

移动对象时空轨迹及社交关系一体化数据模型

张恒才¹ 陆 锋¹ 陈 洁¹

1 中国科学院地理科学与资源研究所资源与环境信息系统国家重点实验室,北京,100101

摘 要:泛地理信息时代来临及 SoLoMo 互联网模式盛行,如何高效管理移动对象位置数据、社交关系数据及地理空间数据成为 GIS 与移动对象数据库研究领域一项全新的技术挑战。首先基于图模型提出了移动对象时空轨迹及社交关系的一体化数据模型 GSM,实现了移动对象所在地理空间、移动对象位置与轨迹、移动对象之间社交关系的时空一体化表达,并定义了该模型的四种基础操作算子;然后基于图数据库系统 Neo4J 实现了 GSM 模型;最后利用真实的 Brightkite、Gowalla 位置社交网络数据验证了 GSM 模型的有效性。所提出的 GSM 模型能够有效避免关系数据库中表连接造成的移动对象管理与查询效率低下的问题,具有良好的执行效率及稳定性,可为基于位置的社交网络应用提供核心技术支持。

关键词:移动对象;数据模型;轨迹;社交网络

中图法分类号:P208

文献标志码:A

近年来,融合 SNS(social network system)与 LBS(location based service)的 SoLoMo 地理信息服务模式逐渐引起了人们的关注,也是新地理信息时代^[1]面临的挑战。SoLoMo 即社交(social)+本地化(local)+移动(mobile),它是 PC、移动设备和真实社会关系的外延,代表着未来互联网发展的趋势。具有 SoLoMo 模式特征的位置服务应用不断兴起,如微信、Twitter、Facebook、LiveJournal、Brightkite、Foursquare、Gowalla 等成为人们沟通交流、分享信息的主要渠道。SoLoMo 强调 Social、Local、Mobile 的无缝整合,其中,Social 为线上虚拟网络空间中人们之间广泛存在的社交关系,如关注、朋友、同事等;Local 为线下真实地理空间中用户位置或活动信息,如签到、消费、旅游等;Mobile 指人们在真实空间的移动。随着“人人都是传感器”^[2]愿景逐渐走向现实,互联网用户数目增长及 O2O(online to offline)商业模式逐渐成熟,SoLoMo 会产生海量的移动对象线上社交关系数据与线下移动位置数据,如何有效存储、管理与查询这些动态变化的数据已成为地理信息系统、移动对象数据库等领域的研究热点与难点问题^[3-4]。

移动对象数据模型是移动对象数据库(mov-

ing objects database, MOD)的基础,是有效管理社交网络系统中用户线下位置活动信息的关键^[5]。目前,移动对象位置数据模型主要的研究成果包括自由空间下的 MOST 模型^[6]、移动对象离散数据模型^[7]、离散时空轨迹 DSTTMOD 模型^[8]、抽象数据类型^[9]等;路网空间下的基于路段的路网空间移动对象模型^[10-13]、基于道路的路网空间移动对象模型^[14-15]、基于分区的路网空间移动对象模型^[16-17]、时空一体化路网空间移动对象模型^[18]等。这些模型并不考虑移动对象之间复杂和动态变化的语义关系,因此也基本不支持 Geo-Social 的联合查询。对于社交网络的研究虽然很多,但大都集中在社交网络分析环节,如研究人们面对面交流模式与社交网络的关系^[19-22]、推断社交网络联系类型^[23-25]、预测新的社交网络关系出现^[26-28]、网络舆情监测^[29]及利用社交网络定量分析人类行为模式^[30]等。或者研究社交网络关系与移动轨迹之间的交互影响,如虚拟世界社交语义关系与真实世界移动轨迹相互影响^[31-34],虚拟社交关系会影响移动对象真实的轨迹行为^[35-36],移动对象的运行轨迹也会影响虚拟社交关系的形成^[37-38]等。在地理社交数据建模研究方面,文献^[39]提出了一种分层用户轨迹数据建模

收稿日期:2014-02-19

项目来源:国家 863 计划资助项目(2012AA12A211,2013AA120305);国家自然科学基金资助项目(41271408, 41101149)。

第一作者:张恒才,博士,博士后。研究方向为互联网空间信息搜索轨迹数据管理与数据挖掘。E-mail: zhanghc@reis.ac.cn

通讯作者:陆锋,博士,研究员,博士生导师。E-mail: luf@reis.ac.cn

方法,支持用户的相似性计算及经典旅行路线推荐。但该模型没有考虑地理空间建模和移动对象间的社交网络联系。

本文尝试在移动对象建模过程中将移动对象间的社交关系作为一种重要的非空间语义信息,提出了一种移动对象时空及社交关系表达模型(geospatial-social-moving trajectory, GSM),以实现地理空间、移动对象轨迹及社交关系的一体化表达,为基于位置的社交网络应用提供核心技术支持。

1 GSM 模型定义

1.1 地理空间剖分

在移动对象数据模型中,如何控制地理空间表达粒度是一个重要问题。如果粒度过大,虽然有助于降低移动对象位置更新代价,但是在MOD的实时查询或位置服务应用中难以准确描述移动对象的实时位置。鉴于此,本文引入Voronoi图对移动对象活动的地理空间进行剖分。Voronoi图定义如下:

$$V(p_i) = \{ q \in \forall q \in p, d(q, p_i) \leq d(q, p_j), i \neq j, p_i, p_j \in P, j = 1, 2, \dots, n \}$$

其中, $V(p_i)$ 为空间点 p_i 的 Voronoi 区域; P 为空间点集,如 POI; n 为空间点集数量; q 为空间中的任意一点; d 为空间距离函数。Voronoi 图是一种基本的数据结构,可以将空间划分成不相交的多边形,如图 1 所示。Voronoi 的优势在于使得划分后的多边形中任一位置的最近邻 POI 都在该多边形内,可以有效控制地理空间建模粒度。

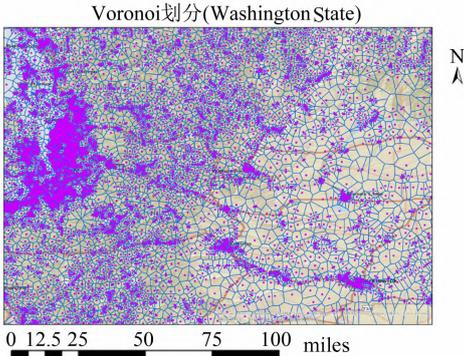


图 1 基于 POI 空间分布的 Voronoi 空间剖分
Fig. 1 Voronoi-based Geographical Space Partition

1.2 GSM 数据模型

GSM 模型结构如图 2 所示,该模型由三个子图结构组成:地理空间图(geospatial graph,GG)、移动对象之间的社交联系图(social graph,SG)及

移动对象轨迹图(moving trajectory graph, MG)。

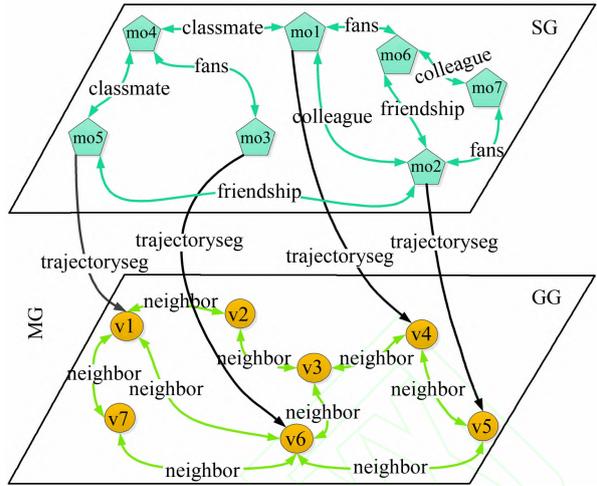


图 2 GSM 模型系统架构
Fig. 2 GSM Model Architecture

定义 1 地理空间图(GG)定义为:

$$GG = (G_{Vertices}, G_{Edges})$$

其中, $G_{Vertices}$ 为根据 POI 空间分布生成的 Voronoi 多边形集合; G_{Edges} 为节点边集合。

定义 2 社交网络图(SG)定义为:

$$SG = (S_{Vertices}, S_{Edges})$$

其中, $S_{Vertices}$ 为移动对象集合; S_{Edges} 为移动对象间社交网络关系集合。

定义 3 移动对象轨迹图(MG)定义为:

$$MG = (M_{Vertices}, M_{Edges})$$

其中, $M_{Vertices}$ 为移动对象 mo 集合与地理空间 Voronoi 剖分多边形集合; M_{Edges} 为移动对象运动轨迹 mo2geoedge 集合。

定义 4 移动对象定义为:

$$mo = (mo_{id}, mo_{name}, param, mloc)$$

其中, mo_{id} 为移动对象编号; mo_{name} 为移动对象名称; $param$ 为该移动对象的其他属性参数集合; $mloc$ 为该移动对象当前位置信息。

定义 5 社交关系定义为:

$$socialedge = (socialid, mo_s, mo_e, s_{time}, e_{time}, socialtype)$$

其中, $socialid$ 为社交关系编号; mo_s 和 mo_e 为具有社交关系的两个移动对象; s_{time} 为该条社交关系的建立时间; e_{time} 为该社交关系的结束时间; $socialtype$ 为社交关系类型。社交网络关系可以分为虚拟世界社交网络关系(如关注关系、粉丝关系等)与真实物理世界社交网络关系(如朋友关系、同学及同事关系等)。

定义 6 签到位置 ckpos 定义为:

$$ckpos = (voronid, moid, pos_x, pos_y, time)$$

其中, voronid 为 Voronoi 多边形编号; pos_x 和 pos_y 为该移动对象的签到位置; time 为签到时间。

定义 7 移动对象轨迹片段. 移动对象轨迹片段 mo2roadedge 定义为:

$$\text{mo2geoedge} = ((\text{mo}, \text{voronoi}), (s_{\text{time}}, e_{\text{time}}, (\text{ckpos}_1, \text{ckpos}_2, \dots, \text{ckpos}_m)))$$

其中, mo 为移动对象; voronoi 为该移动对象所在的 Voronoi 多边形; s_{time} 为该轨迹片段的开始时间, 即进入该 Voronoi 的时间; e_{time} 为该轨迹片段的结束时间, 即离开该 Voronoi 的时间; $\text{ckpos}_1, \text{ckpos}_2, \dots, \text{ckpos}_m$ 为签到位置序列。

1.3 GSM 模型操作

为了在 GSM 模型中实现移动对象、地理空间及位置轨迹数据的联合查询, 需要定义一些基本的模型操作符, 包括截取操作、扩展操作及投影操作。

定义 8 Select(G, C) 截取操作定义为根据指定查询条件 C 对 GSM 模型子图 G 进行截取操作, 其中 $G \subseteq (GG, SG, MG), C \subseteq (\text{attr}_1, \text{attr}_2, \dots, \text{attr}_m), \text{attr}_i$ 为图节点属性, m 为节点属性个数。具体包括: ① Select(GG, C): 对地理空间子图 GG 根据查询条件 C 进行截取操作, 返回满足条件的地理空间子图节点集 NS_{gg} , 且 $NS_{\text{gg}} \subseteq G_{\text{Vertices}}$; ② Select(SG, C): 对社交网络子图 SG 根据查询条件 C 进行截取操作, 返回满足条件的社交网络子图节点集 NS_{sg} , 且 $NS_{\text{sg}} \subseteq S_{\text{Vertices}}$; ③ Select(MG, C): 对位置轨迹子图 MG 根据查询条件 C 进行截取操作, 返回满足条件的位置轨迹子图节点集 NS_{mg} , 且 $NS_{\text{mg}} \subseteq M_{\text{Vertices}}$ 。

例如, 查询子图为北京地理空间, 查询条件为 Name="中国科学院地理科学与资源研究所", 查询语句如下:

$$NS_{\text{gg}} = \text{Select}(GG_{\text{beijing}}, \text{Name} = \text{'中国科学院地理科学与资源研究所'})$$

该截取操作将会返回名称为“中国科学院地理科学与资源研究所”所在的地理空间剖分多边形集合 NS_{gg} 。

定义 9 Expand(NS, d) 扩展操作定义为查询距离子图节点集 NS 中长度为 d 的所有节点集合, 其中 d 为距离阈值, $NS \subseteq (NS_{\text{gg}}, NS_{\text{sg}}, NS_{\text{mg}})$ 。且 $NS_{\text{gg}} \subseteq G_{\text{Vertices}}, NS_{\text{sg}} \subseteq S_{\text{Vertices}}, NS_{\text{mg}} \subseteq M_{\text{Vertices}}$ 。具体包括: ① Expand(NS_{gg}, d): 对地理空间子图节点集 NS_{gg} 根据查询距离阈值 d 进行扩展操作, 返回满足条件的 G_{Vertices} 集合; ② Expand(NS_{sg}, d): 对社交网络子图节点集 NS_{sg} 根据查询距离阈值 d 进行扩展操作, 返回满足条件的 S_{Vertices} 集合;

③ Expand(NS_{mg}, d): 对位置轨迹子图节点集 NS_{mg} 根据查询距离阈值 d 进行扩展操作, 返回满足条件的 M_{Vertices} 集合。

例如, 查询节点集 NS 为社交网络图中 Name = mo₁ 的移动对象, 查询距离阈值 d 分别为 1 或者 2, 查询语句如下:

$$F = \text{Expand}(\text{Select}(SG, \text{Name} = \text{'mo}_1'), 1) \\ FF = \text{Expand}(\text{Select}(SG, \text{Name} = \text{'mo}_1'), 2)$$

当阈值 $d=1$ 时, 扩展操作返回移动对象 mo₁ 的所有朋友集合 F ; 当阈值 $d=2$ 时, 扩展操作返回移动对象 mo₁ 的朋友的朋友 FF 。以图 2 为例, $F = (\text{mo}_2, \text{mo}_4, \text{mo}_6), FF = (\text{mo}_5, \text{mo}_3, \text{mo}_7)$ 。

定义 10 Cross_{G1}(NS_{G2}, TP) 投影操作定义为将子图 $G2$ 节点集 NS_{G2} 投影到其他 $G1$ 子图空间, TP 为时间参数。该操作与 Select 截取操作及 Expand 扩展操作不同, 是在两个子图空间之间进行映射。具体包括: ① Cross_{GG}(NS_{sg}, TP): 对社交网络子图节点集 NS_{sg} 根据时间参数 TP 投影到地理空间子图 GG , 返回满足条件的地理空间子图节点集; ② Cross_{MG}(NS_{sg}, TP): 对社交网络子图节点集 NS_{sg} 根据时间参数 TP 投影到位置轨迹子图 MG , 返回满足条件的位置轨迹子图节点集; ③ Cross_{SG}(NS_{gg}, TP): 对地理空间子图节点集 NS_{gg} 根据时间参数 TP 投影到社交网络子图 SG , 返回满足条件的社交网络子图节点集; ④ Cross_{MG}(NS_{gg}, TP): 对地理空间子图节点集 NS_{gg} 根据时间参数 TP 投影到位置轨迹子图 MG , 返回满足条件的位置轨迹子图节点集; ⑤ Cross_{SG}(NS_{mg}, TP): 对位置轨迹子图节点集 NS_{mg} 根据时间参数 TP 投影到社交网络子图 SG , 返回满足条件的社交网络子图节点集; ⑥ Cross_{GG}(NS_{mg}, TP): 对位置轨迹子图节点集 NS_{mg} 根据时间参数 TP 投影到地理空间子图 GG , 返回满足条件的地理空间子图节点集。

例如, 查找社交网络子图 SG 中名称为 mo₂ 的移动对象在 2013 年 10 月 1 日 22:00 至 22:30 所经过的地理空间及位置轨迹, 查询语句如下:

$$R1 = \text{Cross}_{\text{GG}}(\text{Select}(SG, \text{Name} = \text{'mo}_2'), \\ S_{\text{Time}} = \text{'2013-10-01 22:00:00'}, \\ E_{\text{Time}} = \text{'2013-10-01 22:30:00'}) \\ R2 = \text{Cross}_{\text{MG}}(\text{Select}(SG, \text{Name} = \text{'mo}_2'), \\ S_{\text{Time}} = \text{'2013-10-01 22:00:00'}, \\ E_{\text{Time}} = \text{'2013-10-01 22:30:00'})$$

当投影空间为地理空间子图 GG 时, 返回移动对象 mo₂ 在 30 min 内所经过的地理区域 $R1$; 当投影空间为 MG 时, 返回移动对象 mo₂ 在 30

min 内的位置轨迹 R2。

定义 11 MultiCross_{G1}(NS_{G2}, NS_{G3}, TP) 多维投影操作定义为将子图 G2 的节点集与子图 G3 的节点集投影到子图 G1, TP 为时间参数。具体包括: ① MultiCross_{MG}(NS_{SG}, NS_{GG}, TP): 将社交网络子图节点集 NS_{SG} 与地理空间子图节点集 NS_{GG} 根据时间参数 TP 投影到位置轨迹子图 MG, 返回满足条件的位置轨迹子图节点集; ② MultiCross_{SSG}(NS_{MG}, NS_{GG}, TP): 将位置轨迹子图节点集 NS_{MG} 与地理空间子图节点集 NS_{GG} 根据时间参数 TP 投影到社交网络子图 SG, 返回满足条件的社交网络子图节点集; ③ MultiCross_{GG}(NS_{SG}, NS_{MG}, TP): 将社交网络子图节点集 NS_{SG} 与位置轨迹子图节点集 NS_{MG} 根据时间参数 TP 投影到地理空间子图 GG, 返回满足条件的地理空间子图节点集。

例如, 查询社交网络子图 SG 中名称为 mo₆ 的移动对象在 2013 年 10 月 1 日 12:00 在地理空间区域 V₆ 内所运行的位置轨迹数据, 查询语句为:

```
MR1 = MultiCross( Select( SG, Name = 'mo6'), Select( GG, Name = 'V6'), Time = '2013-10-01 12:01:10')
```

该多维投影操作返回 mo₆ 运行在地理区域 V₆ 内的位置轨迹数据。

2 模型实现与验证

为验证 GSM 模型和模型基本操作符的有效性, 利用 NoSQL 阵营中的图数据库系统 Neo4J (V1.9.5) 与 Java 语言实现了所提出的 GSM 模

型, 并选取了不同的 LBSN 实验数据集, 包括不同的移动对象数量、社交关系数量及签到位置数量, 同时利用成熟的 RDBMS 平台 PostgreSQL (V8.4.18) 实现了移动对象管理, 采用二维关系表存储移动对象轨迹, 进行实验对比。

2.1 实验环境

实验环境为 Dell Power Edge R720, CPU 为 4 核 Intel(R) Xeon(R) CPU E5520 2.27 GHz, 内存 4 G, 运行环境为 64 位的 CentOS Linux 6.4 操作系统。

Neo4J 以图数据结构作为数据库底层存储模型, 基本的存储单元包括节点、关系及属性, 具有完全的事务特征与高效的图结构遍历算法, 对 GSM 模型具有原生的支持功能。

实验所采用的位置签到数据为开放的基于位置的社交网络系统 Brightkite 与 Gowalla 美国站的数据。其中 Gowalla 数据包含 196 591 个移动对象和 6 442 890 个签到位置, 时间范围为 2009 年 2 月 4 日至 2010 年 10 月 23 日; Brightkite 数据包含 58 228 个移动对象和 4 491 143 个签到位置, 时间范围为 2008 年 3 月 21 日至 2010 年 10 月 18 日; 地理空间及 POI 数据来源于美国人口调查局所提供的开放数据。实现区域选择位置签到数据较丰富的华盛顿州 (Washington State, WC)、加利福尼亚州 (California State, CA)、纽约州 (New York State, NY)、德克萨斯州 (Texas State, TX)。图 3 为实验所用签到位置数据的空间分布图, 红色点为 Brightkite 签到; 蓝色点为 Gowalla 签到。将实现数据集分成 8 组, 如表 1 所示。实验中所采用的 GSM 模型操作如表 2 所示。

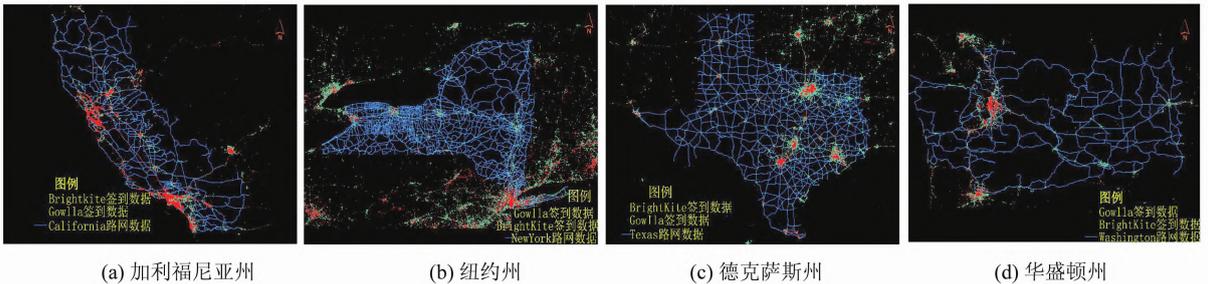


图 3 签到位置空间分布图

Fig. 3 Spatial Distribution of Check-in Data

表 1 实验数据集分组

Tab. 1 Experimental Data

| 编号 | V 数量 | 移动对象数量 | 社交关系数量 | 签到数量 | 编号 | V 数量 | 移动对象数量 | 社交关系数量 | 签到数量 |
|-------|--------|--------|---------|---------|-------|--------|--------|---------|---------|
| WC-GO | 18 655 | 5 058 | 163 622 | 199 469 | NY-GO | 21 657 | 12 388 | 391 144 | 364 940 |
| WC-BK | 18 655 | 797 | 59 194 | 93 816 | NY-BK | 21 657 | 1 789 | 108 678 | 157 020 |
| CA-GO | 57 251 | 16 684 | 513 406 | 739 687 | TX-GO | 40 178 | 17 996 | 524 352 | 984 493 |
| CA-BK | 57 251 | 2 450 | 142 136 | 293 542 | TX-BK | 40 178 | 1 328 | 99 528 | 111 247 |

表 2 实验中采用的 GSM 模型操作

Tab. 2 Tested Operations for the GSM Model

| GSM 模型操作 | 实验案例 |
|------------|--|
| Select | Select(SG, Name='mo _i ') |
| Expand | Expand(Select(SG, Name='mo _i '), 2) |
| Cross | Cross _{MG} (Select(SG, Name='mo _i '), S _{Time} ='2013-10-01 22:00:00', E _{Time} ='2013-10-01 22:30:00') |
| MultiCross | MultiCross _{MG} (Select(SG, Name='mo _i '), Select(GG, Name='V _j '), S _{Time} ='2013-10-01 12:00:00', E _{Time} ='2013-10-01 12:30:00') |

2.2 实验分析

表 3 为不同的实验数据集对 GSM 模型操作执行效率性能的影响。其中运行时间是多次实验的平均值,且不包含索引构建、系统初始化及预处理时间。从表 3 可以看出,总体上,GSM 模型操作符运行效率优于 PostgreSQL 实现方式。

表 3 GSM 模型操作运行时间/ms

Tab. 3 Operational Time for the Operations in the GSM Model/ms

| | Select | | Expand | | Cross | | MultiCross | |
|-------|--------|------|--------|--------|-------|------|------------|------|
| | GSM | PSQL | GSM | PSQL | GSM | PSQL | GSM | PSQL |
| CA-BK | 128 | 269 | 287 | 4 120 | 155 | 454 | 201 | 450 |
| CA-GO | 141 | 352 | 671 | 62 471 | 168 | 535 | 532 | 538 |
| NY-BK | 137 | 266 | 271 | 3 195 | 140 | 345 | 483 | 335 |
| NY-GO | 128 | 324 | 534 | 45 934 | 134 | 399 | 396 | 352 |
| TX-BK | 141 | 258 | 279 | 3 164 | 153 | 349 | 348 | 354 |
| TX-GO | 130 | 351 | 537 | 69 765 | 139 | 617 | 504 | 652 |
| WC-BK | 134 | 249 | 284 | 2 220 | 164 | 322 | 242 | 415 |
| WC-GO | 145 | 277 | 361 | 3 454 | 143 | 310 | 331 | 305 |

此外,随着数据集规模变化(如移动对象数量、社交关系数量、签到位置数量等),GSM 模型

操作符运行效率的波动不大,稳定性较高;相比较,PostgreSQL 实现方式的运行效率的波动较大。实验结果如图 4 所示。图 4(b)中,PostgreSQL 实现方式中的 Expand 操作符的运行效率出现波动的原因在于,实现过程中需要过多的表连接操作,随着社交关系数量的增加,表连接查询耗时增加较快;相反,在 GSM 模型中,Expand 扩展操作可以转化为 GSM 模型节点相连的边的遍历操作,社交关系数量的增长只会导致 GSM 模型 SocialGraph 中边的数量增长,但基本不影响遍历操作效率,体现了 GSM 模型的优势。图 4(c)中的 Cross 操作是在两个子图之间进行投影操作,在传统的 PostgreSQL 实现中,表的联合查询是不可避免的,但在 GSM 模型中,多个子图之间的联合投影同样可以避免表的连接操作,从而提高模型操作符的执行效率。图 4(d)所展示的 GSM 模型中多个子图之间的多维投影操作符 MultiCross 的执行效率随着签到位置数量的增长出现了波动,其原因有两个,一方面是在地理空间子图 GG 中仅构建了地理空间的 R 树索引,缺少对地理空间要素其他属性的索引,影响了多个子图间投影映射的效率;另一方面是在 GSM 模型实现过程中没有采用 Neo4J 的缓存配置与调优。Neo4J 的缓存策略在单子图的 Select、Expand 操作与两子图 Cross 操作中涉及的子图数量少,读写缓存影响不明显,但是在 MultiCross 操作中涉及多个子图之间的频繁操作,缓存配置优化的作用更加重要,造成 MultiCross 操作效率出现波动。在后续研究中,将进一步优化 GSM 模型的实现。

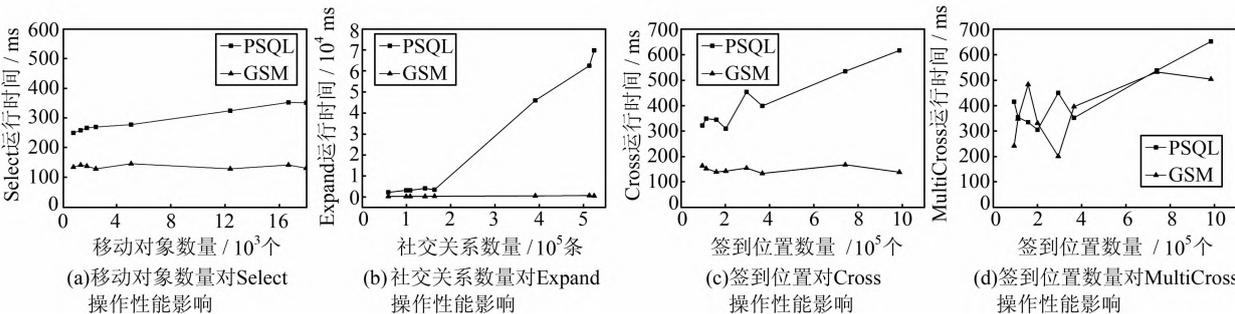


图 4 实验结果

Fig. 4 Experimental Results

2.3 讨论

1) 本文所采用的地理空间分割方法为基于 POI 的 Voronoi 划分,该方法的优势在于移动对象的位置签到数据与 POI 位置比较接近,能够很好地适应位置轨迹数据空间分布不均匀的情况,

但劣势在于造成剖分多边形数量过多,碎多边形现象严重。由于 POI 一般呈现聚集现象,针对该问题,在后续研究中将考虑先采用 POI 空间聚类,然后再进行地理空间剖分的方式。

2) GSM 模型支持对地理空间、移动对象轨

迹及移动对象社交关系的一体化建模,易于扩展,可以有效附加其他背景语义信息,包括移动对象室内位置表达。但在 GSM 模型实现过程中,将大量位置数据存储于图节点的属性中,造成图节点过大,会在一定程度上影响 GSM 模型的查询效率。针对此问题,后续工作中一方面可以采用轨迹压缩或二进制方式减少存储空间,另一方面可以采用指针,将位置数据放在外部存储。

3) 本文实现了 GSM 模型常用的四种基础操作:Select 截取操作、Expand 扩展操作、Cross 投影操作及 MultiCross 多维投影操作。后续研究中将进一步实现其他操作符,如 Union、Intersect、Difference 等。

4) 本文实验所采用的位置签到数据时间间隔较长,且签到时间间隔不规律。后续研究中,可以尝试在 GSM 模型中存储时间间隔在分钟级或者秒级的轨迹数据,如浮动车轨迹数据、手机信令数据等,并开展相关的查询实验,进一步验证与提高 GSM 模型的可靠性。

3 结 语

随着位置服务及社交网络的发展与融合,社交关系表达逐渐成为移动对象建模中不可忽略的因素。本文针对 SoLoMo 地理信息服务模式所亟需的移动对象轨迹与社交关系高效管理的需求,提出了一种移动对象轨迹及其社交关系的时空一体化表达模型 GSM,并给出了 GSM 模型常用的四种基本模型操作的实现方式,弥补了移动对象数据库系统中地理空间及社交网络数据管理的缺陷。GSM 模型首先根据 POI 空间分布,采用 Voronoi 图剖分地理空间,以寻求移动对象空间查询负载均衡和地理围栏的空间范围自适应;然后将剖分后的 Voronoi 多边形抽象为图的节点,移动对象的实时位置表达为移动对象与剖分多边形之间的实时映射关系,即动态边,移动对象之间的社交网络关系表达为移动对象之间的动态边;最后利用图结构实现了移动对象位置、轨迹及社交关系的时空一体化存储。该模型的优势在于实现了移动对象线上线下数据的有效融合,支持 Geo-Social 联合查询,可为基于位置的社交网络应用提供核心技术支持。为了验证 GSM 模型的有效性,本文采用图数据库系统 Neo4J 实现了 GSM 模型,且选择成熟的 PostgreSQL 开展对比实验,验证了 GSM 模型的有效性。

参 考 文 献

- [1] Li Deren, Shao Zhenfeng. The New Era of Geographic Information[J]. *Science in China (Series F: Information Sciences)*, 2009, 39(6):579-587(李德仁, 邵振峰. 论新地理信息时代[J]. 中国科学(信息科学), 2009, 39(6): 579-587)
- [2] Goodchild M F. Citizens as Sensors: the World of Volunteered Geography[J]. *GeoJournal*, 2007, 69(4): 211-221
- [3] Lu Feng, Zheng Nianbo, Duan Yingying, et al. Travel Information Services: State of the Art and Discussion on Crucial Technologies[J]. *Journal of Image and Graphics*, 2009, 14(7): 1 219-1 229(陆锋, 郑年波, 段滢滢, 等. 出行信息服务关键技术研究进展与问题探讨[J]. 中国图像图形学报, 2009, 14(7): 1 219-1 229)
- [4] Li Qingquan, Yang Bisheng, Zheng Nianbo. An Integrated Spatio-Temporal Data Model for GIS-Transportation and Related Applications[J]. *Geomatics and Information Science of Wuhan University*, 2007, 32(11): 1 034-1 041(李清泉, 杨必胜, 郑年波. 时空一体化 GIS-T 数据模型与应用方法[J]. 武汉大学学报·信息科学版, 2007, 32(11): 1 034-1 041)
- [5] Zhang Hengcai, Lu Feng, Chen Jie. Advance in Moving Object Data Modeling Under Geographic Network Environment[J]. *Journal of Geo-Information Science*, 2013, 15(3): 328-337(张恒才, 陆锋, 陈洁. 网络空间移动对象模型的应用与发展[J]. 地球信息科学学报, 2013, 15(3): 328-337)
- [6] Wolfson O, Xu B, Chamberlain S. Moving Objects Databases: Issues and Solutions[C]. *The 10th International Conference on Scientific and Statistical Database Management*, Washington D C, USA, 1998
- [7] Forlizzi L, Güting R H, Nardelli E. A Data Model and Data Structures for Moving Objects Databases [C]. *ACM SIGMOD*, Dallas, Texas, USA, 2000
- [8] Meng X, Ding Z. DSTTMOD: A Future Trajectory Based Moving Objects Database[J]. *Database and Expert Systems Applications*, 2003, 2 736: 444-453
- [9] Güting R H. Moving Object Databases[M]. *Germany: Morgan Kaufmann Publishers*, 2005:10-16
- [10] Vazirgiannis M, Wolfson O. A Spatiotemporal Model and Language for Moving Objects on Road Networks[J]. *Advances in Spatial and Temporal Databases*, 2001, 2 121: 20-35
- [11] Ding Z, Güting R. Modeling Temporally Variable Transportation Networks[J]. *Database Systems for*

- Advanced Applications*, 2004, 2 973:651-724
- [12] Speičevys L, Jensen C S, Kligys A. Computational Data Modeling for Network-Constrained Moving Objects[C]. The 11th ACM International Symposium on Advances in Geographic Information Systems, New York, USA, 2003
- [13] Chen J, Meng X, Guo Y, et al. Modeling and Predicting Future Trajectories of Moving Objects in a Constrained Network [C]. The 7th International Conference on Mobile Data Management, Washington D C, USA, 2006
- [14] Güting R H, Ding Z. Modeling and Querying Moving Objects in Networks[J]. *The VLDB Journal*, 2006, 15(2):165-190
- [15] Xu J, Güting R H. Manage and Query Generic Moving Objects in SECONDO[J]. *The VLDB Endowment*, 2012, 5(12): 2 002-2 005
- [16] Kim D O, Lee K J, Hong D S, et al. An Efficient Indexing Technique for Location Prediction of Moving Objects[J]. *Knowledge-Based Intelligent Information and Engineering Systems*, 2007, 4 693 :1-9
- [17] Kolahdouzan M, Shahabi C. Voronoi-based k Nearest Neighbor Search for Spatial Network Databases [C]. The 30th International Conference on Very Large Data Bases, Toronto, Canada, 2004
- [18] Sandu Popa I, Zeitouni K. Modeling and Querying-Mobile Location Sensor Data[C]. The 4th International Conference on Advanced Geographic Information Systems, Applications, and Services, Valencia, Spain, 2012
- [19] Cattuto C, Van den Broeck W, Barrat A. Dynamics of Person-to-person Interactions from Distributed RFID Sensor Networks[J]. *PLoS ONE*, 2010, 5 (7): e11596
- [20] Isella L, Stehlé J, Barrat A. What's in a Crowd? Analysis of Face-to-Face Behavioral Networks[J]. *Journal of Theoretical Biology*, 2011, 271 (1): 166-180
- [21] Atzmueller M, Doerfel S, Hotho A. Face-to-Face Contacts During a Conference: Communities, Roles, and Key Players[C]. The 2nd International Workshop on Mining Ubiquitous and Social Environments, Athens, Greece, 2011
- [22] Zhuang H, Chin A, Wu S. Inferring Geographic Coincidence in Ephemeral Social Networks[J]. *Machine Learning and Knowledge Discovery in Databases*, 2012, 7 524:613-628
- [23] Tang W, Zhuang H, Tang J. Learning to Infer Social Ties in Large Networks[J]. *Machine Learning and Knowledge Discovery in Databases*, 2011, 6 913:381-397
- [24] Tang J, Lou T, Kleinberg J. Inferring Social Ties Across Heterogenous Networks[C]. The 5th ACM International Conference on Web Search and Data Mining, Seattle, Washington, 2012
- [25] Zhuang H, Tang J, Tang W. Actively Learning to Infer Social Ties[J]. *Data Mining and Knowledge Discovery*, 2012, 25(2):1-28
- [26] Wang D, Pedreschi D, Song C. Human Mobility, Social Ties, and Link Prediction [C]. The 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 2011
- [27] Sadilek A, Kautz H, Bigham J P. Finding Your Friends and Following Them to Where You Are[C]. The 5th ACM International Conference on Web Search and Data Mining, New York, USA, 2012
- [28] Cranshaw J, Toch E, Hong J, et al. Bridging the Gap Between Physical Location and Online Social Networks[C]. The 12th ACM International Conference on Ubiquitous Computing, Seattle, Washington, 2010
- [29] Lee R, Wakamiya S, Sumiya K. Discovery of Unusual Regional Social Activities Using Geo-tagged Microblogs [J]. *World Wide Web*, 2011, 14(4): 321-349
- [30] Cho E, Myers S A, Leskovec J. Friendship and Mobility: User Movement in Location-based Social Networks[C]. The 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Diego, CA, USA, 2011
- [31] Eagle N, Pentland A, Lazer D. Mobile Phone Data for Inferring Social Network Structure [M]. US: Springer, 2008:79-88
- [32] Eagle N, Pentland A S, Lazer D. Inferring Friendship Network Structure by Using Mobile Phone Data[J]. *The National Academy of Sciences*, 2009, 106(36): 15 274-15 278
- [33] Scellato S, Mascolo C, Musolesi M. Distance Matters: Geo-social Metrics for Online Social Networks [C]. The 3rd Conference on Online Social Networks, CA, USA, 2010
- [34] Scellato S, Noulas A, Lambiotte R. Socio-spatial Properties of Online Location-based Social Networks[J]. *ICWSM*, 2011, 11: 329-336
- [35] Li N, Chen G. Analysis of a Location-based Social Network[C]. International Conference on Computational Science and Engineering, Washington D C, USA, 2009
- [36] Noulas A, Scellato S, Mascolo C. An Empirical Study of Geographic User Activity Patterns in Foursquare [C]. The 5th International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, 2011

- [37] Crandall DJ, Backstrom L, Cosley D. Inferring Social Ties from Geographic Coincidences [J]. *The National Academy of Sciences*, 2010, 107(52): 22 436-22 441
- [38] Pham H, Hu L, Shahabi C. GEOSO—A Geo-social Model: From Real-world Co-occurrences to Social Connections[J]. *Databases in Networked Information Systems*, 2011, 7 108:203-222
- [39] Zheng Y, Zhang L, Xie X. Recommending Friends and Locations Based on Individual Location History [J]. *ACM Transaction on the Web*, 2011,5(1):60-67

An Integrated Data Model for Spatial-Temporal Trajectories and the Social Relationships of Moving Objects

ZHANG Hengcai¹ LU Feng¹ CHEN Jie¹

1 State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

Abstract: In the pan geographic information age, with the development of the SoLoMo (social, local and mobile) modes in the mobile internet industry, how to effectively manage massive location-based data, social network and geospatial data has become a major challenge in the moving object databases field of research. In this paper, a spatial-temporal integrated data model for moving object trajectories and social relationships, called the Geo-Social-Moving (GSM) model, is proposed to manage both moving object geographical locations, trajectories and social relationships in an integrated framework. We define a series of common operations including interception, extension, projection and multi-dimensional projection. Then, the GSM model is implemented based on a graph database system. We conducted detailed experiments with two free location-based social network datasets from Brightkite and Gowalla to verified the effectiveness of the GSM model. Results show that the proposed GSM model can avoid query inefficiency caused by relational database table joins with the efficiency and stability to provide support for location-based social network applications.

Key words: moving objects; data model; trajectory; social network

First author: ZHANG Hengcai, PhD. He focus on moving objects database and spatial-temporal data mining. E-mail: zhanghc@lreis. ac. cn

Corresponding author: LU Feng, PhD, researcher, PhD supervisor. E-mail: luf@lreis. ac. cn

Foundation support: The National High Technology Research and Development Program of China(863 Program), Nos. 2012AA12A211, 2013AA120305; the National Natural Science Foundation of China, Nos. 41271408, 41101149.

(上接第 710 页)

neity, and the spatial interactions among cities, a hybrid geographic backbone extraction approach is proposed which is drawn from the gravity model and information entropy theories. This research can increase understanding of the structure of urban systems, radiant ability, attractiveness and openness of cities in the virtual web society.

Key words: social network; spatial structure; geographic backbone; Shannon entropy; gravity model

First author: CHANG Xiaomeng, PhD candidate, specializes in geographic social network, spatio-temporal data visual analysis, and ITS. E-mail: changxiaomeng@gmail. com

Corresponding author: LI Qingquan, PhD, professor. E-mail: liqq@szu. edu. cn

Foundation support: The National Natural Science Foundation of China, No. 41371377; Shenzhen Scientific Research and Development Funding Program, No. ZDSY20121019111146499; Shenzhen Dedicated Funding of Strategic Emerging Industry Development Program, No. JCYJ20121019111128765; CCF-TencentARG20130115.