

访问兴趣相似性 P2P 网络模型

郑晓健^{1,3}, 付铁威², 李彤¹

(1. 云南大学 软件学院, 云南 昆明 650091;

2. 昆明理工大学 计算中心, 云南 昆明 650093;

3. 昆明理工大学 津桥学院 计算机科学与电子信息技术系, 云南 昆明 650106)

摘要:传统资源搜索方法在非结构化对等网络中查找资源时存在的主要问题是检索效率低和通信开销大, 文中提出新的基于访问兴趣相似性的 P2P 网络模型。通过节点的行为特征所表现的访问频谱相似性寻找节点并形成集合, 然后从中选择部分适合的对象建立少量的远程链接, 可以改善传统的非结构化对等网络资源搜索性能。在选择访问频谱相似节点建立远程链接时, 利用了高频访问节点的聚集特性而使工作更加容易。为实现该模型的资源检索, 设计了相应的资源搜索算法。仿真结果表明, 该模型在一定程度上提高了非结构化 P2P 资源搜索的效率, 同时减少了网络中的通信冗余信息量。

关键词: P2P 网络; 搜索模型; 访问频谱; 相似性

中图分类号: TP393.01

文献标识码: A

文章编号: 1673-629X(2015)04-0057-04

doi: 10.3969/j.issn.1673-629X.2015.04.0014

P2P Network Model Based on Interest Similarity

ZHENG Xiao-jian^{1,3}, FU Tie-wei², LI Tong¹

(1. School of Software, Yunnan University, Kunming 650091, China;

2. Center of Computer, Kunming University of Science and Technology, Kunming 650093, China;

3. Department of Computer Science and Electronic Information Technology, Oxbridge College, Kunming University of Science and Technology, Kunming 650106, China)

Abstract: The low efficiency and large communication overhead are the main problems for traditional resource searching method looking for the resource in unstructured P2P network. To solve this problem, a new P2P network model based on interest similarity is given in this paper. The access spectrum similar nodes are found by node behavior and constitute a set of similar nodes. Through establishing the few remote connections among these nodes will improve the traditional unstructured P2P network resources searching performance. Using the high-frequency access node aggregation characteristic makes the work more easily when selecting and accessing the suitable spectrum similar nodes to establish the remote connection. Meanwhile, a new resource searching algorithm is designed for the model of resource retrieval. The simulation results show that the model can improve the resources searching efficiency of unstructured P2P network, while reducing the information redundancy in the network.

Key words: P2P networks; searching model; accessing spectrum; similarity

0 引言

由于 P2P 网络的基本特征符合小世界理论^[1-2], 按小世界理论构建的无结构 P2P 网络将融和规则网和随机网的特点, 能够提高资源查询性能^[3-4]。大量研究表明, 在一些网络节点间建立少量远程链接就可以缩短大多数节点查找资源时的访问路径长度, 从而改善整个网络的综合效能^[3-5]。

节点之间的关系与社会中人与人的关系类似^[6-7], 改变少数人之间的关系会影响到与他们相关的很多人之间的交往关系, 例如婚姻关系就能改变双方家庭成员间的关系。另外, 远程链接节点的选择是有条件的, 即随意建立起来的链接不一定能达到理想的效果^[3-4]。

收稿日期: 2014-05-20

修回日期: 2014-08-26

网络出版时间: 2015-02-23

基金项目: 国家自然科学基金资助项目(61262024); 云南省教育科学基金资助项目(2010C011)

作者简介: 郑晓健(1963-), 男, 硕士, 讲师, 研究方向为对等网络和分布式计算。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20150223.1233.014.html>

1 相关工作

问题的关键在于如何建立更有效的远程链接。文献[3-5]利用网络的结构特性来建立节点的远程链接,取得了明显缩短查询路径长度的效果,而且算法简单、开销比较小,但只利用了节点的静态信息。由于节点信息对其行为的描述作用会随时间的增长而逐步衰减,并在预测节点行为时产生较大偏差^[6-8],因此,应该考虑节点行为和兴趣的动态变化特性。

一些研究者从节点的行为相似性方面着手探索资源检索中可行的方法^[7]。根据 P2P 网络中节点的访问兴趣建立社群,使得具有相似兴趣的节点处于相同社群,且让它们相互之间的距离更接近^[7]。文献[8]提出将节点文档中具有相同关键字的节点聚簇,使节点在检索资源时可从簇内相似的邻居节点那里找到目标。文献[9]提出按节点兴趣相似性对网络节点进行聚簇的方法,再根据查询请求以簇为基础选择兴趣相似的节点转发查询消息,以控制消息扩散范围。在相似性网络模型研究方面,文献[10]提出将兴趣相似度高的节点用短路径相连,构建基于语义的网络覆盖模型。文献[11]提出具有自适应能力的超节点网络模型来描述具有共同兴趣节点的语义相似性。文献[12]利用对等节点推荐方法来构建节点的相似兴趣聚簇。以上方法都取得了一定的效果,但在建立有效的远程链接问题上还是不够理想。

文中提出一种新的基于访问兴趣相似性的 P2P 网络模型,通过对节点的行为和兴趣特征简称节点访问谱线的比较,以发现节点访问谱线的相似性,并建立共有访问谱线相似节点集,再从中选出具有访问谱线稳定性高、兴趣覆盖面宽、交易活跃的节点作为远程连接节点,进而达到缩短查找路径的目的。为了发现适合的远程链接节点,应用了高频访问区域探测法^[6],通过节点访问产生的聚集效应更有效地找出符合要求的节点以建立远程链接。

2 访问谱线检索模型

2.1 基本思想

文献[7]称具有相同或相似兴趣偏好的一组节点为节点群。在节点群内,节点间的相似虽可互相满足一定的需求,但不能满足所有节点的所有需求。当一个节点群所有节点都不能满足查询需求时就要做跨节点群的查询,即要通过跨节点群的远程链接来实现。

另外,节点拥有相似的访问谱线的原因是节点访问了相同的目标节点集合。仿真实验对网络节点形成的访问谱线及其相似性进行了比对,结果表明在资源的有效时间段内节点对网络资源的访问会形成相似的访问谱线片段。实验中将相似兴趣的节点链接起来

后,节点的平均查询路径长度有明显的缩短,如果选择链接的节点的兴趣面越广、交易越活跃,对其他节点的资源检索性能的影响也越大^[6,13-14],那么把这类节点称为泛兴趣节点,利用它们建立远程链接可以提高资源检索性能。

由于节点的访问特性受到网络幂律分布特征的影响^[5],要顺利找到泛兴趣节点可以依靠网络的拓扑结构特征和所产生的访问特征信息来实现^[5-6,15]。文献[6]的研究表明网络中一些具有较多链接而被高频访问的节点周围会聚集着同样被高频访问的节点,并形成高频访问区域。因此为了尽快找到适合的远程链接节点,可以通过网络中节点访问频率的变化来感知高频访问区域,进而可以更快地发现其中的泛兴趣节点。

2.2 模型建立

首先约定在理想网络状态讨论模型,即任何节点都可以随意地访问到网络中的其他节点。设节点集合为 $G = \{v_1, v_2, \dots, v_n\}$, 节点生存期被划分为 n 个时段 $T = \{t_1, t_2, \dots, t_n\}$, v 在时段 t_m 内对 G 中节点的访问次数为 $V(t_m) = (\alpha_{m1}, \alpha_{m2}, \dots, \alpha_{mn})$, 其中 α_{mj} 为在时段 t_m 内 v 对 G 中节点 v_j 的访问次数。

定义 1: 在时段 t_m , 节点 v 访问 G 中节点的次数与此前访问节点的平均数之差称为 v 在时段 t_m 的交易活跃度,即

$$A = \sum_{i=1}^n (\alpha_{mi} - \bar{\alpha}_{mi}), \bar{\alpha}_{mi} = \frac{1}{m} \sum_{i=1}^m \alpha_{mi} \quad (1)$$

若 $A > M$, 则称 v 为交易活跃节点, M 为交易数阈值。

定义 2: 在时段 t_m , v 访问 G 中节点的次数和称为 v 的交易覆盖度,即

$$C = \sum_{i=1}^n \beta_{mi} \quad (2)$$

其中

$$\beta_{mi} = \begin{cases} 1 & \alpha_{mi} \neq 0 \\ 0 & \alpha_{mi} = 0 \end{cases} \quad (3)$$

若 $C > F$, 则称 v 为交易广泛节点, F 为交易覆盖度阈值(一般设置为网络节点单位时段平均交易覆盖度)。

定义 3: 在时段 t_m , 节点 v_i 与 v_j 访问相同节点的差异量与访问节点总数之比称为 v_i 与 v_j 的访问谱线相似度,即

$$\text{Sim}_{ij} = \sum_{i=1}^n \max\left(\frac{\alpha_i - \beta_i - \theta}{|\alpha_i - \beta_i - \theta|}, 0\right) / \sum_{i=1}^n \text{VN}_i \quad (4)$$

其中, $V_i(t_m) = (\alpha_{m1}, \alpha_{m2}, \dots, \alpha_{mn})$, $V_j(t_m) = (\beta_{m1}, \beta_{m2}, \dots, \beta_{mn})$ 为 v_i 与 v_j 在时刻 t 的请求访问次数。

$$\text{VN}_i = \begin{cases} 1 & \alpha_i + \beta_i > 0 \\ 0 & \alpha_i + \beta_i = 0 \end{cases} \quad (5)$$

若 $\text{Sim}_{ij} > S$, 则称 v_i 与 v_j 为访问谱线相似节点, 记为 \tilde{v}_{ij} , S 为访问谱线相似度阈值。

定义4:若 $\exists \tilde{V} \subset G, \tilde{V} = \{v_i | \tilde{v}_{ij}, \forall v_i \in \tilde{V}, \forall v_j \in \tilde{V}\}$, 则称 \tilde{v} 为 G 的共有访问谱线相似节点集。

定理1:节点的访问路径长度与 \tilde{v} 节点在 G 中所占的比例成反比。

证明:假设为 $\forall a \in \tilde{v}$ 建立一个长度为 $R_L = |\tilde{V}|$ 的索引表, 记录节点的 IP 地址, 借助索引表 a 对其他节点的访问路径长度不会超过 $|G| - R_L + 1$, 如果所有节点索引表的内容不重复且都是访问谱线相似节点, 那么节点的访问路径长度不会超过 $|G| / \sum_{a \in \tilde{v}} R_L$ 。所以节点的访问路径长度与 \tilde{v} 节点在 G 中所占的比例成反比, 即 \tilde{v} 节点越多, 节点的访问路径长度越短, 证毕。

定义5:若 $\exists n \in G, n \in \bigcap_{i=1}^k \tilde{V}_i$, 其中 \tilde{V}_i 为 G 的共有访问谱线相似节点集, k 为阈值, 则称 n 为 $\tilde{V}_i (i = 1, 2, \dots, k)$ 的泛兴趣节点。

泛兴趣节点可以缩短检索路径长度的原因是其上建立的跨共有访问谱线相似节点集的远程链接。为 $\tilde{V}_i (i = 1, 2, \dots, k)$ 的节点设置节点索引表, 其中的泛兴趣节点记录其他 \tilde{V}_i 的泛兴趣节点的 IP, \tilde{V}_i 的普通节点保存其所属 \tilde{V}_i 的泛兴趣节点和访问谱线相似节点的 IP, 于是普通节点通过所在 \tilde{V}_i 的泛兴趣节点可向其他 \tilde{V}_i 发送检索消息, 再通过目标 \tilde{V}_i 的泛兴趣节点在局部范围 flooding 消息, 实现节点间的远程查询, 检索路径长度比传统 flooding 将大为缩短, 消息扩散范围也得到控制。

因为, 源 \tilde{V}_i 和目标 \tilde{V}_i 中任意 2 个节点如果不是邻居节点, 它们的检索距离都将大于等于 2。若通过泛兴趣节点在源 \tilde{V}_i 中不用 flooding, 因此可以减小消息扩散范围。

找寻泛兴趣节点方法, 利用泛兴趣节点具有比一般节点更高的交易活跃度即访问频度高、容易在其周边形成高频范围区域的特点^[6]。通过将“找寻消息”收集到的节点访问频度存入消息的访问频度队列(队列长度由对区域直径的统计数据决定), 用 LRU 方法更新队列, 并统计其访问频度总和是否超限, 如果超过规定的高频访问区域的阈值, 则认为存在高频访问区域, 再通过式(1)、(2)、(4)计算区域节点的交易活跃度、交易覆盖度和访问谱线相似度, 找到符合要求的节点作为泛兴趣节点。

2.3 检索算法

检索时消息 M 的传播方式是: 在一般节点以随机漫步方式转发给邻居节点, 遇到有远程链接的节点, 则利用远程链接和邻居节点一起转发, 直到找到目标节点。算法如下:

- 输入:源节点 v_s 发出检索请求 M ;
- 输出:检索到目标节点 v_d 或检索失败。
- (1)接收 M ;
- (2)查询本节点 v 有没有所要找的资源, 有则向 v_d 返回结果, goto(5)
- (3)if TTL=0 then Goto(5)
- (4)if $v \in \tilde{v}$ then
- 由远程链接节点转发 M 和以随机漫步方式向邻节点转发 M
- else
- 以随机漫步方式向邻节点转发 M
- (5)继续侦听消息

3 仿真实验

实验的目的是测试验证模型缩短检索路径长度的有效性。对网络节点形成的访问谱线进行相似性分析, 选择符合模型要求的泛兴趣节点建立远程链接和检索测试, 分析节点检索路径长度的变化。测试网络设置 1 000 个网络节点, 在 10 分钟内每个节点以每秒随机产生 1 个对其他网络节点的访问, 并统计它们的平均查询路径长度。随机抽取 20 个节点, 分析所形成的访问谱线及其相似性, 发现有 3 个节点的谱线满足相似性标准。结果如图 1 所示。不难发现节点的 2、4、6、7、9 频段的访问谱线相似度较高, 表明它们在测试时段内对相同节点进行了满足谱线相似度的访问。接着, 建立访问谱线相似节点的索引表, 挑选符合标准的访问谱线相似节点和泛兴趣节点建立远程链接, 构成不同覆盖度的共有访问谱线相似节点集, 随后再进行同规模的测试(查询将借助远程链接实现), 统计平均查询路径长度的变化。结果如图 2 所示, 可以看出覆盖度的扩大对访问路径的缩短有较大改善。

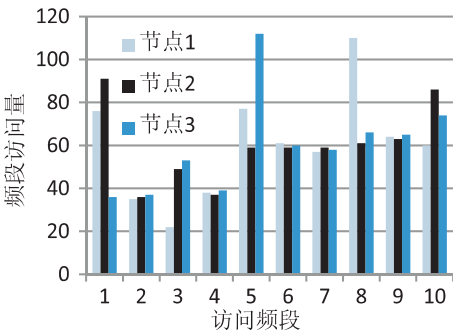


图1 节点访问谱线相似性对比

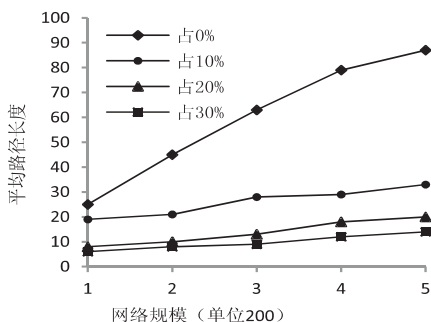


图2 平均访问路径长度

4 结束语

按照网络的小世界特性,提出对节点在资源查找过程中形成的访问频谱进行相似性对比分析,通过发现网络中的访问频谱相似的节点,并将它们组成共有访问谱线相似节点集,再从中筛选出交易活动能力强、交易范围面广的节点和泛兴趣节点,利用它们形成的超出一般节点的交易集合及其交集建立节点间的远程链接。仿真实验对模型的有效性给予了验证,结果表明新方法使网络通信范围和访问路径长度大幅度缩短,改善了网络的资源检索效能,减少了网络带宽资源的消耗。


参考文献:

- [1] Newman M E J, Barabasi Albert-Laszlo, Watts D J. The structure and dynamics of networks [M]. Princeton, New Jersey: Princeton University Press, 2006.
- [2] Hui K Y K, Lui J C S, Yau D K Y. Small-world overlay P2P networks: construction, management and handling of dynamic flash crowds [J]. Computer Networks, 2006, 50 (15): 2727-2746.
- [3] Kleinberg J M. The small-world phenomenon: an algorithm perspective [C]//Proceedings of 32nd annual ACM symposium on theory of computing. Portland, OR: ACM, 2000: 163-170.
- [4] Shen Jingbo, Li Jinlong, Wang Xufa. Effect of long-distance connections selection method on object lookup in P2P networks [J]. Journal of Chinese Computer Systems, 2011, 32 (1): 99-102.
- [5] Xu Haimei, Lu Xianliang, Ge Lijia, et al. Rare resource's sharing mechanism in unstructured P2P networks [J]. Journal of Electronics & Information Technology, 2009, 31 (8): 2028-2032.
- [6] Zheng Xiaojian, Zheng Xiaolan, Li Tong, et al. High frequency access areas discovery algorithm in peer-to-peer network [J]. Computer Engineering and Design, 2014, 35 (3): 780-784.
- [7] Yang Yanchun, Meng Xiangwu. A keyword-based model of similarity measurement on peer group in peer-to-peer networks [J]. Journal of Wuhan Univ (Nat Sci Ed), 2011, 57 (6): 489-493.
- [8] Shi Qingwei, Wu Rongteng. A protocol for text retrieval in structured P2P networks [C]//Proc of international conference on biomedical engineering and computer science. Wuhan: IEEE, 2010: 1-4.
- [9] Yu Ge, Yan Ting. Similarity-based semantics searching in super-peer network model [C]//Proc of international conference on internet technology and application. [s. l.]: [s. n.], 2010.
- [10] Lin C J, Chang Y T, Tsai S C, et al. Distributed social-based overlay adaptation for unstructured P2P networks [C]//Proc of IEEE global internet symposium. Anchorage, AK: IEEE, 2007: 1-6.
- [11] Garbacki P, Epema D H J, van Steen M. The design and evaluation of a self-organizing superpeer network [J]. IEEE Transactions on Computers, 2010, 59 (3): 317-331.
- [12] Mordacchini M, Dazzi P, Tolomei G, et al. Challenges in designing an interest-based distributed aggregation of users in P2P systems [C]//Proc of international conference on ultra modern telecommunications and workshops. St. Petersburg: IEEE, 2009: 1-8.
- [13] Zhang Yuxiang, Zhang Hongke. A load balancing method in superlayer of hierarchical DHT-based P2P network [J]. Chinese Journal of Computers, 2010, 33 (9): 1580-1590.
- [14] Li Zhen, Duan Hancong, Nie Xiaowen, et al. Routing optimization on the layered peer-to-peer management network [J]. Journal of Chinese Computer Systems, 2012, 31 (1): 54-57.
- [15] Huang Yongsheng, Meng Xiangwu, Zhang Yujie. Strategy of content location of P2P based on the social network [J]. Journal of Software, 2010, 21 (10): 2622-2630.

访问兴趣相似性P2P网络模型

作者：[郑晓健](#)，[付铁威](#)，[李彤](#)，[ZHENG Xiao-jian](#)，[FU Tie-wei](#)，[LI Tong](#)

作者单位：[郑晓健, ZHENG Xiao-jian\(云南大学 软件学院, 云南 昆明 650091; 昆明理工大学 津桥学院 计算机科学与电子信息技术系, 云南 昆明 650106\)](#)，[付铁威, FU Tie-wei\(昆明理工大学 计算中心, 云南 昆明, 650093\)](#)，[李彤, LI Tong\(云南大学 软件学院, 云南 昆明, 650091\)](#)

刊名：[计算机技术与发展](#)

英文刊名：[Computer Technology and Development](#)

年，卷(期)：2015(4)

引用本文格式：[郑晓健](#). [付铁威](#). [李彤](#). [ZHENG Xiao-jian](#). [FU Tie-wei](#). [LI Tong](#) [访问兴趣相似性P2P网络模型](#)[期刊论文]-[计算机技术与发展](#) 2015(4)