

基于兴趣和信任评价的 P2P 语义社区模型

王莉, 胡高霞

(太原理工大学计算机与软件学院, 太原 030024)

摘要: 针对大规模网络中 P2P 信息共享面临的资源定位问题和“搭便车”现象, 在考虑节点语义相似度、信任度和活跃强度的基础上, 提出一种语义社区构建方法, 提高 P2P 资源组织的结构性和定位准确性。建立基于信任评价的语义社区动态调整机制, 以减少或消除“搭便车”的节点。

关键词: 对等网络; 语义社区; 信任; 活跃强度; 搭便车

P2P Semantic Community Model Based on Interest and Trust Evaluation

WANG Li, HU Gao-xia

(College of Computer and Software, Taiyuan University of Technology, Taiyuan 030024)

【Abstract】 Aiming at the resource-located problems and free-riding phenomenon during sharing P2P information in large scale network, this paper proposes a semantic community establishing method based on consideration of semantic similarity degree, trust degree and active strength, which improves the structural and resource-located veracity of P2P resource organization. A dynamic evolution mechanism of communities based on trust evaluation is built to reduce and eliminate the free-rider peers.

【Key words】 P2P network; semantic community; trust; active strength; free-riding

1 概述

对等网的非集中性、自治性和容错性等特征使其成为 Internet 上有效的信息共享模型。针对 P2P 网络的资源定位问题, 研究者提出了很多社区模型, 主要包括: (1) 基于信任的社区模型。文献[1]指出信任是用户基于经验和期望对其他用户提供高质量服务或资源能力的评价。它强调受信用户提供的服务能否满足或超过用户预期, 以及用户之间评价的相似程度。Comtella 系统比较用户对不同文档的评价, 以判断用户之间兴趣的相似度, 从而构成社区。(2) 基于“小世界”原理的社区模型。文献[2]认为对等网络具有人为社会属性, 属于“小世界”网络。考虑网络节点隐私性, 通过分析用户公开声明信息和查询记录, 将节点兴趣分为公开兴趣和私密兴趣。利用公开兴趣比较得到社区, 并根据“小世界”现象中无标度网络模型的偏好依附特征, 使新增节点趋于连接到已存在的、具有较高信任度的节点上, 以得到超级节点。(3) 基于语义的 P2P 社区模型。文献[3]提出通过本体相似度匹配来改善消息路由的 P2P 覆盖网络 SemreX。文献[4]建立一种以带权重的词条列表为元数据的方法, 并进行基于模糊集理论的节点相似度计算, 从而构成 P2P 社区。

社区构造为 P2P 中分散资源的管理和定位提供了大粒度结构化方法, 但 P2P 的非集中性导致节点通常更多地表现出自兴趣和自主性, 单个节点对自身网络效用最大化^[5]的追求带来了“搭便车”问题, 即 P2P 中的节点仅从系统中获取其他节点提供的服务, 而不为系统作贡献的行为。因此, 即使得到了合理社区, 如果社区内的资源点只想获取而不想奉献, 则无法实现对资源快速、高效的定位和服务。鉴于此, 本文研究语义社区时充分考虑了“搭便车”问题。

抑制对等网络中“搭便车”行为的普遍原则如下: 根据节点贡献确定它能享受服务的能力。贡献越小的节点从网络中查询、下载资源的能力越弱。文献[6]将抑制机制分为基于节点行为的激励机制、基于博弈论的方法、采用社会网络或经济模型的控制策略等。文献[7]考虑节点间距离、交互情况和查询相似度等提出了一种用户评价机制, 根据在线时间达到的满意度来减少“搭便车”行为, 但该方法对长期在线的成员而言存在不公平性。文献[8]利用社区对成员的期望值和成员实际共享的偏差来遏制“搭便车”节点, 以提高社区服务质量。但在 P2P 中, 很难预测成员的期望值或判断其合理性。

本文以半分布式 P2P 架构为研究对象, 利用节点间的语义关系构建语义社区, 以提高资源组织的合理性和定位准确性。具有一定组织结构的社区有助于对 P2P 中的“搭便车”行为进行定量分析, 因此, 本文机制通过遏制“搭便车”节点数量, 使所有节点可以公平、高效地利用并共享 P2P 中的资源, 从而动态演化社区结构, 使 P2P 网络具有更好的健壮性和可用性, 促进了 P2P 中资源的高效定位。

2 P2P 语义社区模型

P2P 语义社区是指拥有共同兴趣的节点集合, 形式化定义如下:

基金项目: 国家自然科学基金资助项目(60472093); 山西省自然科学基金资助项目(2006021015); 山西省教育厅高校科技开发基金资助项目(2007111); 太原市大学生创新基金资助项目(07010726)

作者简介: 王莉(1971-), 女, 副教授、博士研究生, 主研方向: Web 智能, 知识通信, 软件体系结构; 胡高霞, 硕士研究生

收稿日期: 2008-08-09 **E-mail:** wangli@tyut.edu.cn

定义1(P2P语义社区) 语义社区 SC (Semantic Community) 是一个四元组, 即 $DVC = \langle KP, CKP, SF, \theta \rangle$ 。其中, KP (Knowledge Peers) = $\{n_i, i=1, 2, \dots, m\}$, ($m=|KP|$), 表示 P2P 系统中所有知识节点集; CKP (Community Knowledge Peers) = $\{n_i, i=1, 2, \dots, m\}$, ($m=|CKP|$), 表示经过语义相似度计算后满足语义相似度阈值 θ 的构成语义社区的知识节点集; SF (Similarity Function) = $\langle SF: x, y \rightarrow z \rangle (x, y \in KP, z \in \{0, 1\})$, 表示语义相似度计算函数, 用于计算知识节点间的语义相似度; $\theta \in \{0, 1\}$ 表示检验知识节点是否属于同一语义社区的语义相似度阈值。语义社区的结构如图 1 所示。

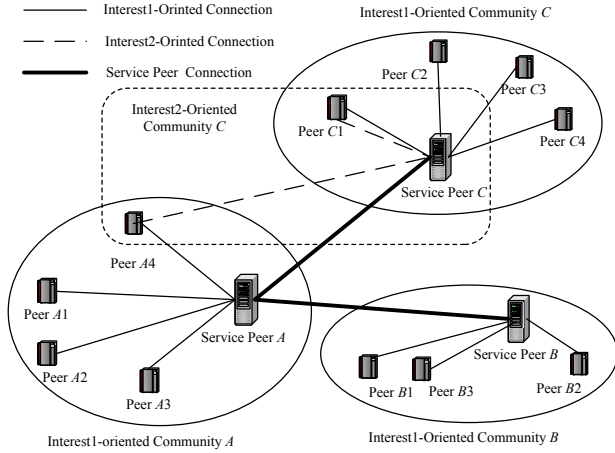


图 1 语义社区的结构

3 语义社区的形成

本文所研究语义社区构建在半分布式 P2P 结构之上, 由超级节点 SP (Service Peer)与普通节点 P (Peer)构成。超级节点上存储了社区中普通节点的信息, 发现算法仅在超级点之间转发, 超级点再将查询请求转发给适当的普通节点。

社区内的活动有 2 类: (1)社区成员间的交互, 即成员为其他成员提供服务, 或接受其他成员服务的行为; (2)社区成员在结束交互后, 对交互活动给出评价反馈(rating)的行为。

3.1 语义社区的形成算法

社区成员具有 3 大基本属性, 即本体知识库(ontology knowledge base)、信任度(rating trust)和活跃强度(active strength)。本体知识库描述节点兴趣和资源的语义信息。信任度是用户对资源的满意程度。活跃强度表示社区成员在社区中的活跃状态, 与成员参与社区活动的行为相关。社区的形成流程如图 2 所示。

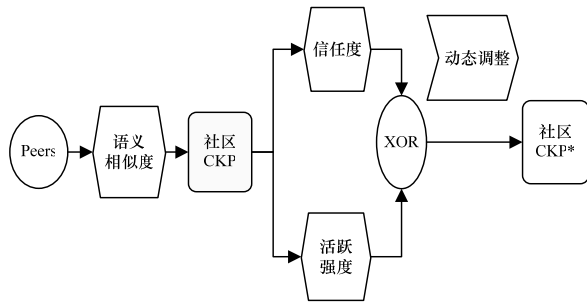


图 2 语义社区的形成流程

在社区的形成过程中, 语义相似度度量很重要。

$P = \{ \langle T_i, \lambda_i \rangle, i=1, 2, \dots, m \}$ 描述 Peer 的语义信息, 其中, T_i 是 Peer 中存储的资源被归入应用本体的类别; λ_i 是被归入 T_i 的资源数量在 Peer 资源集中的统计权重, λ_i 的计算式为

$$\lambda_i = \frac{N_i}{\sum_{j=1}^{|P|} N_j} \quad (1)$$

其中, N_i 是 Peer 中以 T_i 为主题的资源数量; $\sum_{j=1}^{|P|} N_j$ 是 Peer 中的资源总数。

定义 1 中的相似度计算函数 $SF^{[3]}$ 为

$$SF(p_1, p_2) = \sum_{j=1}^{|P_1|} \sum_{i=1}^{|P_2|} [sim(T_i, T_j) \times (\lambda_i \times \lambda_j)] \quad (2)$$

基于上述语义相似度度量方法, 本文给出语义社区生成算法, 具体描述如下:

输入 源 $SP(P_i)$ 的语义描述和 TTL

输出 当前 $P(P_j)$ 的语义索引表

```

while(true) do
    监听网络上的邀请消息;
    if (监听到来自  $P_i$  邀请信息)
        设置  $P_i.TTL = P_i.TTL - 1$ ;
        if ( $SF(P_j, P_i) > Threshold$ )
            将  $P_i$  的语义描述加入  $P_j$  的语义索引表;
             $P_j$  给  $P_i$  发送接受消息, 同意成为该语义社区的成员;
        end if
        if ( $P_i.TTL > 0$ )
             $P_j$  将邀请信息继续转发给它的邻居节点
        end if
    end if
end do

```

社区成员通过超级节点共享信息, 也可以根据语义相关性和信任度推荐其他节点加入社区, 非社区成员也可以通过发送申请消息加入社区。语义社区的基本原则是动态地适应 P2P 网络的拓扑, 使共享相同信息的节点可以形成连接良好的语义社区。

3.2 语义社区的动态调整

为了真正达到 P2P 模式提倡的协作共享理念, 遏制“搭便车”节点数量, 本文提出基于用户评价信任和活跃强度的社区动态调整机制。

定义 2(信任) 社区中的信任是指社区内其他成员对某一成员在社区交互中体现出的服务能力的总体评价, 成员 P_j 的信任记作 $Trust^n = \langle IR_{ij}, \alpha \rangle$ 。 IR_{ij} 是成员 P_i 对 P_j 在交互中体现出的服务能力的评价, $IR_{ij} \in [1, R]$ ($R \in N$) 为最高评价等级数, 1 为最低评价。 α 是 P_i 对 P_j 评价的准确率。 n 表示节点 j 被评价的次数, 即节点 j 提供服务的次数。

定义 3(活跃强度) 成员活跃强度用于表征社区成员的交互历史和活跃状态, 与成员的交互历史以及成员对外提供服务的频度有关。成员 P_i 的活跃强度记为 A_i 。

$$A_i = n_i + r_i \quad (3)$$

其中, n_i 是节点 i 提供服务的次数; r_i 是节点 i 使用服务的次数。

在社区内, 当节点 P_i 和 P_j 进行信息交互时, 假设 P_i 向 P_j 发出查询请求, 如果信息交互成功, 则 P_i 和 P_j 的活跃强度 A_i 和 A_j 都自动增加 1。 IR_{ij} 表示节点 P_i 对 P_j 关于交互完成的评价, 每次节点 P_i 和 P_j 发生交互, P_i 对 P_j 都会给一个评价, 则基于评价的节点 j 的信任度为

$$Trust^n = \alpha \times Trust^{n-1} + (1 - \alpha) \times IR_{ij} \quad (4)$$

为防止恶意评价, 从而使评价更加合理, 设置动态调节因子 α , 其值越小, 表明本次评价越合理, 即所占比例越高。

$$\alpha = \frac{Trust^{n-1} - IR_{ij}}{R} \quad (5)$$

$Trust^0$ 表示初始化的信任度，为了体现资源拥有量、节点资源和社区主节点相似度对信任度值的正向影响以及信任度的规约，信任度初始值定义为

$$Trust^0 = \left[\frac{2(R-1)}{\pi} \arctan\left(\sum_{j=1}^{j \leq |P_i|} N_j - 1\right) \times SF(P_i, SP) + 1 \right] \quad (6)$$

其中， $Trust^0 \in [1, R]$; $1 \leq \sum_{j=1}^{j \leq |P_i|} N_j < +\infty$; $0 \leq SF(P_i, SP) \leq 1$ 。

当节点没提供评价时， $IR_{ij}=0, \alpha=1$ ，则 $Trust^n = Trust^{n-1}$ ，即节点信任度不变，但 n 会自动增加，不会影响判断该节点是否为“搭便车”节点。

在社区内，设置信任度阈值 $Trust_threshold$ 和满意度阈值 $Free_threshold$ 。如果 $\forall i, i \in CKP_j, Trust_i < Trust_threshold$ ，则表示节点 i 共享资源质量很低，社区将给出警告。如果在一段时间内仍没有提高，则将节点在社区中删除。若 $\forall i, i \in CKP_j, \Theta = \frac{n_i}{A_i} < Free_threshold$ ，即节点 i 更多地使用其他节点资源，而很少贡献自己的资源，属“搭便车”节点，社区将会给出警告，如果在一段时间内仍没改善，则将节点在社区中删除。因此，本文提出的动态调整机制能有效遏制了“搭便车”节点，并能提高社区共享资源质量以及社区的健壮性和可用性。

由于社区的动态调整机制基于用户评价，因此用户评价的准确性决定了机制的有效性，尤其是防止恶意节点评价的误差率。在本文算法中，对信任的评价是一个长时间效应的评价。在考虑已有评价和最近时间评价的基础上，证明该评价算法准确性较高且误差率低，能有效防止恶意节点评价带来的影响，具体如下：

证明 设 $Trust^{n-1}=X, IR_{ij}=Y, \theta = X - Y$ ，则 $\alpha = \frac{\theta}{R}, Trust^n = \frac{\theta(X-Y)}{R} + Y$ 。当评价较准确，即 $\theta \approx 0$ 时，可得 $Trust^n = Y = X = Trust^{n-1}$ ，说明本次评价所得信任度和已有评价的误差率为 0。当存在最坏情况的恶意评价，即 $\theta = R - 1$ 时，可得

$$Trust^n = \frac{R-1}{R} \times (X - Y) + Y = X - \frac{X - Y}{R}$$

由于 $X - Y$ 的取值范围为 $0 \leq X - Y \leq R - 1$ ，则

$$0 \leq \frac{X - Y}{R} \leq \frac{R - 1}{R}$$

因此 $Trust^n = Trust^{n-1} - \frac{X - Y}{R}$ ，说明即使有恶意评价，对成员的真实信任度的影响不会超过一个信任评价等级 $\frac{R-1}{R}$ 。可见，本文设置的动态调节因子 α 能有效降低恶意评价的影响。

4 实验

根据提出的语义社区生成及演化算法进行初步实验，并在 Pajek 上搭建起系统模拟环境。初始化 1 000 个节点，每个节点可以有几类不同的兴趣，并分别拥有 1 个或 1 个以上资源。虽然基于兴趣的应用本体包含的类别很多，但每个节点感兴趣的领域一般只有几个，且通常有一个集中研究的领域。在测试中假设每个 Peer 上有 3 个兴趣，并保证其中一个兴趣的权重较高(50%)。为了模拟语义拓扑，让每个节点向网络广播 TTL 等于 3 的语义描述信息，节点间按相似度阈值 0.5 生成语义社区。具体参数设置见表 1。节点监听网络上的邀请信息，如果收到邀请信息，则计算语义相似度，并根据相似

度阈值决定是否加入该社区。

表 1 实验的参数设置

参数	描述	值
Peers	Number of nodes in the network	1 000~6 000
degree	Average degree	3~4
interest	Number of interest	10
max_redource	Number of resource each peer	1~100
similarity_threshold	Threshold of similarity	0.5
TTL	TTL for clustering	3
R	Trust degree	10
α_{opt_res}	Exponent α of resource distribution	1.2
Trust_threshold	Threshold of trust	4
Free_threshold	Threshold of free-rider	0.2

实验步骤如下：

(1)建立节点语义模型 $P = \{ \langle T_i, \lambda_i \rangle, i=1, 2, \dots, m \}$ 。

(2)建立兴趣应用本体，它包括 3 个模块：Life, Career, Entertainment，其高层组织结构如图 3 所示。

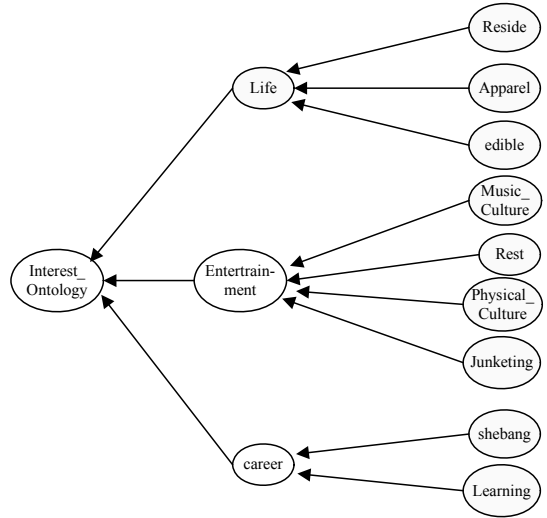


图 3 兴趣应用本体的高层组织结构

(3)语义匹配。根据节点的语义描述信息和应用本体计算节点间的语义匹配度。例如，有 2 个节点 $P_1 = \{ \langle a, 0.2 \rangle \langle b, 0.1 \rangle \langle c, 0.7 \rangle \}$, $P_2 = \{ \langle a, 0.1 \rangle \langle b, 0.3 \rangle \langle c, 0.6 \rangle \}$ ， a, b, c 映射为应用本体中的 3 个概念，其深度分别为 $A = depth(a), B = depth(b), C = depth(c)$ ， a 和 b 的共同父节点中深度最大的节点 w ，其深度为 $W = depth(w)$ ，则 P_1 和 P_2 的语义相似度 $SF(P_1, P_2) = 0.757$ 。

基于上述语义匹配，运用社区形成算法，可以得到语义社区。经过一段时间的交互，社区创建者 SPB 发现 PB24 的信任度为 $2 \ll 4$ ，说明 PB24 共享的资源质量很差，且已经给出警告但仍没有提高，因此，会把 PB24 从社区中删除，并发送信息通知社区中的其他成员。社区创建者 SPB 发现 PB26 的活跃强度 A_i 远大于 PB26 提供服务的次数 n_i ，表示 PB26 是“搭便车”节点，若给出警告后仍然没改善，则把 PB26 从社区中删除，并发送信息通知社区中的其他成员。很多节点为了享用社区资源会提高自己的共享质量和提供服务的次数。

在基于社区的路由检索中，“搭便车”节点会消耗系统处理时间，增加路由跳数，因此，必须剔除“搭便车”节点。

根据实验结果可知，本文提出的语义社区构建与演化机制使真正具有相同兴趣和共享理念的节点聚类在一起，从而有效地共享资源。与文献[3-4]相比，本文考虑语义关系，提高了资源组织和定位的准确性，且遏制了“搭便车”节点的

(下转第 16 页)