

Intelligent Search Study Based on Improved Ant Colony Algorithm in P2P Networks

Jinqi Su, Yulong Guo

College of Economics and Management, Xi'an University of Posts & Telecommunications, Xi'an
Email: jinqisu@163.com

Received: May 5th, 2014; revised: Jul. 11th, 2014; accepted: Aug. 1st, 2014

Copyright © 2014 by authors and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

In order to enhance the practicality of ant colony algorithm and improve the search efficiency of peer-to-peer networks, this paper presents a new approach of unstructured P2P information retrieval based on the polymorphic ant colony algorithm. In order to meet the new file requirement after a while of searching, the conception of generated pheromone is imported to decrease the blindness of pack forwarding in early searching stage. Based on the simulator framework, simulating the flooding, ant colony algorithm, ant colony algorithm with generated pheromone in unstructured peer-to-peer networks, and analyzing the experience data, the experience results indicate that the algorithm is effective and can enhance the performance of peer-to-peer networks.

Keywords

Peer-to-Peer Network, Search, Polymorphic Ant Colony Algorithm, Generated Pheromone

对等网络中改进蚁群智能搜索 算法研究

苏锦旗, 郭玉龙

西安邮电大学经济与管理学院, 西安
Email: jinqisu@163.com

收稿日期: 2014年5月5日; 修回日期: 2014年7月11日; 录用日期: 2014年8月1日

摘要

为了提高蚁群算法在P2P网络资源搜索中存在搜索盲目、搜索效率低的问题，论文将多态蚁群算法和应用到了P2P网络搜索。针对搜索一段时间后网络中发起的对新的文件请求，引入合成信息素的概念，以减少搜索初始阶段消息转发的盲目性。对无结构P2P网络中的洪泛算法、蚁群算法、引入合成信息素后的蚁群算法进行模拟实验，实验结果表明所提出的算法可有效提高P2P网络的搜索性能。

关键词

对等网，搜索，多态蚁群算法，合成信息素

1. 引言

随着网络带宽的不断提高和计算机处理能力的不断增强，Peer-to-Peer(P2P)作为一种端到端的互联网技术，得到了全面而广泛的发展。P2P网络也即对等网络[1]，其所有的节点直接相连且地位功能平等，对等点既是客户机，同时又充当服务器[2]。其应用有资源共享、通信协作和分布计算三大类，其中应用最广泛的就是资源共享，这使得资源搜索成为P2P网络的核心问题之一。

目前，非结构化P2P网络的资源搜索通常采用基于洪泛(Flooding)算法的消息传递机制，但是洪泛算法容易产生大量冗余通信数据包，导致搜索速度慢，系统效率底[3]。为了充分利用网络资源，降低开销，使负载相对均衡，需要一种新的网络模型和搜索算法机制。基于此，蚁群算法(Ant Colony Algorithm, ACA)等启发式计算方法为P2P网络资源搜索提供了新的思路[4][5]。

作为一种群智能技术，与模拟退火算法和遗传算法相比，蚁群算法具有很强的全局优化能力和并行性，能较快得到搜索结果，更适用于动态网络的资源搜索[6][7]。但是，蚁群算法只有一种信息素标识一种资源，而在真实的网络中资源的种类和数量是非常多的，因此有必要用改进的蚁群算法来提高P2P网络搜索效率[8]。国内外学者对此做了一系列的研究[9]-[11]，算法在提高搜索成功率和搜索效率等方面取得明显效果[12]。但在P2P网络中邻居节点的状态对搜索的效率会有影响，一般情况下响应速度快的节点返回查找结果的速度也比较快，但也有可能因为通过某个邻居节点的路径中的其它节点的响应速度慢，从而影响总的响应时间。所以为了能够最大限度的提高搜索的效率，需要对多种情况加以标记(即需要有多种信息素)，目前的研究较少考虑到这方面的，本论文的研究基于此展开。

2. 算法基础

2.1. 多态蚁群算法

对蚂蚁群体生活观察发现，蚂蚁是有明确的分工又相互协调完成复杂的工作。文献[13]以TSP为例将蚁群中的蚂蚁分为三类：工蚁、侦察蚁、搜索蚁。工蚁主要是根据已经标记好的路径取食回巢，侦察蚁群以每个城市为中心做局部侦察，并用侦察素标记侦察结果，搜索蚁根据侦察素和其它一些辅助消息选择最短路径并做必要的标志，以便工蚁根据标记的消息直接取食回巢。

2.2. 算法基本模型

1) 转移概率计算

蚂蚁 $k(k=1,2,\dots,m)$ 运动过程中根据各条路径上的信息量决定其运动方向，这里用禁忌表 $tabu_k(k=1,2,\dots,m)$ 来记录蚂蚁 k 当前所走过的城市，随着 $tabu_k$ 进化过程作动态调整。在搜索过程中，

蚂蚁根据各条路径上的信息量及路径的启发信息来计算状态转移概率。 t 时刻蚂蚁 k 由城市 i 转移到城市 j 的状态转移概率 $p_{ij}^k(t)$ 为:

$$p_{ij}^k(t) = \begin{cases} \frac{[\tau_{ij}(t)]^\alpha \cdot [\eta_{ik}(t)]^\beta}{\sum_{s \in \text{allowed}_k} [\tau_{is}(t)]^\alpha \cdot [\eta_{is}(t)]^\beta}, & j \in \text{allowed}_k \\ 0, & \text{else} \end{cases} \quad (1)$$

式中, $\text{allowed}_k = \{C - \text{tabu}_k\}$ 表示蚂蚁 k 下一步允许选择的城市; α 为信息启发式因子, 反映蚂蚁在运动过程中所积累的信息在蚂蚁运动时所起的作用, 其值越大, 则该蚂蚁越倾向于选择其他蚂蚁经过的路径; β 为期望启发式因子, 反映蚂蚁在运动过程中启发信息在蚂蚁选择路径中的受重视程度, 其值越大, 则该状态转移概率越接近于贪心规则; $\eta_{ij}(t)$ 为启发函数, 其表达式如下:

$$\eta_{ij}(t) = \frac{1}{d_{ij}} \quad (2)$$

式中, d_{ij} 表示相邻两个城市之间的距离。对蚂蚁 k 而言, d_{ij} 越小, 则 $\eta_{ij}(t)$ 越大, $p_{ij}^k(t)$ 也就越大。显然, 该启发函数表示蚂蚁从元素(城市) i 转移到元素(城市) j 的期望程度。

2) 信息素增量计算

信息素增量 $\Delta\tau$ 的计算如下式所示。

$$\Delta\tau = C * \left(\frac{A}{t_r - t_s} \right)^\alpha * \left(\frac{B}{n} \right)^\beta \quad (3)$$

式中 A 、 B 、 C 均为常数, α 、 β 根据情况可做适当调整, n 为消息包经过的节点数即TTL, t_r 、 t_s 分别为节点收到响应包和请求包时的时间戳。

3) 信息素更新计算

在搜索过程中路径上的信息素积累的越来越多, 为了避免信息素过多淹没启发信息, 在每只蚂蚁走完一步或者完成对所有 n 个城市的遍历后, 要对残留信息进行更新处理。这种更新策略模仿了人类大脑记忆的特点, 在新信息不断存入大脑的同时, 存储在大脑中的旧信息随着时间的推移逐渐淡化, 甚至忘记。由此, $t+n$ 时刻在路径 (i, j) 上的信息量可按如下规则进行调整:

$$\tau_{ij}(t+n) = (1-\rho) \cdot \tau_{ij}(t) + \Delta\tau_{ij}(t) \quad (4)$$

$$\Delta\tau_{ij}(t) = \sum_{k=1}^m \Delta\tau_{ij}^k(t) \quad (5)$$

式中, ρ 表示信息素挥发系数, $1-\rho$ 表示信息素残留因子, 为了防止信息的无限积累, ρ 的取值范围为: $\rho \in [0, 1)$; $\Delta\tau_{ij}(t)$ 表示本次循环中路径 (i, j) 上的信息素增量, 初始时刻 $\Delta\tau_{ij}(0) = 0$, $\Delta\tau_{ij}^k(t)$ 表示第 k 只蚂蚁在本次循环中留在路径 (i, j) 上的信息量。

4) 信息素生成计算

邻居节点 i 的信息素 γ_i 的生成按如下的公式进行计算。

$$\gamma_i = \frac{\sum_{j \in P_i} \tau_{ij}}{|P_i|} * \left(\frac{|P_i|}{p} \right)^\alpha \quad (6)$$

式中, P_i 为经由邻居节点 i 查询到的文件的集合, τ_{ij} 为文件 j 经过邻居节点 i 的信息素, p 、 α 为参数, 可以取 $p=1$ 、 $\alpha=1$, 此时也即合成信息素与经过这个节点的所有文件的信息素的均值成正比, 与经过这个节点查找到文件的总数成正比。

3. 基于多态蚁群算法的 P2P 网络资源搜索

3.1. 算法实现

实际 P2P 网络中, 每个节点上可能存放很多不同的文件, 需要引入多种信息素标记所有的文件, 标记邻居节点的信息素表与基本蚁群算法中的相同; 标记文件状态的信息素则需做相应的调整, 单个文件的信息素表记录数据项有文件名、记录邻居节点的标识、信息素和更新时间戳。

由于对每个文件都需要记录其信息素, 所以在信息素记录表项中不仅需要记录邻居节点的名字标识, 还需要记录文件的名字。为方便查找可将每个文件的信息素表链接在一起, 组成链式存储结构, 如图 1 所示。

在图 1 中, 先将每个文件的信息素链接在一起(如 File1), 每个邻居节点均可以返回查询结果; 其他文件也类似, 将可以返回查找结果的邻居节点链接起来。该存储方式查找时可以根据文件名快速找到所有与其相关的邻居节点的信息素状态, 效率相对要高很多; 但是更新邻居节点的状态信息的代价比较大, 当某个节点失效时必须通过文件名遍历链表, 以删除所有与这个节点相关的信息素信息。

3.2. 消息路由过程

当节点收到邻居节点发给自己的查询请求时, 按以下算法进行消息的转发操作。

- 1) 判断消息包的标识符是否和最近一段时间内收到的相同, 若相同, 则将该查询包丢弃;
- 2) 在节点本地进行查找, 若找到按原路返回查询结果, 结束查询;
- 3) 在节点的信息素表中查找请求的文件名的信息素组, 若没有找到, 随机选择一个邻居节点进行转发, 结束查询;
- 4) 用该组信息素数据按公式(1)计算转发概率, 并根据计算出的概率转发查询包, 结束查询。

查询消息转发的流程如图 2 所示。

为了使节点能够判断收到消息包是否重复, 需要在结点上建立消息缓存表, 记录最近一段时间内收到过的消息。消息转发缓存表的记录数据项包括消息包标识符、记录邻居节点的标识和消息转发时间戳。

当节点收到查询包时, 先将该查询包的标识符和缓存表中的记录作比较, 缓存表中存在该条记录, 说明是最近已经转发过的重复消息, 将其丢弃; 没有该条消息的标识符, 则是新的消息, 按照正常的步骤进行转发。为了使节点能够按原路的路径将查询结果消息传回, 节点需要缓存最近一段时间消息的来源邻居节点地址, 当收到查询消息的响应时, 将其转发回去。同时节点周期性的检测接收时间戳, 当超过一定的时间以后, 即使再收到响应消息也失去意义, 所以将超时的记录信息删除, 以减少不必要的冗余消息。

3.3. 信息素的更新

为了提高搜索效率, 需要将搜索信息及时存储到信息素中, 为后面的搜索提供指导。考虑多种不同文件的情况下, 当节点收到查询响应包时信息素的更新算法如下。

- 1) 按照公式(3)计算机信息素变化量 $\Delta\tau$;
- 2) 根据响应包中的文件名从信息素链表中查找该文件的信息素组, 若找到, 按公式(4)将文件信息素更新;
- 3) 生成一个空的信息素组结构, 插入到信息素链表的对应位置, 并将每个邻居节点的信息素初始化为初始值 0。

为了限制信息素无限制的累积增加, 每经过规定的时间间隔需要减少信息素的浓度, 进而模拟信息素的挥发机制。

信息素的更新流程如图 3 所示。

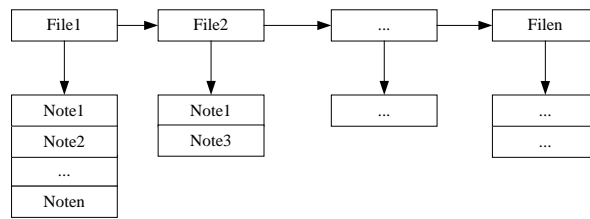


Figure 1. Information storage

图 1. 信息素存储方式

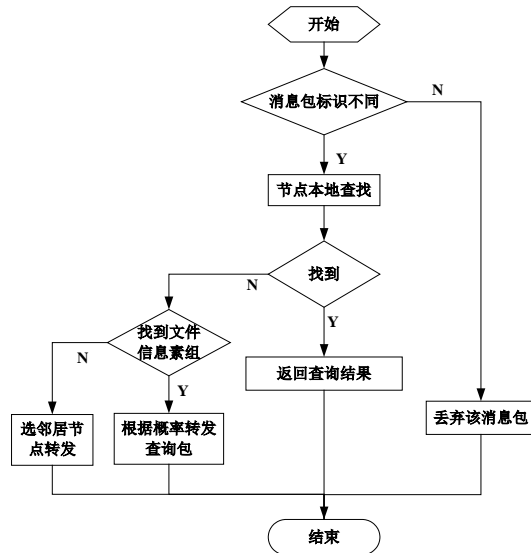


Figure 2. Routing process of polymorphic ant colony algorithm message

图 2. 多态蚁群算法消息路由流程

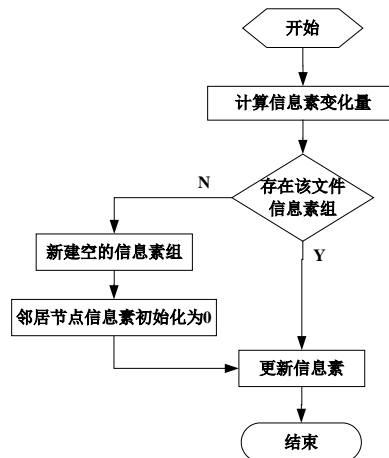


Figure 3. Update process of pheromone

图 3. 信息素更新流程

3.4. 合成信息素及其提取

当某个文件是第一次查找时，按上述算法随机选择一个节点，转发消息时是盲目的，但通常响应速

度较快的节点对其存放的每个文件均可以较快的速度返回查询结果，根据对其它文件的查找速度来间接推断对当前文件的响应速度，并把推断得到的有关响应速度的信息作为间接信息素，用合成信息素来对其进行标识。为了记录合成信息素信息，每个节点上需要建立合成信息素表，合成信息素表的记录数据项由邻居节点标识和信息素组成。引入合成信息素后，消息的路由算法如下。

- 1) 判断消息包的标识符是否存在于消息缓存表中，若在，将该包丢弃；
- 2) 在本地进行查找，若找到，转发给来源邻居节点；
- 3) 查找对应文件名的信息素组，若未找到，在结点上查找任意文件的信息素；
- 4) 根据公式(1)计算转移概率，转发消息包；
- 5) 在结点上查找任意文件的信息素，若不存在，随机选择一个邻居节点转发；
- 6) 根据公式(6)计算合成信息素，按概率进行转发；
- 7) 随机选择一个邻居节点转发；
- 8) 直接将该包丢弃；
- 9) 转发给来源邻居节点；
- 10) 结束。

消息转发的执行流程如图4所示。

4. 实验及结果

实验对洪泛算法、多态蚁群算法、引入合成信息素后的蚁群算法分别进行模拟，在同等条件下统计成功返回的结果数和网络中产生的总消息包，实验结果如图5和图6所示。

从图5中可以看出，随着搜索次数的增加，由于洪泛算法采用盲目搜索，即使搜索已经进行了很多次，仍然和刚开始一样进行随机转发，成功返回的消息包数并没有明显的变化。采用蚁群算法，由于随着搜索的进行各个节点上逐渐建立起来信息素，再搜索时信息素为消息包的转发提供了参考信息，减少了转发的盲目性，成功返回的结果数也随之越来越多。但蚁群算法在某个文件搜索的初始阶段消息的转

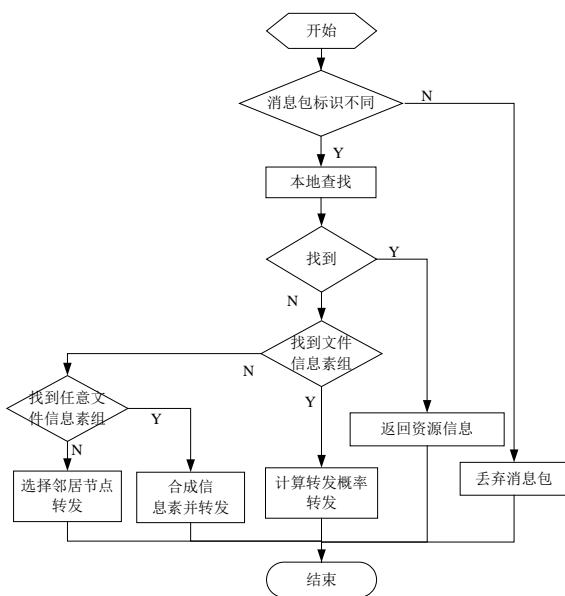


Figure 4. Routing process of ant colony algorithm message including synthetic pheromone

图4. 引入合成信息素的蚁群算法消息路由流程

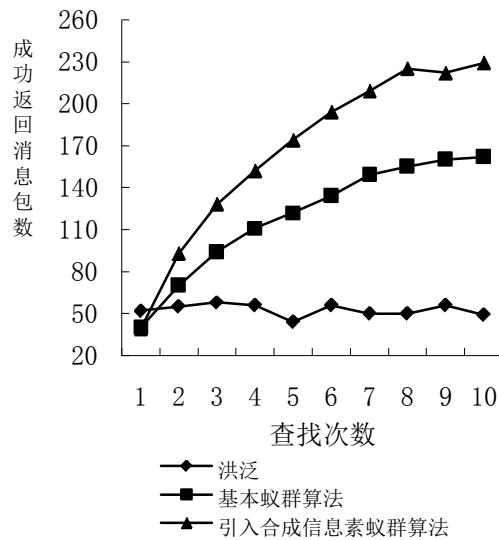


Figure 5. Change of the number for the successful return message with the query times

图 5. 成功返回消息数随查询次数变化情况

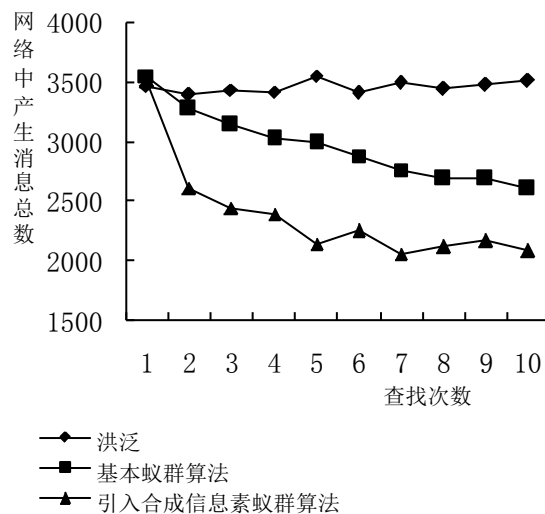


Figure 6. Change of the total message number with the query times

图 6. 产生消息总数随查询次数变化情况

发仍具有很大的盲目性，合成信息素的引入使得在搜索的初始阶段也能在一定程度上减少消息转发的盲目性，所以成功返回的消息包数有了进一步的提高。

图 6 反映的是在同等条件下网络中产生的总的消息数与查询次数的关系。从图中可以看出，由于洪泛采用盲目搜索算法，整个搜索过程中网络的状态没有明显的改变，产生的消息量基本稳定。采用蚁群算法，搜索使得信息素表逐渐建立起来，转发时可以根据信息素进行，盲目性降低，随着搜索的进行，网络中产生的不必要的查询包降低，总消息数也在减少。引入合成信息素后，由对搜索进行了一段时间对刚开始的搜索提供了一定的指导，减少了在某个文件搜索初始阶段转发的盲目性，产生的总的消息包也有一定程度的降低。

综合分析看出，蚁群算法在提高搜索的成功率和降低网络中的消息数两方面均优于洪泛机制，引入

合成信息素后 P2P 网络的性能进一步提升。

5. 结束语

P2P 搜索中的洪泛机制由于转发的盲目性,产生的冗余消息比较多,影响了搜索的性能。引入蚁群算法后,利用以往的查找历史信息,指导未来的搜索过程,减少了搜索的盲目性,提高了效率。为了让蚁群算法更好的适应 P2P 网络的情况,将多态蚁群算法的思想应用到 P2P 网络中,并在应用的过程中做了适当的修改,使得蚁群算法更适合于 P2P 网络:当进行一个从未进行过的查找时,可以根据对以往其它文件的查找结果情况做预先的估计,以减少查找的盲目性,为此引入合成信息素的概念及提取策略,减少刚开始进行的查找消息转发的盲目性,以提高总体搜索性能。

基金项目

国家自然科学基金项目(71173248),陕西社会科学基金(13Q081),西安邮电大学青年教师科研基金项目(ZL2012-30)。

参考文献 (References)

- [1] 吴国庆 (2008) 对等网络技术研究. *计算机技术与发展*, **7**, 100-103.
- [2] Kojima, K. (2003) Grouped peer-to-peer networks and self-organization algorithm system. *IEEE International Conference on Systems, Man and Cybernetics*, **3**, 2970-2976.
- [3] 钱宁, 吴国新 (2010) 无结构化 P2P 网络资源搜索机制研究综述. *计算机科学*, **4**, 7-10.
- [4] Abdulhai, B., Pringle, R. and Karakoulas, G.J. (2003) Reinforcement learning for true adaptive traffic signal control. *Journal of Transportation Engineering*, **129**, 278-285.
- [5] 夏启志, 谢高岗 (2005) 无结构 P2P 网络搜索方法及其改进. *计算机应用研究*, **9**, 256-260.
- [6] Dorigo, M., Maniezzo, V. and Coloni, A. (1996) Ant system: Optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man and Cybernetics, Part B: Cybernetics*, **26**, 29-41.
- [7] Wu, C.-J., Yang, K.-H. and Ho, J.-M. (2006) An ant search algorithm in unstructured peer-to-peer networks. *Proceedings of the 11th IEEE Symposium on Computers and Communications*, 26-29 June 2006, 429-434.
- [8] 段海滨 (2006) 蚁群算法原理及其应用. 科学出版社, 北京.
- [9] 苏玉, 毛力 (2010) 基于蚁群算法的非结构化 P2P 搜索机制的研究. *计算机工程与设计*, **5**, 939-941.
- [10] 李春秀, 刘方爱 (2012) 基于蚁群算法的非结构化 P2P 网络资源搜索策略. *计算机工程与应用*, **4**, 97-99.
- [11] Li, J.Q., Pan, Q.K. and Xie, S.X. (2008) Research on peer selection in peer-to-peer networks using ant colony optimization. *Proceedings of the Fourth International Conference on Natural Computation*, 18-20 October 2008, 516-520.
- [12] 蔡康 (2012) 基于改进型蚁群算法的 P2P 网络资源搜索的研究. *电信科学*, **3**, 32-42.
- [13] 徐精明, 曹先彬, 王煦法 (2005) 多态蚁群算法. *中国科学技术大学学报*, **1**, 59-65.