

网格资源的组织与发现研究*

杨清平^{1,2}, 邓小清¹, 蒲国林^{1,2}, 邱玉辉²

(1. 四川文理学院 计算机科学系, 四川 达州 635000; 2. 西南大学 语义网格实验室, 重庆 400715)

摘要:分析了网格资源的组织形式,阐述了蚁群算法的基本原理,提出了一种基于蚁群算法的网格资源发现方法。其中,将用户请求本体看作蚂蚁,查找的资源即搜索的目标视为食物,食物源就是存在搜索目标的节点。蚂蚁寻找食物的过程就是网格资源的发现过程。

关键词: 网格; 蚁群算法; 资源发现

中图分类号: TP181

文献标识码: A

文章编号: 1672-6693(2008)04-0070-04

网格是构筑在互联网上的一组新兴技术,它将高速互联网、高性能计算机、大型数据库、传感器、远端设备等融为一体,并进行了信息的融合,使人们能够透明地使用资源的整体能力,并能按需获取所有信息。网格的主要任务是在动态变化的网络环境中共享资源和协同解决问题。网格技术最终目的是把整个网络整合成一台巨大的超级计算机,实现 Internet 上计算资源、信息资源、知识资源等所有资源的连通、共享和互操作^[1-3]。网格技术的核心是网格资源管理,而网格资源发现则是网格资源管理中的基本组成部分,它为其他网格资源管理提供满足需求的可用资源集^[4]。

1 蚁群算法

蚁群算法(ACO, Ant Colony Optimization),又称蚂蚁算法,首先出自意大利学者 Marco Dorigo 的博士论文中。其灵感来源于自然界中真实蚁群在寻找食物过程中发现路径的行为。

1.1 真实蚂蚁的觅食行为

1) 活动范围。蚂蚁观察到的范围是一个方格世界,蚂蚁有一个参数为速度半径,那么它能观察到的范围就是以半径为单位的方格里,并且能移动的距离也在这个范围之内。

2) 活动环境。一只蚂蚁所在的环境中有障碍物、信息素和其它蚂蚁。信息素有两种,一种是找到食物的蚂蚁洒下的食物信息素,另一种是找到蚁巢的蚂蚁洒下的蚁巢信息素。每个蚂蚁都仅仅能感知

其方格世界内的环境信息,并且信息素以一定的速率消失。

3) 觅食规则。每只蚂蚁在自己的方格世界里感知。首先感知是否有食物,如果有就直接过去。然后感知是否有信息素,如果有就朝信息素多的地方走,并且每只蚂蚁多会以小概率犯错误,从而并不总是往信息素最多的点移动。蚂蚁找寻蚁巢的规则和上面一样,只不过它是对蚁巢信息素做出反应。

4) 移动规则。每只蚂蚁都朝信息素最多的方向移动,当周围没有信息素指引的时候,它会按照自己原来运动的方向惯性地运动下去。为了防止原地转圈,蚂蚁会记住并尽量避开最近刚走过的那些点。

5) 避障规则。如果蚂蚁要移动的方向有障碍物挡住,它会随机地选择另一个方向,并且有信息素指引的话,它会按照觅食的规则行为。

6) 播撒信息素规则。每只蚂蚁在刚找到食物或者蚁巢的时候散发的信息素最多,并随着它走远的距离,播撒的信息素越来越少。

从规则可知,蚂蚁之间并没有直接的关系,每只蚂蚁都只和环境发生交互,是信息素把各个蚂蚁联系起来。

1.2 人工蚁群系统

人工蚁群算法是在自然界中真实的蚁群行为的基础上提出的一种模拟进化算法。为了区别于真实蚂蚁群体系统,称这种算法为“人工蚁群算法”。

人工蚁群与真实蚁群间既有共性,又有差异。其共性如下:蚂蚁间的合作,任务都是找寻连接起点

* 收稿日期: 2008-03-25

资助项目: 西南大学研究生科技创新基金(No. 2006011)

作者简介: 杨清平(1954-) 男, 副教授, 研究方向为人工智能、服务计算。通讯作者: 邱玉辉, E-mail: yhqiu@swu.edu.cn.

和终点的最短路径,通过信息素进行间接通讯;人工蚁群也利用了蚂蚁觅食中的正反馈的特点进行信息素的更新。另一方面,人工蚁群又有一些真实蚂蚁不具备的特性:人工蚁之间是离散的,它们的移动是从一个状态跃迁到另一个状态;人工蚁拥有一个内部的状态,这个私有的状态记忆了蚂蚁过去的行为;人工蚁释放一定量的信息素,它是蚂蚁所建立的问题解决方案优劣程度的函数;人工蚁群释放信息素的时间可以根据实际情况而定,而真实蚂蚁是在移动的同时释放信息素;人工蚁可以在建立了一个可行的解决方案之后再行进行信息素的更新;人工蚁群被赋予了一些智能化的特性,如前瞻性、局部优化、原路返回等。

1.3 蚁群算法基本原理

蚁群算法基本原理如图1。

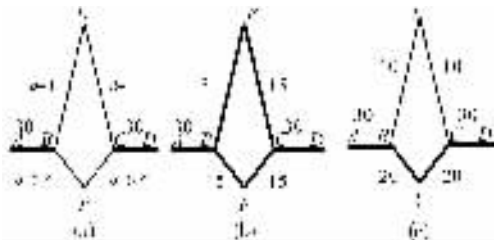


图1 蚁群算法基本原理

设A为蚁穴,D为食物源,EF为一障碍物。由于障碍物的存在,蚂蚁只能经由A经E或F到达D,或由D到达A,各点之间的距离如图1(a)所示。假设每个时间单位有30只蚂蚁由A到达D点,有30只蚂蚁由D到达A点,蚂蚁过后留下信息量为1。在初始时刻,由于路径AF、FC、BE、EC上均无信息存在,位于A和D的蚂蚁可以随机选择路径,从统计学角度可以认为蚂蚁以相同的概率选择BF、FC、BE、EC,如图1(b)所示。经过一个时间单位后,在路径BFC上的信息量是路径BEC上信息量的2倍。又经过一段时间,将有20只蚂蚁由B、F和C点到达K,如图1(c)所示。如此继续下去,沿线路BFC上移动的蚂蚁越来越多,最终将完全选择路径BFC。这就是巢穴到食源的最短路线,蚂蚁根据线路上留下信息素浓度大小,确定在路线上移动的方向,蚁群向信息素浓度重的线路集聚的现象称为正反馈。蚂蚁算法正是基于正反馈原理的启发式算法^[5]。

2 网格资源组织

网格资源是指所有能够通过网格远程使用的实体,包括计算机软件、计算机硬件、设备和仪器等。

它具有异构性、动态性、自治性等特点。

2.1 资源的组织形式

资源组织指资源在网络中如何分布和存储。资源信息组织受到很多因素的影响。文献[6]等提出网格资源信息组织分域管理模型,本文中是按资源属性分类,并为每一个资源分类建立一个域。如图2所示,网格资源模型定义为无向图,边表示信息素。

根据资源的分类描述,对于网格中的每一类资源,注册该类资源的信息节点形成一个连通的覆盖网络,并将每类资源交由相应资源管理域管理,形成若干的资源管理域。每个功能节点都至少拥有一个同属分类邻节点,这样资源信息就可以尽量在同属分类的节点间扩散,从而改善资源发现以及资源信息更新的性能。

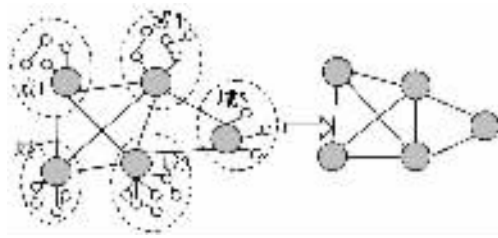


图2 域资源组织

2.2 资源的加入

节点的加入算法:文献[7]中提出了基于P2P资源发现的改进方案。在该方案中,根据某种兴趣关系聚合重要的节点形成覆盖网络来有效组织P2P网络中的资源。文献[8]在文献[7]的基础上提出改进的资源动态自组织算法,该算法的目标是:减小查找范围,减小冗余消息的数量,使具有相似兴趣的结点尽可能地链接在一起,形成网格资源虚拟组织。本文借鉴文献[8]中的资源自组织算法,自动形成网格资源域。

算法如下。

HostPeerSelection_Algorithm(HostPeer p)

For all my hostpeers

Find indirect hostpeer s with maximum
 $RQueryHits_p(s)$

Find immediate hostpeer q with least
 $RQueryHits_p(q)$

If $RQueryHits_p(s) \geq RQueryHits_p(q)$

Send a direct connection from hostpeer p to host-peer s

If $numConnection_p \geq$

$MAX_CONNECTIONS_p$ Then remove

hostpeer q with least $RQueryHits_p(q)$

其中 $RQueryHits_p(q)$ 表示结点 p 从邻接结点 q 收到的查询命中消息数与从所有结点收到的查询命中消息数的比率; $numConnection_p$ 表示结点 p 当前链接数; $MAX_CONNECTIONS_p$ 表示结点 p 的最大链接数。

2.3 资源的更新

1) 资源节点主动更新。如果资源节点客户程序在脱机情况下对资源服务的状态作了调整,则在资源节点接入网络的时候,会向注册中心发出状态更改请求,一旦重新介入网络,资源节点就会判断是否现在的资源描述和注册的描述存在区别,如果存在区别就会向注册中心发出更改状态及信息请求。

2) 注册中心被动更新。当资源节点没能正常关机或断网等,没有正确地改变资源服务的状态。在这种情况下注册中心会定时向资源节点发出状态询问请求,如果没有得到客户端的应答,就会将资源服务的状态改变。

3 基于蚁群算法的网格资源发现

资源发现是根据资源请求者的资源请求描述,从网格上为请求者找到满足描述要求的合适资源的过程。资源发现可以形式化地描述为一个函数 $f: S_{reg} \rightarrow S_{id}$ 。该函数的输入变量是资源需求描述,输出变量是一个资源的唯一标识符。函数 f 是资源发现功能的表示^[9],资源发现就是要构造一个实现函数 f 的系统。

本文主要利用蚁群算法进行域间的资源发现。

3.1 设计思想

其主要思想是:将用户请求本体看作蚂蚁,查找的资源即搜索的目标视为食物,存在搜索目标的节点就是食物源。源节点发出搜索请求,就相当于派出蚂蚁到网络中寻找食物,根据一定的规则即转发策略进行转发。网格中的节点指的是域服务器节点,每个节点都维护一张邻接点表和信息素表,当蚂蚁到达时,先看节点是否有需要的食物也就是在域内搜索资源的过程,如果有就返回一个命中消息包,可看作派出一只找到食物的蚂蚁沿原路返回源节点,沿途释放信息素,即修改节点的信息素表。负责搜索食物的蚂蚁,根据节点中的信息素浓度,选择离目标近的邻居继续爬行。

3.2 算法分析

3.2.1 关键数据结构 1) 邻接点表(AT) 在每个域服务器节点中用 AT 来存放与其他的域服务器节点的邻接关系 2) 访问节点表(VNT) :VNT 就是蚂蚁

在每访问一个节点就将访问的节点记录下来,存放在每只蚂蚁上,如 $VNT_k(s)$ 表示第 k 只蚂蚁已访问过的 s 节点; β) 信息素 (τ_{ij}): 信息素就是蚂蚁在运动时在路径中释放出一种特殊的分泌物,蚂蚁就是通过它来寻找路径。当它们碰到一个还没有走过的路口时,就随机地挑选一条路径前行,同时释放出与路径长度相关的信息素。

3.2.2 网格资源信息素 蚁群算法为一个通用算法,而在不同领域的信息素的影响因素不同,本文分析了网格资源发现领域,提出了适合网格资源发现的信息素。1) 信息素的初始化:在新资源加入的时候要附带资源的固有信息包括:数据传输时间 s_{ij} ,传输的正确率 e ;在初始化信息素时还要考虑带宽 b 、响应时间 c 、带宽利用率 p 等。所以边 (i, j) 上初始的信息素为 $\gamma_{ij} = bpce/s_{ij}$; γ_{ij} 为资源的固有信息;在资源加入时要初始化信息素: $\tau_{ij} = \gamma_{ij}$ 。2) 信息素的更新:在资源发现过程中,当蚂蚁访问一个节点就更新信息素 $\tau_{ij}(t+n) = \tau_{ij} + (1-\rho)(sim_{ij} + \Delta\tau_{ij}(t))$;其中 $sim_{ij} = sim(A, B)$ A, B 分别为资源请求本体和域服务器资源本体匹配度; ρ 为信息挥发系数, ρ 越小则挥发越快, $1-\rho$ 表示信息素残留因子, ρ 的取值范围为 $\rho \in [0, 1)$; $\Delta\tau_{ij}(t) = \sum_{k=1}^m \Delta\tau_{ij}^k(t)$ $\Delta\tau_{ij}^k(t)$ 表示边 (i, j) 上信息素的增量; $\Delta\tau_{ij}^k(t)$ 表示第 k 只蚂蚁访问过 (i, j) 边留下的信息素; $\Delta\tau_{ij}^k(t) = Q/d_{ij}$; Q 为常数; d_{ij} 为 (i, j) 间的距离。

3.2.3 转发策略(VS) 转发策略就是蚂蚁如何选择下一个访问节点的策略。蚂蚁服务器在派发蚂蚁的时候,先根据服务器节点的信息素来决定转发策略。转发策略为:访问 AT 找到当前节点的邻接点,然后在所有的邻接点中根据转发概率 $P_{ij}^k = \tau_{ij}^\alpha(t) \cdot \eta_{is}^\beta(t) / \sum_{s \in VNT} \tau_{is}^\alpha \cdot \eta_{ij}^\beta(t)$ 来进行选择。其中 $\eta_{ij}(t)$ 为启发函数,在本文中记为常数 C 。本文中的移动策略如下式

$$P_{ij} = C \frac{(\tau_{ij})^\beta}{\sum_{j \in VNT(i)} (\tau_{ij})^\beta}$$

$\beta > 0$ 说明该条路上的信息素最多,走过的蚂蚁就越多,而蚂蚁再次选择此条路径的概率就越大。即 $\beta > 0$ 选择倾向于快速的链路。在本文中 $\beta < 0$,即蚂蚁总是选择较为陌生的路径,从而达到快速搜索的效果。

3.3 算法设计

定义标志变量 S_{reg} , $S_{reg} = 0$ 代表最低资源发现, $S_{reg} = 1$ 代表贪婪资源发现。算法伪代码如下。

step1 系统初始化 ;

step1.1 将 τ_{ij} 初始化 ,其中初始值和节点间的通信能力相关 ;

step1.2 将 VNT 置空 ;

step2 将 k 只蚂蚁随机转发在 n 个节点上 ;

step3 根据蚂蚁所在点的本地目录匹配 ,根据蚂蚁内部消息判断 ,若是最低资源发现 ,则转回生成蚂蚁的节点 ;

step4 更新 VNT τ_{ij} goto step8 ;

step5 :每只蚂蚁遍历节点 ;

step6 :访问 VNT ,判断是否遍历所有的节点 ,若是 ,则 goto step7 ;

step6.1 蚂蚁根据 VS 选择下一个访问节点 ;

step6.2 蚂蚁移动到所选择的节点 ;goto step3

step7 根据蚂蚁的一次遍历比较 ,查找到近似最优满足用户需要的资源 ;

step8 返回资源地址或者资源备份给用户 ;结束。

4 结束语

本文用域来组织网格资源 ,并且通过域节点的加入和更新操作使动态资源保持最新。对于域间资源的发现则使用改进的群算法 ,尽量避免重复的搜索过程 ,实现资源的快速搜索。

Research into Organization and Discovery of Grid Resource

YANG Qing-ping^{1,2} , DENG Xiao-qing¹ , PU Guo-lin^{1,2} , QIU Yu-hui²

(1. Dept. of Computer Science , Sichuan University of Arts and Science , Dazhou 635000 ;

2. China Semantic Grid Laboratory of Southwest University , Chongqing 400715 , China)

Abstract : The core of grid technique is grid resources management , and the grid resources discovery is the basic constitute part in it. The grid resources discovery provides a valid set to meet the demands for other grid resources managements. This paper uses the organized form of domain resources , which classify resources through the resources attribute , and build up a domain for each resources classification. We research the joining and renewing of domain resources and the discovery of the resources with ant colony optimization. We treat the customer request ontology as ant , the searching resources as the food and the node which includes searching target is a food source. The source node which sends out a search request equals to send ants to look for food in grid network , and transmits according to the certain rule , namely transmitting strategies. What the node in the grid is mentioned is a domain server node. Each node supports an abutment nodes form and an information hormone form. When an ant arrives , it sees if the node first has the food of demand. This is also the process which searches resources in the domain. The ant which is responsible for searching food crawls toward the neighbor node which is the nearest to the target according to the information hormone density in the node.

Key words : grid ; Ant Colony optimization ; resources discovery

参考文献 :

- [1] 旭龙. 网格资源监控及其资源描述的研究[D]. 兰州理工大学 , 2003.
- [2] 马燕 , 邱玉辉 , 李明勇. 一种基于移动 Agent 的网格资源发现机制研究[J]. 重庆师范大学学报(自然科学版) , 2007 , 24(2) : 18-20.
- [3] 应宏. 网格系统的组成与体系结构分析[J]. 西南师范大学学报(自然科学版) 2004 , 29(4) : 586-591.
- [4] CZAJKOWSKIY K , FITZGERALDZ S , FOSTER I. Grid Information Services for Distributed Resource Sharing[R]. In : Proceeding of IEEE HPDC-10 , 2001.
- [5] 高尚 , 杨静宇. 群智能算法及其应用[M]. 北京 : 中国水利水电出版社 , 2006.
- [6] 殷锋 , 李志蜀 , 付强 , 等. 基于关联规则的网格资源分域管理[J]. 四川大学学报 : 工程科学版 , 2006 , 38(3) : 129-134.
- [7] KOBAYASHI H , TAKIZAWA H , INABA T , et al. A Self-organizing Overlay Network to Exploit the Locality of Interests for Effective Resource Discovery in P2P Systems[C]. // IEEE. Sendai , Japan : Proceedings of the 2005 Symposium on Applications and the Internet (SAINT'05). Tohoku : Inf Synergy Center , Tohoku University , 2005 : 246-255.
- [8] 熊金波 , 张珊珊. 网格资源发现中资源虚拟组织构建方法研究[J]. 计算机工程与设计 , 2007 , 28(12) : 2863-2866.
- [9] 徐志伟 , 冯百明 , 李伟. 网格计算技术[M]. 北京 : 电子工业出版社 , 2004.