花费的通信时间。 $C_{1i}, C_{2i}, C_{3i}, C_{4i}, C_{5i}, C_{6i}$  分别为其相应项的系数,代表了执行此类任务时该项指标的影响能力。

由于传统策略中仅考虑到了单个指标的影响,往往无可避免的会产生“单点失效”现象,使得任务丢失,并且失效的机器长时间内无法使用,影响整个云系统的正常运行。本文所提出的负载均衡建模策略综合考虑了多种指标的影响,并用  $C$  来表示每项指标对执行任务的依赖程度。另外,对于不同类型的任务,  $C$  的取值也会有很大差异。因此本文按照任务对指标的依赖程度的不同进行分类,通过 MAGA 算法为每种类型的任务分别求得一组较优的  $C$  值。

## 3 用于数字油田云计算系统中参数估计的多智能体遗传算法

从多智能体的角度出发,把遗传算法中的个体看做是一个具有感知、竞争协作和自学习能力的智能体,通过智能体与环境以及智能体间的相互作用达到全局优化的目的。这是多智能体遗传算法(MAGA)的根本思想。基于这一思想, MAGA 将智能体对环境的感知和反作用的能力与遗传算法的搜索方式相结合,使得其收敛速度远远高于传统算法。

对数字油田云计算系统中的负载均衡的建模最终可以转化为一个求解函数最小值的线性规划问题。在任务种类增多时,线性函数的维数成倍增长。因此采用 MAGA 建模数字油田云计算中的负载均衡是可行的。

基于上述模型(1),不同类型的任务,所对应的  $C$  值各不相同。首先定义一个智能体  $a$  表示待优化函数(1)的一个候选解,  $a$  即为满足条件的  $C$  值:

$$Energy(a) = -t(a) \quad (2)$$

令所有智能体生存在一个网格环境中,这个网格环境被称为智能体网格,记为  $T$ 。网格大小为  $T_{size} \times T_{size}$ ,其中  $T_{size}$  为整数。每个智能体固定于一个格点  $T_{ij}$  上,每个智能体只能和相连的邻域发生相互作用,如图1所示:

在这个智能体网格中,每个智能体都与其相邻的智能体展开竞争或合作,以便能够获得更大的能量。一个智能体与其邻域发生作用后便将其信息传给了它们,在这种方式下,信息逐渐扩散到整个智能体网格。

随后设计4个智能体遗传算子,邻域竞争算子、邻域正交交叉算子、变异算子和自学习算子。其中邻域竞争算子和邻域正交交叉算子分别实现智能体的竞争和合作行为;变异算子和自学习算子实现了智能体利用自身知识的行为。

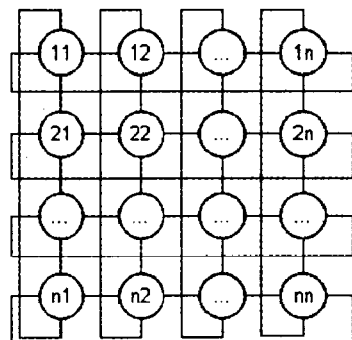


图1 智能体网格

(a)邻域竞争算子:

若格点  $T_{ij}$  满足下式,则为胜者;否则,为败者。

$$Energy(T_{ij}) > Energy(T_{ij}^{max}) \quad (3)$$

若  $T_{ij}$  为胜者,则可继续存活在网格上;否则,必须死亡,空出的节点将被  $T_{ij}^{max}$  所占据。

(b)邻域正交交叉算子:

邻域正交交叉算子将正交交叉算子作用在  $T_{ij}$  与  $T_{ij}^{max}$  确定的搜索空间上。利用正交矩阵  $T_{M_2}(Q_2^T) = [b_{ij}]_{M_2 \times F}$  产生  $M_2$  个智能体,然后利用  $M_2$  个智能体中最大的一个替代  $T_{ij}$ 。该算子中没有参数调整。考虑到该算子计算复杂度,智能体会依概率执行正交交叉操作。

(c)变异算子的设计:

产生一个新的智能体  $T_{ij}^{new} = (e_1, e_2, \dots, e_n)$ , 其中

$$e_k = \begin{cases} l_k, & U(0,1) < 1/n \\ l_k + G(0,1/t), & \text{否则} \end{cases} \quad (4)$$

其中:  $t$  是进化代数,  $G(1,1/t)$  是高斯分布的随机数。然后用  $T_{ij}^{new}$  代替  $T_{ij}$ 。

(d)自学习算子的设计:

在自学习算子中产生一个智能体网格  $sL$ , 其大小为  $sT_{size} \times sT_{size}$ , 其上的所有智能体  $sT_{i,j}$  均由下式产生:

$$sT_{ij} = \begin{cases} l_{ij}, & i' = \text{land}' = 1 \\ sT_{ij}^{new}, & \text{otherwise} \end{cases} \quad (4)$$

其中:  $sT_{ij}^{new} = (e_{ij1}, e_{ij2}, \dots, e_{ijn})$

#### 4 算法步骤

在本次试验中各相关参数设置如下:

智能体网格大小  $L_{size} = 5$ ;

邻域竞争算子中策略选择概率  $P_0 = 0.3$ ;

邻域交叉算子执行概率  $P_c = 0.1$ ;

变异算子执行概率  $P_m = 0.1$ ;

终止条件为达到的最大迭代代数,本次试验取最大代数为1000代。

另外,用  $T^i$  表示第  $i$  代的智能体网格,  $T^{i+1/3}$  和  $T^{i+2/3}$  是  $T^i$  和  $T^{i+1}$  间的中间代智能体网格。  $Best(i)$  是  $T^0$ ,

$T^1, \dots, T^i$  中最优的智能体。 $CBest(i)$  是第  $i$  代中最优的智能体。

- Step 1: 初始化  $T^0$ , 更新  $Best(0)$ , 令  $i \leftarrow 0$ ;
- Step 2: 对  $T^i$  中的每个智能体执行邻域竞争算子, 得到  $T^{i+1/3}$ ;
- Step 3: 对  $T^{i+1/3}$  中的每个智能体, 若  $U(0,1) < P_c$ , 则将邻域正交叉算子作用在其上, 得到  $T^{i+2/3}$ ;
- Step 4: 对  $T^{i+2/3}$  中的每个智能体, 若  $U(0,1) < P_m$ , 则将变异算子作用在其上, 得到  $T^{i+1}$ ;
- Step 5: 从  $T^{i+1}$  中找出  $CBest(i)$ , 并将自学习算子作用在其上;
- Step 6: 如果  $Energy(CBest(i)) > Energy(Best(i))$ , 则令  $Best(i+1) \leftarrow CBest(i+1)$ ; 否则令  $Best(i+1) \leftarrow Best(i)$ ,  $CBest(i+1) \leftarrow Best(i)$ ;
- Step 7: 如果终止条件满足, 输出  $Best(i)$  并停止; 否则, 令  $i \leftarrow i+1$ , 并转 Step 2。

5 仿真实验与分析

仿真实验分为 4 组, 即定义了 4 种不同的实验环境。所用的硬件环境完全相同。在每组实验中, 对 UT 策略、Spring 策略、MAGA 策略这三种方式所构建的数字油田云计算负载模型提交完全相同的任务。三种策略在系统执行任务时的最大 CPU 使用率、平均 CPU 使用率、最大内存使用率、平均内存使用率以及任务执行成功率 5 个方面进行了比较。实验结果如下表 1~表 4 所示:

表 1 100 个任务并发提交至数字油田云系统(10 种不同类型的任务)

负载策略	最大 CPU 使用率(%)	平均 CPU 使用率(%)	最大内存使用率(%)	平均内存使用率(%)	任务执行成功率(%)
UT	11	9.76	34.74	24.31.	100
Spring	23	11.37	24.79	23.43	100
MAGA	14	11.39	27.75	23.77	100

表 2 500 个任务异步提交至数字油田云系统(10 种不同类型的任务)

负载策略	最大 CPU 使用率(%)	平均 CPU 使用率(%)	最大内存使用率(%)	平均内存使用率(%)	任务执行成功率(%)
UT	21	15.67	37.38	29.43	100
Spring	38	16.52	28.13	27.99	100
MAGA	27	15.94	32.39	29.03	100

表 3 1000 个任务异步提交至数字油田云系统(20 种不同类型的任务)

负载策略	最大 CPU 使用率(%)	平均 CPU 使用率(%)	最大内存使用率(%)	平均内存使用率(%)	任务执行成功率(%)
UT	55	37.46	79.45	40.41	100
Spring	96	47.91	53.17	49.17	99.8
MAGA	58	37.22	54.44	47.73	100

表 4 5000 个任务异步提交至数字油田云系统(20 种不同类型的任务)

负载策略	最大 CPU 使用率(%)	平均 CPU 使用率(%)	最大内存使用率(%)	平均内存使用率(%)	任务执行成功率(%)
UT	78	74.61	91.76	79.33	99.98
Spring	100	89.57	78.74	69.03	85.06
MAGA	79	73.93	77.93	68.82	100

由上面 4 组实验结果可以看出,在任务量较小时,三种策略并无太大差异。但随着任务数量及类型的增加,特别是在任务增加到 1000 个时,UT 策略由于着重于考虑 CPU 的使用率,而未考虑到内存使用情况,从而使得系统内存花销不均衡;Spring 策略正好相反,它仅考虑了内存的使用情况,而未考虑 CPU 使用情况,造成了 2 个单点失效状况的发生。在任务增加到 5000 时,UT 策略和 Spring 策略均发生了单点失效的情况;而 MAGA 策略由于兼顾了 CPU、内存以及其他的一些指标的影响,故在数据量较大且任务种类较多时仍能保持整个系统的稳定运行。

## 6 结论

本文提出了使用 MAGA 来构建数字油田云计算中的负载均衡模型的方法。数字油田云系统的规模一般较大,而 MAGA 在解决函数优化方面具有较快的收敛速度,特别是在较大数据集上,依然保持良好的性能。通过对 UT 策略、Spring 策略、MAGA 策略生成的负载模型进行比较分析,可以发现,用 MAGA 生成的负载均衡模型较 UT 策略、Spring 策略生成的负载均衡模型更能够保证数字油田云系统的稳定性和健壮性。特别是在提交的任务数量较大且并发任务较大的情况下,MAGA 生成的负载均衡模型明显优于其他两种算法,而且这种模型不易使系统发生“单点失效”。由此说明本方法用于构建数字油田云计算负载均衡策略模型是可行及有效的。

## 致谢

本文得到国家自然科学基金项目(40872087)的支持。

## 参考文献

- 王鹏. 走近云计算[M]. 北京:人民邮电出版社,2009
- 许骏. 面向服务的网格计算[M]. 北京:科学出版社,2009
- 焦李成,刘静,钟伟才. 协同进化计算与多智能体系统[M]. 北京:科学出版社,2006
- Buyya R, Yeo CS, Venugopal S. Market - oriented. cloud computing: Vision, hype, and reality for delivering it services as computing utilities [C]. Keynote at the 10th IEEE International Conference on High Performance Computing and Communications (HPCC - 08). IEEE, 2008
- TAKAGIH. Interactive evolutionary computation: fusion of the capabilities of EC optimization and human evaluation [J]. Proceedings of the IEEE, 2001, 89(9)
- 陆青,梁昌勇,黄永青,张俊岭. 面向旅游行程规划的交互式多智能体遗传算法[J]. 计算机应用研究, 2008, 25(11)
- 张伟丰,郑建国. 混合多智能体遗传算法[J]. 计算机工程与应用, 2010, 46(3)
- 孟红云,刘三阳. 求解多目标优化问题的多智能体遗传算法[J]. 西北大学学报(自然科学版), 2005, 35(1)

## 作者简介

程国建,博士,教授,研究方向为机器学习、模式识别、计算智能,工程应用领域包括智能油藏工程、生物

特征识别、商务智能、图像处理等。

# 基于三网融合的统一多屏信息化平台

吕雪花 杨永刚

(上海贝尔股份有限公司 上海 201206)

新疆油田公司通讯公司 克拉玛依 834000)

**摘 要** 三网融合趋势下的统一多屏信息化平台,解决了目前油田多个网络之间的资源共享和互动协作问题,如利用手机作为灵活机动的控制屏,高效便捷的实时连接和控制油田企业的各类通信和生产设备,用于提高企业的生产能力,并同时提供高性能的视频互动的方案和应用。本文对统一多屏信息化平台如何构建数字智慧油田进行了必要性和可行性的分析。最后,本文还如何推进多屏信息化平台和建立数字智慧油田提出了建议。

**关键词** 三网融合 多屏 统一信息化平台

## 1 三网融合和多屏时代

### 1.1 三网融合的背景和意义

2010年1月13日,国务院总理温家宝主持召开国务院常务会议,决定加快推进电信网、广播电视网和互联网三网融合。所谓“三网融合”,是一种广义的、社会化的说法,在现阶段它并不意味着电信网、计算机网和有线电视网三大网络的物理合一,而主要是指业务应用的融合。三大网络通过技术改造,能够提供包括语音、数据、图像等综合多媒体的通信业务,通过一张网络可以完成以往分别电信网、广播电视网和互联网的日常信息处理,包括:

- 信息服务将由单一业务转向文字、语音、数据、图像、视频等多媒体综合业务;
- 有利于极大地减少基础建设投入,并简化网络管理,降低维护成本;
- 将使网络从各自独立的专业网络向综合性网络转变,网络性能得以提升,资源利用进一步提高;
- 三网融合是业务的整合,它不仅继承了原有的话音、数据和视频业务,而且通过网络的整合,衍生出了更加丰富的增值业务类型,如图文电视、VOIP、视频邮件和网络游戏等,极大地拓展了业务提供的范围。

### 1.2 多屏的定义和特点

“多屏”是指以用户为核心,以业务体验为导向,随时随地使用任何设备(电视、PC、手机以及其他移动设备)访问多媒体内容服务,并获得一致的用户体验,实现“多屏”通信、“多屏”电视、“多屏”内容共享、“多屏”推送等功能。

简单地讲,通过多屏业务,用户可以将原来独立的几项活动串联起来。例如,在家里通过IPTV订购某段视频后,可以先在电视上收看,然后在PC上接着看,如果没看完,出门后可以通过手机观看剩余的节目。用户是感受不到设备、网络之间的切换,以及重复认证的。多屏业务允许设备之间的无缝切换,无需用户多次登录认证。此外,系统还可以记忆用户的喜好与习惯,针对用户的设备,主动对内容进行适配。并可以根据用户偏好和场景,调整用户观看的体验。

在独立业务环境里,上述过程只能通过各个终端来实现,即用户一旦注册某服务,整个用户体验仅限于该终端环境,切换到其他终端是无法使用的。但在多屏业务环境里,上述用户体验可以在任意终端里进行。特别重要的是,多屏业务不仅仅是可以在任意终端观看视频,而且可以有其他体验,例如恢复视频内容。