

# 情景感知的 P2P 移动社交网络构造及发现算法

曹怀虎<sup>1)</sup> 朱建明<sup>1)</sup> 潘 耘<sup>2)</sup> 李海峰<sup>1)</sup>

<sup>1)</sup>(中央财经大学信息学院 北京 100081)

<sup>2)</sup>(中国传媒大学计算机学院 北京 100024)

**摘 要** 如何通过情景感知,获得用户的行为特征,从而自动发现潜在的社会关系,是移动社交网络重要的研究内容之一.该文提出了情景感知的移动 P2P 社交网络系统架构、聚合模型及发现算法,将用户的位置信息、环境特征、运动轨迹等引入到聚合算法中,智能地聚合成潜在的 P2P 社交网络,根据用户需求自主发现匹配的社会关系,避免了社交活动的盲目性和随意性.最后对该方案和算法进行了理论分析及实验验证,结果表明该文所提出的方案和算法具有较高的响应速度、准确率及用户满意度.

**关键词** 移动社交网络;情景感知;聚合模型;社交发现;P2P;移动互联网

**中图法分类号** TP393 **DOI 号:** 10.3724/SP.J.1016.2012.01223

## Context-Aware P2P Mobile Social Network Structure and Discovery Algorithm

CAO Huai-Hu<sup>1)</sup> ZHU Jian-Ming<sup>1)</sup> PAN Yun<sup>2)</sup> LI Hai-Feng<sup>1)</sup>

<sup>1)</sup>(School of Information, Central University of Finance and Economics, Beijing 100081)

<sup>2)</sup>(School of Computer, Communication University of China, Beijing 100024)

**Abstract** It is one of important topics in social network research how to get the user's behavior feature and discover actively potential social relations by automatic context-aware. This paper presents a context-aware mobile P2P social network framework, aggregation model and the discovery algorithm. The user's location information, environmental characteristics etc. are introduced to the aggregate algorithm, which aggregate intelligently to potential P2P social network and discover the suited relationship according to user's demand. Thereby, the social blindness and randomness are avoided. The theoretical analysis and experimental results show that the proposed approach and the algorithm have a higher response speed, accuracy and user's satisfaction.

**Keywords** mobile social network; context-aware; aggregation model; social discovery; P2P; mobile Internet

## 1 引 言

社交网络(Social Networks)起源于网络交友,

用户使用社交网络服务来组织、维护现有的社会关系,发现新的社会关系,从而作为一个社会人在网络上呈现自己,进行相关的社交活动<sup>[1-2]</sup>.随着智能手机、平板电脑、笔记本电脑等便携移动设备的普及和

收稿日期:2011-12-09;最终修改稿收到日期:2012-05-11. 本课题得到国家自然科学基金(61103198,61100112,60970143)、教育部科学技术重点项目基金(109016)、教育部人文社会科学研究青年基金(11YJCZH006)、中央财经大学学科建设基金(CUFEIE201107)资助. 曹怀虎,男,1977年生,博士,副教授,主要研究领域为网络体系结构、移动互联网、网络计算. E-mail: caohhu@163.com. 朱建明,男,1965年生,博士,教授,主要研究领域为网络生存性、网络安全等. E-mail: tyzjm65@163.com. 潘 耘,女,1974年生,博士,副教授,主要研究领域为网络体系结构、移动互联网. E-mail: pany@cuc.edu.cn. 李海峰,男,1979年生,博士,讲师,主要研究领域为数据挖掘、数据管理. E-mail: mydlhf@126.com.

移动通信技术的发展<sup>[3]</sup>,具有高度分布、自治、拓补动态变化等特性的移动无线网络正逐渐成为社交网络的基础,传统依靠 Web 支撑、集中式的社交网络,很难适应目前动态、异构的分布式移动网络环境,也无法满足自主发现、点对点共享数据等新需求,同时,人们已不再满足于单纯的交友活动,希望能够根据用户所处的环境、个人爱好、行为特征等情景信息,主动、智能地发现潜在的社会关系,进而获得更多、更有价值的娱乐、交流、合作机会<sup>[4]</sup>.因此,研究如何利用情景感知信息,构造潜在的社交网络,根据用户的兴趣和偏好,自主地发现相关度较高的社会关系,这对于社交网络的进一步发展具有重要的意义.

在移动环境下,只有通过情景感知,社交网络用户才能够了解彼此的状态,从而自主发现潜在的社会关系.本文试图将用户的位置信息、环境特征、行为特征、兴趣偏好等作为情景感知的内容,通过智能聚合,形成潜在的 P2P 社交网络,根据用户的需求自主发现潜在的社会关系.本文主要研究路线如图 1 所示,首先进行 P2P 移动社交网络的拓补构造模型、系统架构、网络聚合模型的研究;在此基础上提出情景感知的 P2P 移动社交网络聚合算法和社交发现算法;最后,进行算法性能的理论分析和仿真实验测试分析.

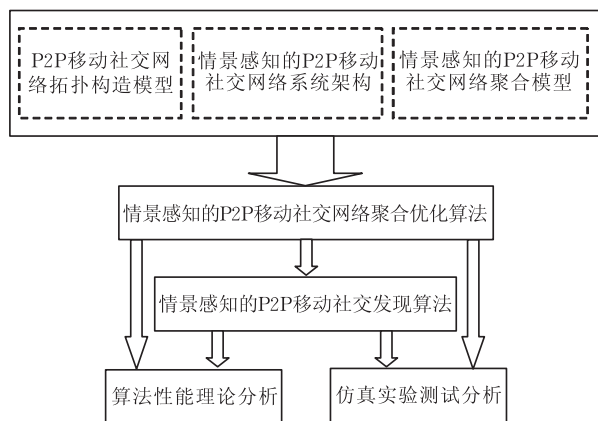


图 1 主要研究路线

本文第 1 节介绍研究背景、研究路线、文章总体结构;第 2 节对社交网络及情景感知技术的研究现状进行回顾和分析;第 3 节讨论移动覆盖 P2P 社交网络的覆盖(overlay)拓补结构和系统架构;第 4 节从相似度函数的定义入手,建立移动 P2P 社交网络的聚合模型;第 5 节提出情景感知的移动 P2P 社交网络的聚合及发现算法;第 6 节对算法性能进行理论分析;第 7 节对所提出的方案和算法进行仿真实验及对比分析;最后,第 8 节给出本文的结论和未

来的工作.

## 2 相关工作

社交网络是对真实社交活动的虚拟反映,既来源于现实的社交关系,又具有不同于现实交往的新特征.这一领域的主要研究方向及成果有:(1)对社交网络全网关系图的拓补性质及信息传播特性的研究<sup>[5]</sup>;(2)基于社交网络数据挖掘的用户行为研究<sup>[6]</sup>;(3)基于 Web2.0 的社交网络流量特点及群组结构的研究<sup>[7]</sup>;(4)社交网络隐私策略、可信性等研究<sup>[8]</sup>.以上研究主要是基于集中式的社交网站,关于移动模式下的 P2P 社交网络的研究目前还较少.

情景感知的目的是利用人机交互或传感器提供给计算设备关于人和环境的信息,并让计算设备给出相应的决策.这里的情景是指一系列的环境特征,这个环境特征决定了用户的行为.包括计算情景、用户情景、物理情景、时间情景、社会情景等.情景感知实际上反映了应用从“以计算机为中心”到“以人为中心”的转变,无论是桌面计算机还是移动设备,都应当围绕用户环境自适应地提供个性化的服务.关于情景感知的研究主要有情景的获取、情景的表示、情景的使用等<sup>[9]</sup>.

为了提高社交网络的服务水平,研究人员已经开始将情景感知技术引入到社交网络中.目前,主要是利用搜索引擎、数据挖掘技术,通过感知用户的属性特征、使用习惯等来提升社交网络的服务.现有的研究大多是基于 Web 的,如 Schmidt 等人<sup>[10]</sup>通过整合用户浏览 Web 页面时的信息,感知个人页面浏览者的行为特征. Acquist 等人<sup>[11]</sup>以 Facebook 为例,探讨了社交网络中的感知、信息共享和隐私问题,指出用户对社交网络性质、结构和个人隐私设置等内容感知的重要性. Matsuo 等人<sup>[12]</sup>基于搜索引擎建立了一个社交网络提取系统“POLYPHONET”,通过 GOOGLE 搜索引擎实现了关系的提取、群体结构的检测. Mika<sup>[13]</sup>建立了一个名为 Flink 的系统,从网页、个人资料等信息资源提取个人信息,并进行基于语义的推理,实现了在线社会网络的提取、聚集、分析和可视化.

从以上研究成果可以看出:结合情景感知技术来研究社交网络是主要研究方向,目前的研究主要基于 Web 历史数据,基于实时、移动的用户情景感知的社交网络研究还很少.在现实的社会交往过程中,用户通过面对面的交互,可以从相互的接触和非正式的交流中收集所需的情景信息.但是在社交网

络环境下,或者不提供情景感知信息,或者非常有限,这样导致社交网络用户对网络内其它用户的情景感知程度普遍较低,丧失了很多互动、合作、参与的机会.目前,移动设备将拥有越来越多的传感功能,包括视频、语音、GPS、重力感应、加速度感应、六轴向陀螺、红外感应等,获得节点的感知信息已不再是困难和昂贵的事情.因此,本文将通过引入更多的情景感知信息,如 GPS、轨迹、物理环境、兴趣偏好等,来探讨 P2P 移动社交网络的构造方法及潜在社交关系的发现算法.

### 3 移动 P2P 社交网络覆盖拓扑结构和系统架构

移动 P2P 社交网络拓扑架构如图 2 所示,各种移动设备接入的方式不同,如手机用户可能通过 3G 基站接入 Internet,平板电脑、笔记本电脑可能更多的是通过 WiFi 接入 Internet.它们之间形成了一个底层的混合通信网络,同时这些移动设备之间还可以通过蓝牙技术进行短距离共享数据.这些移动设备通过运行授权的客户端软件,在底层混合通信网络的基础上,通过情景感知,自主聚合形成一个潜在的虚拟 P2P 覆盖社交网络(Overlay Social Network),在这个虚拟 P2P 覆盖社交网络的基础上,可以进一步实现主动的社交发现服务,向用户推荐满足其要求的社交关系.

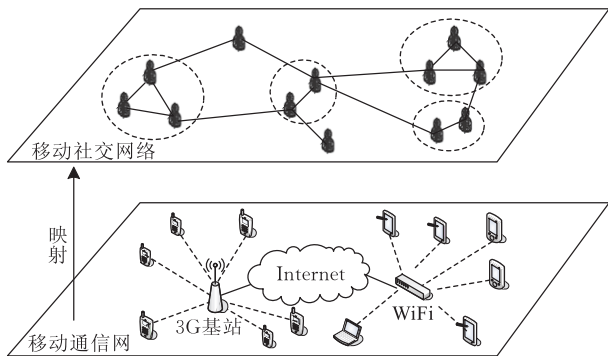


图 2 移动 P2P 社交网络拓扑结构

移动 P2P 社交网络的系统架构如图 3 所示,移动客户端具有情景感知、自动聚合、社交发现、社交维护等功能.服务器端具有注册管理、隐私控制、行为记录、用户偏好等功能.用户通过向服务器注册,获得授权的客户端软件,用户一旦运行该软件,在硬件的支持下,就可以获得情景感知的数据,如 GPS 位置、温度、光照、速度等,当然这些都必须是在隐私控制许可的前提下,然后在一定条件(如某地理位置范围内)下,运行分布式的聚合算法<sup>[14]</sup>,自动形成覆盖

的社交网络拓扑结构,也就是潜在的社交网络.之后,可以根据用户的兴趣、偏好等,推荐匹配的社交关系,用户可以决定和哪些人交往,进行自己的社交活动(建立好友目录、开始对话、共享数据等).

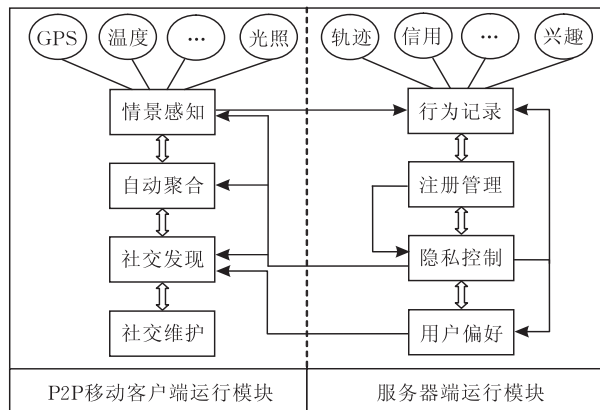


图 3 移动 P2P 社交网络系统架构

### 4 移动 P2P 社交网络的聚合模型

根据上一节所述,移动 P2P 社交网络首先要根据用户情景感知的信息进行自动聚合,形成潜在的覆盖社交网络.如果用图论来表示移动节点的聚合问题(Mobile Node Clustering, MNC),则一个混合网络可以表示为无向图  $G=(V, E, W_G)$ ,其中顶点集  $V$  为移动节点的集合,边集  $E$  为移动节点间可能的链路的集合,  $W_G(u, v) (u, v \in V)$  表示移动节点  $u, v$  间聚合矢量距离(相似度),移动节点的聚合问题就是将移动节点在其矢量距离空间上划分为不同的子图  $g_1(v_1, e_1, w_1), g_2(v_2, e_2, w_2), \dots, g_M(v_M, e_M, w_M)$  ( $M$  为聚集数,  $v_i, e_i, w_i$  分别为子图的节点、链路、节点间的矢量距离集合,  $i=1, 2, \dots, M$ ),且满足下面的条件:

- ①  $v_1 \cap v_2 \cap \dots \cap v_k = \emptyset$ ;
- ②  $v_1 \cup v_2 \cup \dots \cup v_k = V$ ;
- ③  $w_i(u, v) < W_G(u, v'), (u \in g_i, v \in g_i, v' \notin g_i, i=1, 2, \dots, M)$

(1)

聚合的方法是通过每个移动节点之间相似度的计算,将节点组织成若干聚集,相似度依赖于节点特征矩阵.划分聚集方法可以描述为:给定  $L$  维空间的  $N$  个节点  $X = \{x_1, x_2, \dots, x_N\}$ ,将  $N$  个节点分成  $M$  个节点子集  $C_1, C_2, \dots, C_M$  (聚心分别为  $y_1, y_2, \dots, y_M$ ),使得在一个聚集中的节点彼此尽可能相似,一般有  $C_i \cap C_j = \emptyset (i \neq j)$ ,且  $\bigcup_{i=1}^M C_i = X$ .划分聚集的方法是将每一聚集描述为一个代表元,对于每一节点,依据该节点与某一聚集代表元的相似程度,决定该节点属于那一个聚集.

本文的节点聚合优化模型可归结为下面的求解

极小化问题:

$$\min D(\mathbf{X}, \mathbf{Y}) = \min \left( \sum_{i=1}^N \sum_{j=1}^M \left( \alpha \sqrt{\sum_{l=1}^L (x_{il} - y_{jl})^2} + \beta \frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\| \times \|\mathbf{Y}_j\|} + \gamma \frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\|^2 + \|\mathbf{Y}_j\|^2 - \mathbf{X}_i^T \mathbf{Y}_j} \right)^2 \right) \quad (2)$$

式中:  $\mathbf{X}_i = \{x_{i1}, x_{i2}, \dots, x_{iN}\}$  为节点  $x_i$  的特征向量 (具有  $L$  个分量),  $N$  为节点总数,  $\mathbf{Y}_j = \{y_{j1}, y_{j2}, \dots, y_{jL}\}$  为聚心节点  $y_j$  的特征向量 (具有  $L$  个分量),  $M$  为聚心节点总数,  $\sqrt{\sum_{l=1}^L (x_{il} - y_{jl})^2}$  表示节点  $x_i$  与节点  $y_j$  的环境相似度, 采用欧氏距离表示,  $\frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\| \times \|\mathbf{Y}_j\|}$  表示运动轨迹相似度, 采用几何角度相似性函数表示,  $\frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\|^2 + \|\mathbf{Y}_j\|^2 - \mathbf{X}_i^T \mathbf{Y}_j}$  表示用户历史偏好相似度, 采用 Tanimoto 测度表示其特征共性.  $\alpha, \beta, \gamma$  表示权重因子, 在实际应用中可以根据实际需要调整其数值. 由于式(2)为一非凸优化问题, 目前还没有十分有效的办法求出全局最优解, 下一节将给出一种可行的求解算法.

## 5 移动 P2P 社交网络聚合及发现算法

### 5.1 基于确定性退火技术的移动 P2P 社交网络聚合算法

确定性退火技术是美国加州理工学院的 Rose  $F(y_1, y_2, \dots, y_M, \delta) =$

$$\begin{cases} \frac{1}{M} \sum_{i=1}^N \sum_{j=1}^M \left( \alpha \sqrt{\sum_{l=1}^L (x_{il} - y_{jl})^2} + \beta \frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\| \times \|\mathbf{Y}_j\|} + \gamma \frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\|^2 + \|\mathbf{Y}_j\|^2 - \mathbf{X}_i^T \mathbf{Y}_j} \right)^2, & \delta = 0 \\ -\frac{1}{\beta} \sum_{i=1}^N \ln \sum_{j=1}^M e^{-\delta \left( \alpha \sqrt{\sum_{l=1}^L (x_{il} - y_{jl})^2} + \beta \frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\| \times \|\mathbf{Y}_j\|} + \gamma \frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\|^2 + \|\mathbf{Y}_j\|^2 - \mathbf{X}_i^T \mathbf{Y}_j} \right)^2}, & 0 < \delta < +\infty \\ \sum_{i=1}^N \sum_{j=1}^M \left( \alpha \sqrt{\sum_{l=1}^L (x_{il} - y_{jl})^2} + \beta \frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\| \times \|\mathbf{Y}_j\|} + \gamma \frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\|^2 + \|\mathbf{Y}_j\|^2 - \mathbf{X}_i^T \mathbf{Y}_j} \right)^2, & \delta = +\infty \end{cases} \quad (4)$$

式中:  $\mathbf{X}_i = \{x_{i1}, x_{i2}, \dots, x_{iL}\}$  为节点  $i$  的特征向量 (具有  $L$  个分量),  $y_1, y_2, \dots, y_M$  分别代表聚集  $C_1, C_2, \dots, C_M$  的聚心节点,  $\mathbf{Y}_j = \{y_{j1}, y_{j2}, \dots, y_{jL}\}$  为聚心节点  $y_j$  的特征向量 (具有  $L$  个分量,  $j=1, 2, \dots, M$ ),  $N$  为节点总数,  $M$  为聚心节点总数, 其它参数含义同式(2).

$x_i \in C_j$  的概率  $P$  如下:

$P(x_i \in C_j) =$

$$\frac{e^{-\delta \left( \alpha \sqrt{\sum_{l=1}^L (x_{il} - y_{jl})^2} + \beta \frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\| \times \|\mathbf{Y}_j\|} + \gamma \frac{\mathbf{X}_i^T \mathbf{Y}_j}{\|\mathbf{X}_i\|^2 + \|\mathbf{Y}_j\|^2 - \mathbf{X}_i^T \mathbf{Y}_j} \right)^2}}{\sum_{k=1}^M e^{-\delta \left( \alpha \sqrt{\sum_{l=1}^L (x_{il} - y_{kl})^2} + \beta \frac{\mathbf{X}_i^T \mathbf{Y}_k}{\|\mathbf{X}_i\| \times \|\mathbf{Y}_k\|} + \gamma \frac{\mathbf{X}_i^T \mathbf{Y}_k}{\|\mathbf{X}_i\|^2 + \|\mathbf{Y}_k\|^2 - \mathbf{X}_i^T \mathbf{Y}_k} \right)^2}} \quad (5)$$

博士于 1990 年首先提出的, 它是按自然法则计算的一个重要分支. 对于求解以下极小化问题非常有效:

$$\min E = E(x) \quad (3)$$

这里  $x$  可以是连续的、离散的或混合的,  $E(x)$  可看作是某一系统的能量. 确定性退火技术将极小化问题式(3)看作是求解一物理系统能量极小的状态. 首先构造一自由能函数  $F(x, T)$ , 这里  $x$  是函数变量,  $T$  表示温度. 由以上分析知, 在某一温度下, 系统状态的变化总是朝着自由能减少的方向进行, 当系统达到平衡态时自由能函数达到极小. 文献[15]证明了当  $F(x, T)$  为连续映射时, 它的全局极小点  $x_{\min}(T)$  为  $T$  的连续映射. 设  $T \rightarrow \infty$  时,  $F(x, T)$  的全局最优点极易求出, 而  $F(x, 0) = E(x)$ . 确定性退火技术, 在每一温度  $T$ , 以系统在  $T = T + \Delta T$  时自由能函数极小的状态  $x_{\min}(T + \Delta T)$  作为初始点, 通过求解  $\min F(x, T)$  的极小点来模拟系统达到平衡态的过程. 随着  $T$  的减小,  $F(x, T)$  的全局极小点不断变化, 当  $T$  的变化  $\Delta T$  很小时, 可认为  $x_{\min}(T)$  位于  $x_{\min}(T + \Delta T)$  所在的局部极小区域内, 故可以用  $x_{\min}(T + \Delta T)$  作为初始点求解  $F(x, T)$  的极小值. 当  $T$  连续减小速度合理时, 可以认为  $\lim_{T \rightarrow 0} x_{\min}(T)$  为问题(2)的全局极小点.

对于本文的社交网络聚合问题  $\min D(\mathbf{X}, \mathbf{Y})$ , 利用确定性退火技术, 定义自由能函数:

式中:  $x_i$  表示某节点 ( $i=1, 2, \dots, N$ ),  $C_j$  为某一聚合子集,  $j, k=1, 2, \dots, M$ , 其它参数含义同式(4).

当  $\delta=0$  时,  $F(y_1, y_2, \dots, y_M, \delta)$  关于  $(y_1, y_2, \dots, y_M)$  为连续可微的凸函数, 按传统的优化方法极易求出全局最优解<sup>[16-17]</sup>; 当  $\delta \in [0, +\infty]$  时,  $F(y_1, y_2, \dots, y_M, \delta)$  关于  $y_1, y_2, \dots, y_M$  为连续可微,  $F(y_1, y_2, \dots, y_M, \delta)$  的极小点与极小值为  $\delta$  的连续映射. 在  $F(y_1, y_2, \dots, y_M, \delta)$  的极小点处, 一阶必要条件满足, 即

$$\frac{\partial F(y_1, y_2, \dots, y_M, \delta)}{\partial Y_j} = 0,$$

( $j=1, 2, \dots, M, M$  为聚集数)

将式(4)代入上式得

$$y_j = \frac{\sum_{i=1}^N x_i P(x_i \in C_j)}{\sum_{i=1}^N P(x_i \in C_j)} \quad (6)$$

其中  $P(x_i \in C_j)$  可由式(5)确定.

当  $\delta=0$  时,有

$$P(x_i \in C_j) = \frac{e^{-\delta \left( \alpha \sqrt{\sum_{l=1}^L (x_{il} - y_{jl})^2} + \beta \frac{x_i^T y_j}{\|x_i\| \times \|y_j\|} + \gamma \frac{x_i^T y_j}{\|x_i\|^2 + \|y_j\|^2 - x_i^T y_j} \right)^2}}{\sum_{k=1}^M e^{-\delta \left( \alpha \sqrt{\sum_{l=1}^L (x_{il} - y_{kl})^2} + \beta \frac{x_i^T y_k}{\|x_i\| \times \|y_k\|} + \gamma \frac{x_i^T y_k}{\|x_i\|^2 + \|y_k\|^2 - x_i^T y_k} \right)^2}} = \frac{1}{M} \quad (7)$$

式中,  $j=1, 2, \dots, M$ ,  $M$  为聚集个数.

从而,  $y_j = \frac{1}{M} \sum_{i=1}^N x_i$  为  $\delta=0$  时  $F(y_1, y_2, \dots, y_M, \delta)$  的全局最优点(因为对于凸规化,局部极小点必为全局极小点),  $y_j$  为  $\mathbf{X}$  中所有点的加权平均.

由式(6),可以得到求最优  $y_i$  的迭代公式:

$$y_j^{(k+1)} = \frac{\sum_{i=1}^N x_i p(x_i \in C_j^{(k)})}{\sum_{i=1}^N p(x_i \in C_j^{(k)})} \quad (8)$$

在每一  $\delta$ (对应温度  $T$ ,  $T \propto \frac{1}{\delta}$ ), 可按迭代式(8)来模拟聚合系统的平衡态, 本文下面将给出利用确定性退火技术的分布式移动 P2P 社交网络聚合 DCMPDA 算法(Distributed Clustering for Mobile P2P Social Network Deterministic Annealing algorithm).

由于影响移动节点聚合的因素较多(如地理位置、运动轨迹、个人偏好等), 用符号  $F_i(x_{i1}, x_{i2}, \dots, x_{iL})$  来表示这些感知信息(这里  $i=1, 2, \dots, N$ ,  $N$  为节点数,  $L$  表示感知因素个数, 即特征向量维数), 其中  $F_i$  为节点  $x_i$  的影响因素的集合,  $x_{iL}$  为  $i$  个节点的第  $L$  个影响因素. 由于  $x_{i1}, x_{i2}, \dots, x_{iL}$  中有些不是基数变量, 比如说序数变量如移动速度、温度高低、各种先验指标、用户感受等. 为了给聚合提供合理依据, 需要从这些数据中提取出有效的可比较数据, 必须将这些半定性、半定量的问题转化为定量的数据. 取某个因素集合  $F_i(x_{i1}, x_{i2}, \dots, x_{iL})$ , 采用 AHP(Analytical Hierarchy Process)方法得到对应各个因素  $x_{i1}, x_{i2}, \dots, x_{iL}$  的权系数  $w_i$ , 它满足如下条件:

$$w_i \geq 0, \sum_{i=1}^L w_i = 1 \quad (9)$$

将变量  $y_j$  表示成  $x_{i1}, x_{i2}, \dots, x_{iL}$  的线性组合:  $y_j = w_1 x_{i1} + w_2 x_{i2} + \dots + w_L x_{iL}$ ,  $w_1, w_2, \dots, w_L$  为线

性系数. 这样便求出了对应  $F_i$  的量化可比数据  $y_j$ . 依次类推推出所有  $F_i (i=1, 2, \dots, N)$  对应的量化数据  $y_j$  便获得一个量化数组  $y (y_1, y_2, \dots, y_N)$ .

按确定性退火技术的聚合算法中主要迭代式(8)聚合的时候, 使用的是分布式聚合方式, 采用主从式编程模式, 预先随机聚合, 将式(8)的计算分散到各个子节点上去执行. 下面是并行分布式程序的主要框架.

主结点程序:

1. 先随机聚合, 形成初始聚集;
2. 输入聚集初始数据;
3. 取  $y_i$  的初始值为所有聚集点的中心即  $y_1^0 = y_2^0 = \dots =$

$$y_M^0 = y_j = \frac{1}{M} \sum_{i=1}^N \sum_{l=1}^L w_l x_{il};$$

4. 产生  $Q$  个任务并将要聚合的节点集  $X$  送给各个任务(不妨设  $Q=M$ ), 其中一个留在主结点上, 其它在从结点上;
5.  $\delta = \delta_0 + c\delta_0$ ;
6. 将旧的聚集中心即  $\delta$  送往各个子任务;
7. 依次接收各子任务回送的计算结果构成新的聚集中心  $y_1, y_2, \dots, y_M$ , 若终止准则满足, 则停, 输出结果(聚集数、聚心及对应的  $\delta$ ), 否则重置初值, 对  $\forall i, y_i^0 = y_i, \delta_0 = \delta$  转步 4.

从结点程序:

1. 初始化;
2. 接收主结点数据;
3. 按式(8)进行迭代, 直到迭代终止;
4. 回送计算结果.

## 5.2 移动 P2P 社交网络维护及社交发现算法

### 5.2.1 移动 P2P 社交网络维护

在移动社交网络中, 由于节点的移动性、无线信号不稳定及外界的干扰, 都可能造成节点失效或者退出系统. 因此, 必须提供自适应的系统维护机制. 节点聚合算法结束后, 可以在聚集内推选一个盟主节点和  $B-1$  个候选盟主节点, 用来维护这个覆盖 P2P 拓扑结构, 同时负责和其他聚集之间的交互. 如果有新节点加入, 则按式(2)选择最相似聚集优先加入, 节点退出时, 将告知盟主节点.

假设每个盟主节点失效的平均概率为  $P$ , 则  $B$  个盟主节点都失效的概率为  $P^B$  ( $1$  个为主盟主节点, 其余为备份盟主节点). 因此, 至少有一个盟主节点不失效的概率为  $(1-P^B)$ , 可以根据不同的情况利用这种方法粗略地估计所需的备份盟主节点的数目. 一般来说, 为了减小开销,  $B$  应取一个比较小的值, 这就意味着不能排除聚集中所有盟主节点都失效的情况. 因此, 每个聚集内节点每隔时间  $T_B$  将向自己聚集内的盟主节点发送一个“Hello”消息, 如果  $T_R$  时间后仍未收到聚集盟主节点发送的“alive”消息, 则认为该盟主节点失效, 立即发起盟主节点选举

算法。

因此,首先要解决网络初始化阶段的盟主节点选举问题,然后解决动态情况下,当所有盟主节点失效后,通过何种选举算法在聚集中选举出新节点,作为聚集中新的盟主节点。

本文借鉴了局域网的 IEEE 802.1D 标准,引入了基于生成树(Spanning Tree)的选举算法。即通过先“生长(growing)”,然后再“收缩(shrinking)”生成树的方法,达到从聚集中选举出新盟主节点的目的。

### (1) 选取标准

**定义 1.** 对每个节点  $i$ ,  $T(i)$  为一个非负函数,表示节点  $i$  在社交网络中的会话时间。

**定义 2.** 对每个节点  $i$ ,  $H(i)$  为一个非负函数,表示节点  $i$  的能量、信号强度和计算能力等。

节点间除了在计算能力、能量、信号强度等能力上存在差异外,在会话时间及行为上也存在极端异构的情况。而一个节点作为盟主节点,一方面应有较强的计算能力、能量、信号强度,另一方面更重要的是能为聚集内其它节点提供稳定的服务,以避免系统的震荡。所以本文用  $A(i)$  表示节点  $i$  的能力,  $A(i) = \omega T(i) + H(i)$ ,  $\omega > 1$ ,也即对每个节点,赋予节点会话时间相对较高的权值。从定义可以看出,如果将聚集看作一个无向连通图  $G$ ,其中  $G$  中每个节点赋予一个标识符  $A(i)$ ,则选举算法的目的就是要在  $G$  内选举出  $A(i)$  不减排序中前  $B$  个节点作为新的主盟主节点和备份盟主节点。

### (2) 社交网络初始化阶段的盟主选举

首先考虑网络初始化阶段的盟主选举,可以由系统指定任一个节点  $S$  发动盟主选举, $S$  为选举算法的源节点,选举算法主要包括以下 3 步:

①“生长”以  $S$  为根的生成树。首先, $S$  将向自己邻居节点集合中的每个节点发出选举消息(election message)。而聚集中除了源节点  $S$  之外的其它任意节点  $i$ ,如果没有父节点,则  $i$  将把收到的第 1 个选举消息的发送端作为自己在生成树中的父节点,然后  $j$  将向自己除了父节点之外的邻居节点集合中的所有其它节点发出选举消息,依此类推。

②生成树向根  $S$ “收缩”。如果某个节点  $j$  收到选举消息,并且  $j$  已经有父节点,则节点  $j$  将立即向发送端发出一个应答消息(ACK message),并且在应答消息中加入自己的节点能力信息。而聚集中除源节点  $S$  之外的其它任意节点,一旦收到所有自己发出的选举消息的应答消息后,选出自己以及所有应答消息中节点能力不减排序中前  $B$  个节点,将该

信息加入到应答消息中,然后再向自己的父节点发送该应答消息,依次类推。

③广播新盟主节点信息。最后当选举算法的源节点  $S$ ,也就是生成树的根节点收到所有自己发出的选举消息的应答消息后,将选出自己以及所有应答消息中节点能力不减排序中前  $B$  个节点,就得到了新的主盟主节点和备份盟主节点,并将新的盟主节点的节点标志信息向聚集中所有节点广播(Broadcast message),至此整个网络初始化阶段盟主选举算法结束。

### (3) 盟主节点失效后动态情况下的选举算法

在实际的移动 P2P 社交网络中,可能几个节点同时发现所有盟主节点失效,于是同时发起了盟主节点选举算法,并且任意节点还可能在选举算法过程中失效,这将导致其选举过程的不一致或无效,因此动态选举算法主要解决两个问题:

①如何处理多个源节点的情况;

②如何处理选举过程中节点失效的情况。

**定义 3.** 二元组  $SI(i) = \langle num, id \rangle$  表示一个节点的选举算法标识符,其中  $num$  是一个正整数,代表节点  $i$  发起的选举算法个数, $id$  是节点  $i$  的标识符。

**定义 4.**  $SI(i) > SI(j) \Leftrightarrow (num_i > num_j) \vee ((num_i = num_j) \wedge (id_i > id_j))$ 。

首先考虑并发的情况。为了避免网络中同时存在多个选举算法,也就是存在多个源节点的情况,每个选举算法的源节点  $S$  在发起选举算法时,都会先计算选举算法标识符(定义 3),即对  $SI(S) \rightarrow num$  加 1,同时把自己的节点标识符加入到  $SI(S) \rightarrow id$  中,然后将  $SI(S)$  加入到发出的选举消息中。

当聚集中某个节点收到一个选举消息并且已经设置了父节点时,将会比较新的选举消息标识符和当前的选举算法标识符。如果两者相等,则立即向发送端发出一个应答消息 ACK,并且在应答消息中加入自己的节点能力信息。如果大于当前自己参加的选举算法标识符,将立即停止当前的选举算法,加入到新的选举算法中,重新开始“生成树”的过程。而如果小于当前自己参加的选举算法标识符,则直接丢弃该消息。同时为了与选举消息保持一致,每个节点的应答消息中也同样包含选举算法标识符,以便自己的父节点辨别不同的应答消息。通过这种方法,可以保证最后网络上只存在唯一的一个有效选举算法。

并发情况下选举消息处理算法的伪代码如下,其中 STP 表示节点在生成树中的父节点(Spanning Tree Parent Peer),LSI(Local source index)保存节

点当前正在参与的选举算法的选举消息标识符。

### 算法 1. 并发选举算法.

输入: 移动社交网节点信息

输出: 盟主节点信息

```
On_Receive_Election(Node Sender, Node Receiver,
                    Message e){
```

```
IF (this.STP != NULL){
  IF (LSI > SI(e)){
    //Discard this election message
    exit;}
  ELSE IF(LSI == SI(e)){
    //Send ACK message to source immediately
    add(ACK, this, C);
    Send_ACK(sender, ACK, LSI);}
  ELSE{
    //Join the new election algorithm
    this.STP = source;
    LSI = SI(e);
    Create_Child_Tree(this.STP, e);}
  ELSE{
    //Never receive any election message before
    this.STP = source;
    LSI = SI(e);
    Create_Child_Tree(this.STP, e);}
```

## 5.2.2 移动 P2P 社交网络社交关系发现算法

当某个节点  $i$  发起社交关系发现请求时, 先在聚集内按照 P2P 的 Flood<sup>[18]</sup> 协议进行查找, 如果能找到则结束查询请求. 否则根据每个聚集内盟主节点逻辑标识, 将请求转发给盟主节点, 由盟主节点按照 Flood 的方式将请求转发给其它聚集的盟主节点, 转入其它聚集按照同样的方式继续查找, 为了保证算法的性能, 可以在查找过程中设置一个 Flood 次数门限值, 超过这个值则结束查找, 发现算法伪代码如下.

### 算法 2. 聚集内社交发现算法.

输入: 社交请求

输出: 聚集内社交推荐

```
Send_LookupReq(friend)
    //send a lookup request to find the friend
    r = nil; n' = nil;
    IF(find_caches(friend) is not nil)
        return;
    IF(find_rtable(friend) is not nil)
        r = get_rtable(friend);
    ELSE
        r = get_approximate(friend);
    n' = r.nextHop;
    Send_Req(friend, n');
    Recv_LookupReq(friend)
    //receive a lookup request from predecessor
```

```
    r = nil; n'' = nil;
    IF(find_caches(friend) is not nil)
        Send_LookupHit(pref, friend, context);
    return;
    IF(find_rtable(friend) is not nil)
        r = get_rtable(friend);
    ELSE
        r = get_approximate(friend);
        n'' = r.nextHop;
        Send_Forward(friend, n'');
算法 3. 聚集间社交发现算法.
    输入: 社交请求
    输出: 聚集间社交推荐
    Inter_lookupReq(friend)
        //run a lookup request on inter cluster
    r = nil; n' = nil; b' = nil;
    IF(find_caches(friend) is not nil)
        Send_LookupHit(src, friend, resource);
    return;
    IF(find_local(friend) is not nil)
        n' = get_local(friend);
        Send_Forward(friend, n');
    ELSE
        IF(find_rtable(friend) is not nil)
            r = get_rtable(friend);
            b' = r.bnid;
        ELSE
            b' = find_successor(friend);
            Send_Core_Forward(friend, b');
        Recv_LookupHit(pref, friend, context)
            //receive a lookup hit message
        update_ltable(pref, friend);
        IF(src is not equal to me)
            Send_LookupHit_Backlookup(pref, friend, context);
        IF(find_caches(friend) is nil)
            append_caches(context);
```

## 6 社交发现算法性能的理论分析

从上节算法描述可知, 无论是系统初始化聚集, 还是在新节点加入 P2P 移动社交网络时, 总是向相似度更高的某个聚集点靠近, 直到找到所属聚集或者创建新聚集为止, 所以节点聚集算法总是收敛的, 其查找开销决定了聚集和发现算法的性能. 因此, 下面将分析算法的平均查找开销.

如果将所有聚集的盟主节点组成一个无向图  $G_s$ , 图中每个节点代表一个聚集的盟主节点, 则新节点在聚集盟主节点间的平均查找开销, 等于  $G_s$  中任意两点间的平均距离  $D_{\text{average}}$  与  $G_s$  中相邻节点间平均传输开销的乘积. 聚集算法的平均查找开销与  $G_s$  中

任意两点间的平均距离  $D_{\text{average}}$  成正比. 不失一般性, 假设  $G_s$  共有  $S$  个盟主节点, 每个盟主节点的聚集邻居个数为  $m+1$  ( $m>2$ ), 也即  $G_s$  中每个节点的度数为  $m+1$ . 计算  $G_s$  中任意两点间的平均距离  $D_{\text{average}}$ , 实际就等同于从  $G_s$  任取一节点, 然后计算该节点到  $G_s$  中所有其它节点距离的平均值. 所以不妨从  $G_s$  中任取一个节点  $A$ , 以节点  $A$  作为树根来构造一颗树  $T_A$ .  $A$  将它的  $m+1$  个聚集邻居节点都加入到树  $T_A$  中作为  $A$  的孩子节点, 于是该  $m+1$  个孩子节点就组成了树  $T_A$  的第 1 层. 然后每个孩子节点又将自己除父节点之外的其它  $m$  个聚集邻居加入到树中作为自己的孩子节点, 于是这  $(m+1) \times m$  个节点就组成了树  $T_A$  的第 2 层, 依次类推, 直到  $G_s$  中所有的节点都加入到树  $T_A$  中.

由于从树根  $A$  到树  $T_A$  中第  $k$  层节点的距离是  $k$ , 若记  $N(k)$  代表树  $T_A$  中第  $k$  层的节点个数, 树  $T_A$  中最大层数为  $K_{\text{max}}$ , 则得到从  $A$  到  $T_A$  中任一节点的平均距离是

$$\sum_{k=1}^{K_{\text{max}}} \frac{k \times N(k)}{S} \quad (10)$$

式中  $S$  表示盟主节点个数.

不过由于聚集邻居的随机性, 所以在  $T_A$  中存在很多重复的情况, 即某个节点可能有多个父节点, 所以上述计算得到的从  $A$  到  $T_A$  中任一节点的平均距离并不等于  $D_{\text{average}}$ . 若记  $U(k)$  代表  $T_A$  中第  $k$  层中以前从未出现过的唯一节点个数, 则  $G_s$  中节点间真正的平均距离是

$$D_{\text{average}} = \sum_{k=1}^{K_{\text{max}}} \frac{k \times U(k)}{S} \quad (11)$$

为计算  $U(k)$ , 将树  $T_A$  中所有节点从根节点  $A$  开始编号, 编号从 1 开始, 顺序为从第 1 层、第 2 层到第  $K_{\text{max}}$  层, 每层则从左至右.

若记第  $k$  层最后一个节点的编号为  $L(k)$ , 则有

$$\begin{aligned} L(k) &= 1 + (m+1) + (m+1) \times m + \dots + \\ & (m+1) \times m^{k-1} = \\ & 1 + \frac{(m+1)(m^k - 1)}{m-1} = \\ & \frac{(m+1)m^k - 2}{m-1} \end{aligned} \quad (12)$$

设  $f(t)$  为编号  $t$  的节点加入树  $T_A$  后,  $T_A$  中不重复节点个数, 则有

$$f(t) - f(t-1) = \begin{cases} 1, & t \text{ 是新节点} \\ 0, & t \text{ 是旧节点} \end{cases} \quad (13)$$

设  $x(t)$  是编号  $t$  的节点为新节点的概率, 因为在节点  $t$  加入  $T_A$  之前, 由  $f(t)$  定义可知  $G_s$  中已经有  $f(t-1)$  个节点加入到  $T_A$  中, 而剩余节点的加入是随机的, 所以有

$$x(t) = \frac{S - f(t-1)}{S} \quad (14)$$

式中  $S$  表示盟主节点个数,  $f(t-1)$  表示编号为  $t-1$  的节点加入到树  $T_A$  后,  $T_A$  中不重复节点的个数.

记  $\bar{f}(t)$  为  $f(t)$  的期望, 对式(13)和(14)同时取期望得到

$$\begin{aligned} \bar{f}(t) - \bar{f}(t-1) &= \frac{S - \bar{f}(t-1)}{S} \Rightarrow \\ \bar{f}(t) &= 1 + \frac{S-1}{S} \bar{f}(t-1) = \\ & 1 + \frac{S-1}{S} + \left(\frac{S-1}{S}\right)^2 + \dots + \left(\frac{S-1}{S}\right)^t \bar{f}(0) \end{aligned}$$

而显然  $f(0)=0$ , 所以有

$$\bar{f}(t) = S \left[ 1 - \left(\frac{S-1}{S}\right)^t \right] \approx S(1 - e^{-\frac{t}{S}}), S \rightarrow +\infty \quad (15)$$

又由  $U(k)$ 、 $L(k)$  和  $f(t)$  定义可知

$$U(k) = \bar{f}(L(k)) - \bar{f}(L(k-1)) \quad (16)$$

于是由式(10)和(11)得到

$$\begin{aligned} D_{\text{average}} &= \sum_{k=1}^{K_{\text{max}}} k \times \frac{[\bar{f}(L(k)) - \bar{f}(L(k-1))]}{S} = \\ & \frac{k_{\text{max}} \bar{f}(L(k_{\text{max}})) - \bar{f}(L(k_{\text{max}-1})) - \dots - \bar{f}(L(0))}{S} \end{aligned} \quad (17)$$

又根据定义有

$$\lim_{k \rightarrow +\infty} \bar{f}(L(k)) \rightarrow S \quad (18)$$

式中  $S$  表示盟主节点个数,  $L(k)$  表示树  $T_A$  第  $k$  层最后一个节点的编号,  $\bar{f}(L(k))$  表示编号为  $L(k)$  的节点加入到树  $T_A$  后,  $T_A$  中不重复节点的个数的数学期望.

所以由式(17)和(18)得到

$$\begin{aligned} D_{\text{average}} &\leq \sum_{k=1}^{+\infty} k \times \frac{[\bar{f}(L(k)) - \bar{f}(L(k-1))]}{S} = \\ & \sum_{k=0}^{+\infty} \left( 1 - \frac{\bar{f}(L(k))}{S} \right) \end{aligned} \quad (19)$$

式中,  $D_{\text{average}}$  表示任意两个盟主节点之间的距离.

将式(15)和(12)代入式(19)得到

$$D_{\text{average}} \leq \sum_{k=0}^{+\infty} e^{-\frac{L(k)}{S}} = \sum_{k=0}^{+\infty} e^{-\frac{(m+1)m^k - 2}{(m-1)S}} =$$



$$\sum_{k=0}^{\log_m S} e^{-\frac{(m+1)m^k-2}{(m-1)S}} + \sum_{k=\log_m S+1}^{+\infty} e^{-\frac{(m+1)m^k-2}{(m-1)S}} <$$

$$\log_m S + 1 + \sum_{k=0}^{+\infty} e^{-\frac{(m+1)Sm^k-2}{(m-1)S}} \leq$$

$$\log_m S + 1 + \sum_{k=0}^{+\infty} e^{-m^k} \leq$$

$$\log_m S + 1 + \sum_{k=0}^{+\infty} e^{-mk} =$$

$$\log_m S + 1 + \frac{1}{1-e^{-m}} < \log_m S + 3.$$

这就证明了聚合算法和查找算法总是收敛的，并且其查找开销是  $O(\log_m S)$ ，其中  $S$  是 P2P 移动社交网络中盟主节点的个数， $m$  是每个盟主节点所在聚集内成员的个数。

## 7 算法的仿真实验及分析

### 7.1 实验环境设置

基于混合网络模型建立了实验原型系统，移动

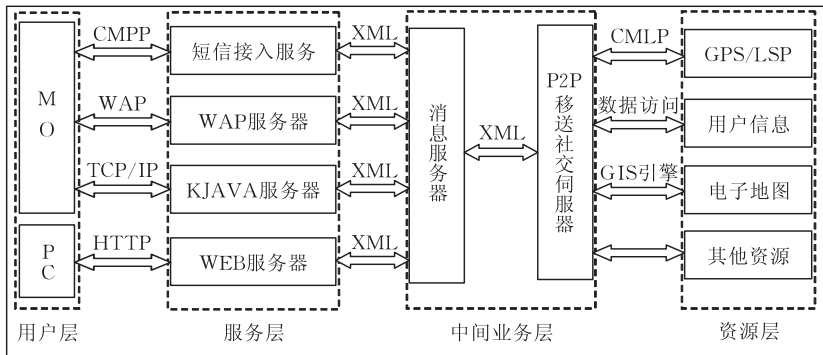


图 5 P2P 移动社交网络服务平台架构

客户端基于 KJAVA 平台，采用 Java 和 XML 语言实现了如第 3 节中图 3 所示的主要功能，使用消息机制在各个服务中传递数据。图 4 展示了服务器端的主要处理流程，图 5 展示了 P2P 移动社交网络服务平台架构，图 6 展示了智能手机客户端界面，图 7 显示了数据库中移动用户相关的感知信息。移动客户端所使用的移动设备有智能手机、平板电脑、笔记本等，接入方式有 3G 网络、WiFi。

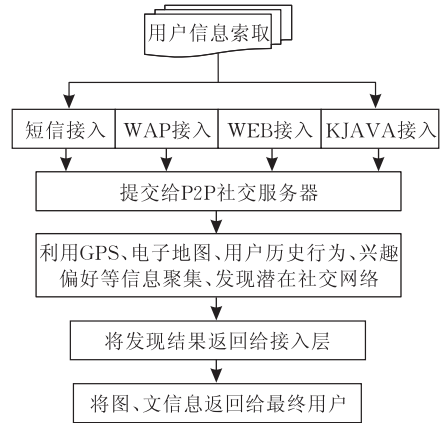


图 4 P2P 移动社交网络服务器端处理流程



图 6 客户端界面

用户ID	经度	纬度	速度	轨迹	温度	光照
2009312145	116.6449	39.54679	5	t Excel 工作表	30	122
2009312146	116.5449	39.13456	8	t Excel 工作表	26	126
2009312147	116.6549	39.23558	67	t Excel 工作表	29	556
2009312150	116.6455	39.23445	15	t Excel 工作表	26	245
2009312155	116.6478	39.45678	20	t Excel 工作表	28	845
2009312165	116.2255	40.14679	30	t Excel 工作表	25	236
2009312167	116.4689	39.13567	15	t Excel 工作表	32	4567
2009312177	116.8653	39.87652	6	t Excel 工作表	26	345
2009312179	116.4476	39.54564	70	t Excel 工作表	31	1676
2009312180	116.5666	39.22226	75	t Excel 工作表	25	756
2009312186	116.5278	39.35678	20	t Excel 工作表	31	556
2009312199	116.8745	39.46789	30	t Excel 工作表	24	345

图 7 数据库用户信息

### 7.2 聚合算法的仿真实验

本文所提出的基于确定性退火技术的分布式移动 P2P 社交网络聚合 DCMPDA 算法是一种分布式

并行算法,要在移动设备上运行.由于移动设备能量和运算能力有限,所以该算法的开销、收敛速度和通信量对系统的整体性能至关重要.为此,我们进行了大规模的仿真实验来检验其性能.模拟实验在网络拓扑生成器 BRITE(Boston University Representative Internet Topology Generator)的 Java 开源版本以及离散事件驱动模拟程序包 Simjava 基础上,通过 Java 语言实现,忽略了实际网络传输过程中的延迟、拥塞、丢包等细节问题.整个模拟实验包括节点物理拓扑生成、聚集算法、离散事件驱动模拟程序、情景感知信息模拟、消息传递以及发现算法等 6 个主要部分,中间节点的物理拓扑利用 BRITE 生成. BRITE 是波士顿大学开发的一个通用拓扑产生器,它实现了 Waxman、BA、BA-2、GLP 多种拓扑生成算法,并可以在 AS(自治域)级和路由器级产生拓扑图,用自顶向下和自底向上两种方法来产生 Transit-Stub 层次模型.为尽量接近现实的移动混合网络拓扑结构,本文模拟实验中的节点物理拓扑图按照 GLP 生成算法在 BRITE 路由器级生成,并且随机注入地理位置、温度、光照、运动轨迹等感知信息.

在仿真系统中运行传统的  $k$ -means 聚合算法和本文所提出的社交网络聚合 DCMPDA 算法,所得到的聚合拓扑图分别如图 8、图 9 所示,可见 DCMPDA 算法所得到的聚合结果规模适度,比较接近现实的社交关系.算法运行所花费的时间开销如图 10 所示,可以看出随着节点个数的增加,DCMPDA 聚合算法所花费的时间开销增长平缓,具有良好的扩展性,而  $k$ -means 聚合算法则呈现指数增长的趋势. DCMPDA 算法消息总量随节点数的变化如图 11 所示,由于 DCMPDA 算法是一种分布式的并行算法,其交互的消息总量随着节点数的增加呈现快速增长的趋势,如何减少其交互的消息总量将是下一步需要研究的问题.

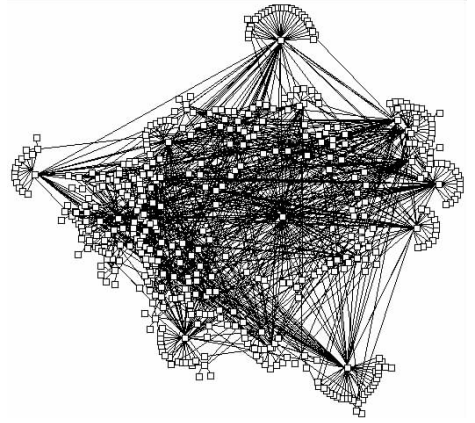


图 9 DCMPDA 算法聚合结果

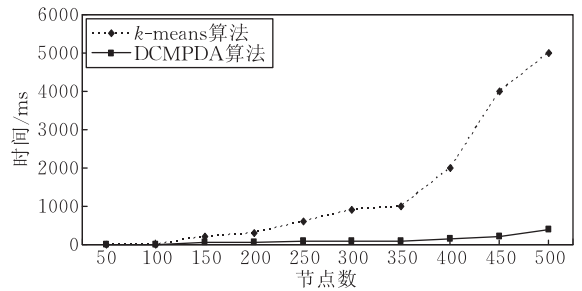


图 10 DCMPDA 算法和  $k$ -means 算法性能比较

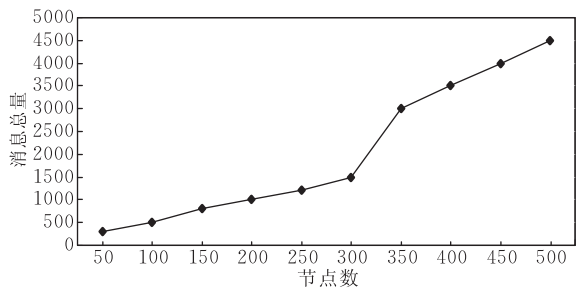


图 11 DCMPDA 算法消息总量随节点数的变化

### 7.3 社交发现响应速度

仍然采用 7.2 节的实验环境,对社交发现进行仿真测试.一种是在预先聚合的基础上按照本文所提出的算法来发现潜在的关系,另一种是直接在所有用户里遍历查找.推荐朋友总数门限值设为 20.结果如图 12 所示,可以看出随着用户数的增加,本

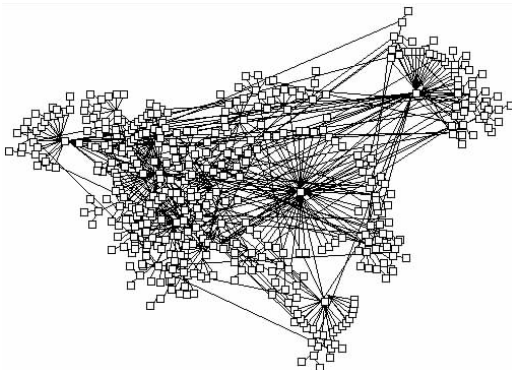


图 8  $k$ -means 算法聚合结果

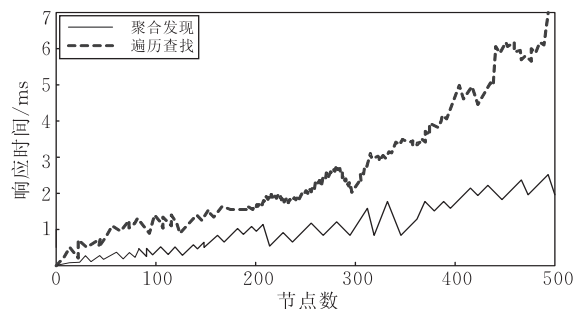


图 12 本文聚合发现算法与遍历查找算法响应速度比较

文聚合发现算法的响应时间增长缓慢,而随机遍历查找则呈现出指数增长的态势.究其原因,是因为本文的算法是一种分布式并行算法,受问题规模的影响不大.

#### 7.4 实际移动用户社交网络聚合及社交发现测试

我们选择了 30 个实际的移动用户作为测试对象,注册下载了 P2P 移动社交客户端软件,同时运行该软件,进行了实际的情景感知聚合及社交发现测试,图 13 展示了通过情景感知自动聚合的结果,虚线连接表示潜在的社交网络,实线表示的是根据用户提交的请求,自动发现的潜在朋友.如图所示,可见满足其要求的朋友多数都属于潜在社交聚集.这表明,自动感知聚合的潜在社交网络与用户可能感兴趣的社交关系具有极大的相关性,情景感知的自动聚合将加快自动社交发现的速度,对于社交发现具有极大的帮助,避免了社交的盲目性和随机性.

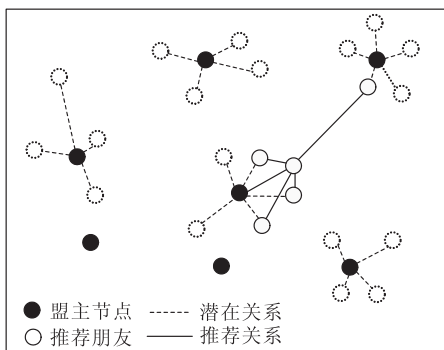


图 13 移动用户聚合及社交发现结果

#### 7.5 情景感知聚合发现用户满意度评价、查准率和查全率比较

为了测试社交发现的实际效果,随机选取了不同的人群,进行了多次测试.每次测试选择 30 个测试对象,对结果进行了满意度的问卷调查和人工访谈,并对多次测试的结果进行统计分析.将本文情景感知的社交发现和根据用户注册信息遍历查找进行对比,其查全率、查准率和用户满意度如图 14 所示,可以看出,情景感知的聚合发现具有很高的查全率、查准率和用户满意度.

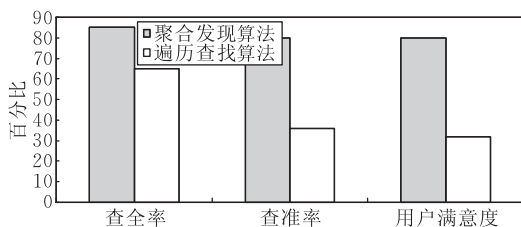


图 14 查全率、查准率和用户满意度比较

## 8 结论和未来的工作

如何通过情景感知,获得用户的行为特征,从而自动发现潜在的社会关系,是移动社交网络重要的研究内容之一.本文提出了情景感知的移动 P2P 社交网络的系统架构、聚合模型及发现算法,将用户的位置信息、环境特征、运动轨迹等引入到聚合算法中,智能地聚合形成潜在的 P2P 社交网络,根据用户的需求自主发现潜在的关系,避免了社交活动的盲目性和随意性.并进行了算法的理论分析及实验验证,结果表明本文所提出的方案及算法具有较高的响应速度、准确率及用户满意度,同时该系统具有一定的负载均衡和自适应能力.

目前本文的实验系统是一个原型系统,规模较小,下一步需要进行更大规模的实际应用测试.目前我们的工作采用的定位技术基于 GPS,但实际的定位技术有很多,包括移动基站定位等;覆盖的 P2P 社交网络用户之间通信采用 Flood 协议,这个协议是非结构化的,其产生的消息总量过多,是否适合大规模应用,还需要进一步的测试;采用新的通信协议实现(如基于场景的无线点对点通信:如蓝牙协议),会提高信道的使用效率.此外,隐私控制和信用管理是社交网络构造和发现的重要影响因素,这也是将来需要进一步研究的内容.

### 参 考 文 献

- [1] Cranshaw J, Toch E, Hong J et al. Bridging the gap between physical location and online social networks//Proceedings of the 12th ACM International Conference on Ubiquitous Computing. New York, USA, 2010: 119-128
- [2] Kimmerle J, Cress U. Group awareness and self presentation in computer-supported information exchange. International Journal of Computer-Supported collaborative Learning, 2008, 3(1): 85-97
- [3] Li Gang, Sun Hong-Mei, Li Zhi et al. Resource constrained Web services. Chinese Journal of Computers, 2010, 33(2): 193-207(in Chinese)  
(李刚, 孙红梅, 李智等. 资源受限 Web 服务, 计算机学报, 2010, 33(2): 193-207)
- [4] Wu Jian-Ping, Wu Qian, Xu Ke. Research and exploration of next-generation Internet architecture. Chinese Journal of Computers, 2008, 31(9): 1536-1548(in Chinese)  
(吴建平, 吴茜, 徐格. 下一代互联网体系结构基础研究与探索. 计算机学报, 2008, 31(9): 1536-1548)
- [5] Li Zhi, Zhang Hong, Liu Feng-Yu. Friend cluster based delay tolerant routing protocol in social networks. Computer Science, 2012, 39(2): 26-28(in Chinese)  
(李陟, 张宏, 刘凤玉. 一种基于好友簇的社交网络中的实验容忍路由协议. 计算机科学, 2012, 39(2): 26-28)
- [6] Tam Van Nguyen, Wontaek Lim, Huy Anh Nguyen et al. Context awareness framework based on contextual graph//

- Proceedings of the 5th International Conference on Wireless and Optical Communications Networks. Surabaya, Indonesia, 2008: 1-5
- [7] Buder J, Bodemer D. Supporting controversial CSCL discussions with augmented group awareness tools. *International Journal of Computer-Supported Collaborative Learning*, 2008, 3(2): 123-139
- [8] Qiao Xiu-Quan, Yang Chun, Li Xiao-Feng, Chen Jun-Liang. A trust calculating algorithm based on social networking service users' context. *Chinese Journal of Computers*, 34(12): 2403-2413 (in Chinese)  
(乔秀全, 杨春, 李晓峰, 陈俊亮. 社交网络服务中一种基于用户上下文的信任度计算方法. *计算机学报*, 2011, 34(12): 2403-2413)
- [9] GU Jun-Zhong. Context aware computing. *Journal of East China Normal University (Natural Science)*, 2009, 5: 1-20 (in Chinese)  
(顾君忠. 情景感知计算. *华东师范大学学报(自然科学版)*, 2009, 5: 1-20)
- [10] Schmidt A, Gellersen H-W. Visitor awareness in the Web// *Proceedings of the 10th International Conference on World Wide Web*. Hong Kong, China, 2001: 745-753
- [11] Acquisti A, Gross R. Imagined communities: Awareness, information sharing, and privacy on the Facebook. *Privacy Enhancing Technologies*, 2006, 3(1): 36-58
- [12] Matsuo, Morib J, Hamasaki M. POLYPHONET: An advanced social network extraction system from the Web// *Proceedings of the 15th International Conference on World Wide Web*. Edinburgh, UK, 2006: 397-406
- [13] Mika P. Flink: Semantic Web technology for the extraction and analysis of social networks. *Journal of Web Semantics*, 2005, 3(2): 211-223
- [14] Ren Y, Zhang S D, Zhang H K. Theories and algorithms of coverage control for wireless sensor networks. *Journal of Software*, 2006, 17(3): 422-433 (in Chinese)  
(任彦, 张思东, 张宏科. 无线传感器网络中覆盖控制理论与算法. *软件学报*, 2006, 17(3): 422-433)
- [15] Yang Guang-Wen, Li Xiao-Ming, Wang Yi-He et al. Deterministic annealing. *Chinese Journal of Computers*, 1998, 21(8): 765-768 (in Chinese)  
(杨广文, 李晓明, 王义和等. 确定性退火技术. *计算机学报*, 1998, 21(8): 765-768)
- [16] Lin Chuang, Li Yin, Wan Jian-Xiong. Optimization approaches for QoS in computer networks: A survey. *Chinese Journal of Computers*, 2011, 34(1): 1-14 (in Chinese)  
(林闯, 李寅, 万剑雄. 计算机网络服务质量优化方法研究综述. *计算机学报*, 2011, 34(1): 1-14)
- [17] Wang Peng, Luo Jun-Zhou, Li Wei et al. A cooperative game model for traffic engineering and overlay routing based on trustworthy and controllable network. *Chinese Journal of Computers*, 2010, 33(9): 1664-1674 (in Chinese)  
(王鹏, 罗军舟, 李伟等. 基于可信可控网络的流量工程与覆盖网路由的合作博弈模型. *计算机学报*, 2010, 33(9): 1664-1674)
- [18] Chopra D, Schulzrinne H, Marocco E, Ivov E. Peer-to-peer overlays for real-time communication: Security issues and solutions. *IEEE Journal on Communications Surveys & Tutorials*, 2009, 11(3): 4-12



**CAO Huai-Hu**, born in 1977, Ph. D., associate professor. His research interests include computer network architecture, mobile Internet and network computing.

**ZHU Jian-Ming**, professor, Ph. D. supervisor. His research interests include wireless network, network security and network survivability.

**PAN Yun**, born in 1974, Ph. D., associate professor. Her research interests include computer network architecture and mobile Internet.

**LI Hai-Feng**, born in 1979, Ph. D., lecturer. His research interests include data mining and data management.

## Background

With popularity of smart mobile phones, tablet, notebook computers and other portable handheld mobile devices and development of mobile communication technology, people hope that the social network can find actively potential friends according to the user's environment, personal preferences, behavioral characteristics. Thus, it is important to study how to construct the potential of social network structure according to the scene and behavior characteristics and find some highly relevant friends based on user's interests and preferences.

In the mobile environment, through the context-aware, the interactive users can understand each other's state within the social network and self-discover new or old relationships. This paper presents a context-aware mobile P2P social network framework, aggregation model and the discovery algorithm. The user's location information, environmental characteristics etc. are introduced to the aggregate algorithm, which intelligently aggregate automatically to potential P2P social network and discover the suited relationship according

to user's demand, the social blindness and randomness are avoided. The theoretical analysis and experimental results show that the proposed approach and the algorithm have a higher response speed, accuracy and user satisfaction. We will further improvement the efficiency and practicality of this system in the future.

This group's research interests are currently focused on mobile social network, mobile Internet, cognitive network, wireless network, and network security etc. The long term goal is build a context-aware intelligent and self-adaptive social network, and discover potential commercial value and provide users with high-quality social services. This work was supported by the National Natural Science Foundation of China (Nos. 61103198, 61100112, 60970143), the Key Project of Chinese Ministry of Education (No. 109016), the Ministry of Education in China Project of Humanities and Social Sciences (No. 11YJCZH006) and the Discipline Construction Foundation of Central University of Finance and Economics (No. CUFEIE201107).