

# 一种基于移动 Agent 的分布式信息检索优化模型

王家伟<sup>1</sup> 赵毅<sup>2</sup> 徐毅<sup>1</sup>

(重庆交通大学计算机信息学院 重庆 400074)<sup>1</sup> (重庆交通大学教育技术中心 重庆 400074)<sup>2</sup>

**摘要** 本文以移动 agent 在信息检索系统中为应用背景,设计了基于多移动 agent 的分布式信息检索优化模型。在模型的求解过程中,提出了求解的启发式策略,以寻找使总的检索时间最小的路由。目标是获得在一次信息检索中的通信、处理资源耗费最小的优化结果。并和其它启发式策略进行了比较,以达到优化的目的。

**关键词** 移动 agent,启发式策略,信息检索,路径优化

## Distributed Information Retrieval Optimal Model Based on Mobile Agents

WANG Jia-wei<sup>1</sup> ZHAO Yi<sup>2</sup> XU Yi<sup>1</sup>

(College of Computer and Science, Chongqing Jiaotong University, Chongqing 400074, China)<sup>1</sup>

(Educational Technology Center, Chongqing Jiaotong University, Chongqing 400074, China)<sup>2</sup>

**Abstract** Based on the using of Mobile Agents for information retrieval, an optimal model of information retrieval has been designed. During solving the model, we present a heuristic algorithm for seeking routes of the mobile agents that minimize the total completion time. The goal to system design is producing an "optimal" answer with minimal use of communication and processing resources. To reach the goal of optimization, we compare our heuristic algorithms with others. Simulation examples show that the model is effectively.

**Keywords** Mobile agent, Heuristic algorithm, Information retrieval, Path optimization

## 1 引言

随着计算机网络,尤其是 Internet/WWW 的迅速发展,Web 已经发展成一个全球的、巨大的、分布的和共享的信息空间,用户要想准确地获取自己所需要的信息资源,就必须借助于 Web 检索工具去进行信息的查询和分析<sup>[1]</sup>。然而,由于传统的 Web 检索工具在很大程度上只是为了方便用户进行信息查找,并没有考虑查找的方式是否快捷,因此其检索工具大多采用的都是低效的检索模式<sup>[2,3]</sup>,这对于飞速增长的 Internet 上通信量来说,已经远远不能适应网络信息服务发展的要求。尽管目前 Internet 的带宽在不断增大,但其发展速度还是跟不上 Internet 通信量的增长速度,因此也极大地制约了 Internet 网络资源的优化控制技术的发展。

为了更加合理、有效、快速地利用 Internet 上巨大的计算资源,达到优化组合网络信息资源的目标,应运而生的移动 Agent 技术便成为了解决这个问题的一个最为有效的方法。因此不少专家学者从不同的角度采用 Agent 技术对网络信息检索系统进行了研究<sup>[4-7]</sup>,取得了一定的成果,可遗憾的是这些信息检索系统都是从单一的角度来考虑网络信息资源的检索,而在日益复杂的网络系统中这种搜索技术往往难以解决信息密集的问题,其信息检索的速度也受到了极大的限制。为了解决这个问题,本文从移动 Agent 信息检索技术出发,结合计算机网络设计了一种加快信息检索速度的分布式多 Agent 信息检索模型。在本模型中,客户端需要同时发起多个同质的移动 agent,每个 agent 需要访问数据提供者(主机)的一个子集,这些主机存储有需要检索的相关信息。移动 agent 在访问完成之后会把检索到的相关信息返回到客户端。这样移动可以直接访问服务器资源,避免了大量数据在网络中的传输,从而降低了系统对网络带宽的依赖。

## 2 基于移动 Agent 的信息检索原理

基于移动 Agent 信息检索系统的整体工作过程是数据查询请求通过客户机或源主机的用户查询界面发起,源主机接受用户的查询请求,完成查询任务分解工作和进行查询调度,并实现查询结果的整合,最后将查询的结果传送给用户查询界面。因此,整个系统作为一个 Region,其中每个主机都是一个 Agency,每个主机都独立维护自己的路由表。首先,用户通过用户界面提交数据检索请求;然后源主机的 Agency 根据该检索请求进行任务分解,或从本地检索数据,或创建一个移动 agent,并将检索任务交给移动 agent;移动 agent 根据路由规划迁移路径并迁移到目的主机 1 上;该移动 agent 按检索目标执行相应的程序代码,检索目的主机 1 的数据库,并把符合要求的结果传给源主机;然后按照规划好的路径,移动 agent 移动到下一站点目的主机 2;MA 在目的主机 2 上也执行相应的程序检索该主机的数据库,将符合要求的结果根据 Agent 通信的方式以消息机制传送给源主机的用户界面 Agent。依次下去,直到检索完规划路径中的所有目的主机后,移动 agent 迁移回到源主机并返回数据。在远程目标主机完成检索的同时,远程主机和源主机之间的网络连接可以断开,各主机也可以同时进行自己的任务。在完成所有任务的数据检索之后,源主机根据用户的查询要求进行查询结果的过滤和整合,最后将符合用户要求的结果数据传给客户端用户,并由用户界面显示结果。工作流程如图 1。

## 3 基于移动 Agent 的信息检索模型

显然,在整个移动 Agent 完成信息检索的工作流程中其核心工作是要完成信息检索。因此,我们需要找到完成检索任务的最小时间耗费解决方案,即定义一个求问题最优解的

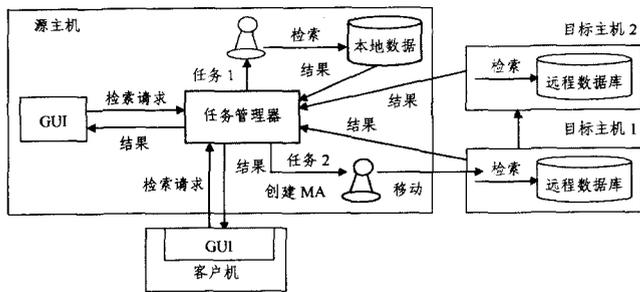


图1 基于移动 Agent 信息检索系统工作过程

方案；目标是找到特定数量的移动 agent 的遍历路由，以达到使总的检索时间最小。基于这种思想，假定一个主机节点上数据由一个 agent 完成检索，并且已知一个移动 agent 检索路由的最后一个主机节点的累计检索数据量。因此，该问题可以抽象为一个无向图  $G(V, E)$ ，其中  $V = \{0, 1, 2, \dots, n\}$  是节点集合， $E$  为图  $G$  的边集，定义  $V'$  为客户端检索数据需要的所有主机节点的集合，即  $V'$  是  $V$  除发起节点（节点 0）的节点子集。不失一般性，用节点 0 表示发起 agent 的客户端，建立如下模型，使  $T$  达到最小化：

$$T = \max_K \sum_{i=0}^N \sum_{j=1}^N \left( \frac{1}{t_{ij}} \cdot S_{ik} \cdot \lambda_{ijk} + t_{1ij} + t_{2i} \right), k=1, 2, \dots, K \quad (1)$$

$$S_{ik} = \begin{cases} L, & \text{当 } i=0 \\ S_{\pi(i)k} \cdot \lambda_{\phi(i)k} + l \cdot z_{ik}, & \text{当 } i \in V' \\ S_{\pi(i)k} \cdot \lambda_{\phi(i)k}, & \text{其它情况} \end{cases} \quad (2)$$

其中  $T$  表示信息检索所花的总时间； $K$  表示发起同质的移动 agent 数目； $k$  表示第  $k$  个移动 agent； $S_{ik}$  表示 agent  $k$  在访问节点  $i$  之后的累计大小； $t_{ij}$  表示边  $i \rightarrow j$  间的传输速率； $L$  表示移动 agent 的初始大小； $l$  表示从一个节点检索的数据量的大小； $z_{ik}$  是一个二进制决策变量，1 表示 agent  $k$  在节点  $i$  有数据检索；反之则为 0； $\lambda_{ijk}$  为一个二进制决策变量，表示 agent  $k$  是否有直接从节点  $i \rightarrow j$  的访问。1 表示有，0 表示没有； $\phi(i)$  表示在路由中节点  $i$  的前驱节点和  $i$  之间的传输延迟； $t_{1ij}$  表示在节点  $i$  测试节点  $i$  和节点  $j$  之间的传输速率所花费的时间； $t_{2i}$  表示节点  $i$  完成检索任务的能力，用时间表示；该属性作为各主机节点路由的附加属性，在第一次完成检索时，可以根据经验赋初始值，在一个移动 agent 实际访问该节点后更新该属性的值。

定义更新函数为：

$$t_{2i} = Q_i / T_i$$

其中  $Q_i$  表示在节点  $i$  检索的数据量的大小； $T_i$  表示完成  $Q_i$  的数据量检索所花费的时间。

#### 4 信息检索的启发式策略

求解本问题的目标是使  $K$  个同质 Agent 完成检索任务的最后一个移动 agent 的耗时最小。该模型的特点主要表现在一个边的遍历时间代价依赖于检索数据集的累计，而这个值是不固定的，而且该优化模型是解决一个非完全图，边的权不固定，即使一个节点无相关数据，移动 agent 也可能去访问该节点。为解决该问题，专门设计了一种启发式策略：基于节点服务信任度的移动 Agent 的路径选择策略。

##### 4.1 基于节点服务信任度的移动 Agent 的路径选择算法

移动 Agent 在网络环境中访问网络节点服务的过程中，必须考虑各节点服务内容、服务能力及网络环境的动态性、网络路径等因素。各节点的服务能力会随着当前负载的变化而

变化；对移动 Agent 来说，各节点当前的服务能力之间又不具有直接的可比性且难以预知。因此，用移动 Agent 对节点的信任度作为衡量服务节点当前服务能力，为后续移动 Agent 提供路径选择的参考。前面移动 Agent 对节点服务越信任，后续同一类型的移动 Agent 就越可能访问该节点。同时移动 Agent 在路径的选择时不能只考虑当前路径段的长度，还要考虑该路径段是否为优化路径的一部分和节点间的传输速率，即使两个服务节点间的路径较短也不一定意味着移动 Agent 选择该路径后会使得它完成任务经过的距离最短，即全局路径最短。因此对路径也用信任度来衡量，后续移动 Agent 可以根据前面 Agent 的信任度选择路径，前面 Agent 对路径的信任度由它是否为一个较优路径的一部分决定的。另外，由于网络负载是动态变化的，当一段网络拥塞时，Agent 的迁移速度大大降低，而延误它完成自己的任务。因此，Agent 还需要考虑路径当前的负载状态。综合上述所有因素，移动 Agent 在网络中的选路算法如式(3)所示：

$$p_{ij}(t) = \begin{cases} \beta_j \cdot \frac{[\mu_j(t)]^{\tau_{ij}(t)}}{\sum_k [\mu_k(t)]^{\tau_{ik}(t)} \cdot t_{1ik}}, j \in V', k \in V' \\ 0, j \notin V' \end{cases} \quad (3)$$

其中， $p_{ij}(t)$  表示在  $t$  时刻移动 Agent 选择由服务节点  $i$  转移到服务节点  $j$  的概率； $t_{1ij}$  是前面总模型中定义的在节点  $i$  测试和节点  $j$  之间的传输速率所花费的时间，用于表示该段路径当前负载状态； $\mu_j(t)$  表示以前访问过服务节点  $j$  的所有移动 Agent 对它的总体信任度； $\tau_{ij}(t)$  是前面经过节点  $i$  和  $j$  间路径的所有移动 Agent 对该段路径的总体信任度； $V'$  表示移动 Agent 在节点  $i$  时下一步允许选择的服务节点的集合； $\beta_j$  表示节点  $j$  的数据内容和整个检索目标内容的相似度； $\delta$  表示节点当前服务能力和路径当前状态对移动 Agent 进行路径选择影响力大小的参数。

##### 4.2 节点服务信任度更新算法

当移动 Agent 访问一个服务节点时，将该节点的应用程序需要的数据返回给应用程序。首先，我们定义了服务质量的概念：

$$\phi_i = \frac{M_i}{Q_i + S_j} \quad (4)$$

其中， $\phi_i$  表示当前节点  $i$  的服务质量， $Q_i$  表示 Agent 在节点  $i$  检索到的数据量； $S_j$  表示 Agent 本次检索过程中在访问节点  $i$  以前所获得的总数据量； $M_i$  表示在节点  $i$  中检索的数据而没有在  $S_j$  中的数据量，该数据量可以通过集合的差运算完成。

为了及时反映当前节点服务能力的变化，需要定义一个服务质量的区间： $A = [\phi_1, \phi_2]$ ，当服务质量在  $A$  中时，就增加该节点的服务信任度，否则就减小该节点的服务信任度作为惩罚，因为 Agent 访问了它却没有获得需要的数据而造成了浪费。移动 Agent 对节点服务信任度的更新算法如(5)，由 Agent 在该服务节点获得的数据量占总需要数据量的百分数和服务质量决定。

$$\mu'_i(t) = \begin{cases} \mu_i(t) + \rho \cdot \phi_i \cdot \frac{Q_i}{T_i S_j} (1 - \mu_i(t)), \phi_i \in A \\ \mu_i(t) - \rho \cdot \phi_i \cdot \frac{\min(Q_i)}{S_j} (1 - \frac{1}{T_i}) (1 - \mu_i(t)), \\ j=1, \dots, n; \phi_i \notin A; \dots \\ 0, \text{其他} \end{cases} \quad (5)$$

其中， $\mu'_i(t)$  表示移动 Agent 在  $t$  时刻访问服务节点  $i$  后的信任度，Agent 用这个信任度来更新服务节点  $i$  的总体信任度；

$\mu_i(t)$ 表示以前访问过服务节点  $i$  的所有移动 Agent 对它的总体信任度;参数  $\rho$  决定每个 Agent 对服务信任程度的影响因子; $Q_i$  表示 Agent 在节点  $i$  检索到的总数据量; $\phi_i$  表示节点  $i$  当前的检索服务质量; $S_j$  表示 Agent 在本次检索过程中获得的总数据量,包含在节点  $i$  检索的数据量; $T_i$  表示 Agent 访问节点  $i$  所花费的时间; $n$  是 Agent 在本次检索过程中所访问的节点数量。

### 4.3 路径信任度的更新算法

移动 Agent 对所经路径的信任度更新算法见式(6):

$$\tau'_{ij} = \begin{cases} \tau_{ij}(t) + \rho \cdot \frac{(1-\tau_{ij})}{Ll_{ij}} \cdot (1 - \frac{l_{ij}}{S_j}), & \text{Agent 访问节点 } j \\ 0, & \text{其它} \end{cases} \quad (6)$$

其中, $L$  为移动 Agent 在本次迁移中当前走过路径的总长度; $l_{ij}$  是移动 Agent 当前走过节点  $i$  和节点  $j$  间路径的长度。Agent 对路径信任度与节点间的当前距离成反比,与节点间的传输速率成正比,即当前这段路径越拥塞信任度就越低。

## 5 仿真试验

为了测试该模型的性能,对三个不同节点数目的节点集进行求解。三个节点集的数目分别为 10,20,50 个,测试选择

表 1 Agent 数目=1

	计算 Agent 最后完成搜索时间(ms)		算法耗时(ms)	
	对照算法	启发式策略	对照算法	启发式策略
V1(n = 10)	315	305	563	584
V1(n = 20)	1903	1876	5129	3473
V1(n = 50)	5664	5452	73417	53402

表 2 Agent 数目=2

	Agent 最后完成搜索时间(ms)		算法耗时(ms)	
	对照算法	启发式策略	对照算法	启发式策略
V1(n = 10)	243	236	325	341
V1(n = 20)	835	871	2764	1971
V1(n = 50)	2912	2906	41550	3057

表 3 Agent 数目=6

	Agent 最后完成搜索时间(ms)		算法耗时(ms)	
	对照算法	启发式策略	对照算法	启发式策略
V1(n = 10)	177	184	117	124
V1(n = 20)	610	642	1328	976
V1(n = 50)	1946	2264	20337	16624

(上接第 246 页)

由表 1 可以看出,文件大小也同样和视频特性有关,视频运动变化越剧烈得到的文件越大。无论采用哪种无线信道,传输开销都和视频文件大小成正比。对于不同的无线传输信道,WiFi,Bluetooth 传输速率较快,而且上传下载速率没有明显变化;而 GPRS 传输速率要慢许多,1 分钟的视频短片上传时间超过 3 分钟,需要注意的是下载时间在 1 分钟以内,若采用视频点播方式几乎没有时间开销。

**结束语** 本文设计和实现了一种基于智能手机的视频共享系统。在该系统中,通过视频发布客户端视频采集、压缩和视频获得客户端视频解码、播放等功能,以及视频服务器的支持,实现了智能手机用户间的视频共享。测试结果表明该系统有较强的实用性,播放视频主观感受良好,传输开销也在可接受的范围。鉴于目前我国有超过千万的智能手机用户,而且智能手机用户数量在今后几年内将会出现井喷式的增长,

每次发起的 Agent 数目分别为 1,3,6 个。对照算法采用文献 [3]提出的解决车辆路径问题的串行存储路由生成算法(Sequential route\_building algorithm based on saving criterion)。

由实验结果可以看出,随着问题求解规模的增加,算法耗时增加较快,由于改进的启发式策略对访问路径的构造是动态生成的,避免了很多无效路径的搜索,有效地减少了算法对问题求解所耗时间,特别在求解规模较大时,比对照算法的求解时间减少比较明显。同时,求得问题解决方案的质量同对照算法得到的问题解决方案相差不大(由 Agent 最后完成搜索时间衡量解决方案质量)。

**结束语** 本文提出的基于多个同质移动 Agent 完成信息检索模型及检索过程中使用的启发式策略考虑了网络的动态性、链路速率的动态性,避免了链路速率随时间和负载变化对求解方案的影响,同时把每个数据节点的服务能力列入考虑因素,更能符合真实应用环境。在构建搜索路由时采用了基于网络节点服务信任度和路径信任度的动态构建,减少了问题求解的时间耗费。

但该启发策略还有一些地方存在不足,如在 Agent 数量与节点总数比率(K/N)大于 0.4 时,生成的解决方案质量有显著下降,笔者将在下一步作进一步研究。

## 参考文献

- [1] Kleinrock L. Nomadicity: Anywhere, Anytime in a Disconnected World. *Mobile Networks and Applications*, 1996, 1: 351-357
- [2] Mole R H, Jameson S R. A sequential route building algorithm employing a generalized savings criterion. *Operational Research Quarterly*, 1976, 27: 503-511
- [3] Clarke G, Wright J W. Scheduling of vehicles from a central depot to a number of delivery points. *Operations Research*, 1964, 12: 568-581
- [4] 顾椿,王小平,曹立明. 基于 Mobile Agent 的信息检索系统的结构及相关技术[J]. *计算机工程*, 2004
- [5] 卢菁,等. Mobile Agent 在信息检索中的应用研究. *东南大学学报(自然科学版)*, 2003, 33
- [6] 于淑惠. 一个基于移动 Agent 的信息检索系统[J]. *现代图书情报技术*, 2004(10)
- [7] 张至柔,罗四维,陈歆,等. 移动 Agent 在网格中的路径优化算法研究. *计算机研究与发展*, 2006, 43(5): 791-796

因此该系统将会有很好的应用前景。

## 参考文献

- [1] Wiegand T, Sullivan G, Luthra A. Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H. 264/ISO/IEC 14496-10 AVC), JVT-G050, Mar 2003
- [2] Schulzrinne H, Rao A, Lanphier R. Real Time Streaming Protocol (RTSP). IETF RFC 2326, 1998
- [3] Wenger S, Hannuksela M M, Stockhammer T, et al. RTP Payload Format for H. 264 Video. IETF RFC 3984, August 2004
- [4] Harrison R. Symbian OS C++ for mobile phones (volume 1) [M]. London: Symbian Press, 2004
- [5] 毕厚杰. 新一代视频压缩编码标准——H. 264/AVC. 第一版. 北京: 人民邮电出版社, 2005
- [6] 马建, 陈建, 牛建伟, 等. 智能手机操作系统编程——Symbian S60 系列. 北京: 科学出版社, 2005