

面向路网的不确定移动范围查询技术*

谷峪, 郭娜, 于戈

(东北大学 信息科学与工程学院, 辽宁 沈阳 110819)

通讯作者: 谷峪, E-mail: guyu@ise.neu.edu.cn

摘要: 随着移动定位技术和物联网技术的不断发展,时空查询技术受到了广泛关注.在实际的应用中,对象的移动方向和轨迹常受到空间网络限制并且位置信息往往带有不确定性.在以一般性的概率分布函数形式表示位置的不确定性的基础上,提出一种基于分割区间的概率查询增量处理模型和查询优化方法.考虑采用概率分布近似中心作为目标对象的估计位置,近似地解决普遍位置不确定性的问题,以较小的精度损失换取效率上的极大提高.最后,采用真实的路网数据集和模拟的对象分布,验证了模型和算法在效率和准确性方面均表现突出.

关键词: 路网空间;不确定;移动范围查询;分割区间;概率

中图法分类号: TP311 文献标识码: A

中文引用格式: 谷峪,郭娜,于戈.面向路网的不确定移动范围查询技术.软件学报,2013,24(6):1243-1262. <http://www.jos.org.cn/1000-9825/4320.htm>

英文引用格式: Gu Y, Guo N, Yu G. Uncertain moving range query techniques in road networks. Ruan Jian Xue Bao/ Journal of Software, 2013, 24(6): 1243-1262 (in Chinese). <http://www.jos.org.cn/1000-9825/4320.htm>

Uncertain Moving Range Query Techniques in Road Networks

GU Yu, GUO Na, YU Ge

(School of Information Science and Engineering, Northeastern University, Shenyang 110819, China)

Corresponding author: GU Yu, E-mail: guyu@ise.neu.edu.cn

Abstract: With the continuous development of the mobile positioning technology and the Internet, spatio-temporal query processing has drawn more and more attention. In the real situation, the directions and trajectories of mobile objects are usually restricted by an underlying spatial network, and the position information is usually uncertain. Based on the general probability distribution function (PDF) used to represent the uncertainty of the positions, incremental processing model and optimization methods for probabilistic query based on split intervals are proposed. By taking the probability distribution approximate center as the estimated position of the target objects, the general position uncertainty problem is solved and the efficiency is improved with a minor cost of accuracy. Finally, based on the real-life road network dataset and synthetic object distribution, the accuracy and efficiency of the proposed models and algorithms are verified.

Key words: road network; uncertain; moving range query; split interval; probability

随着无线通信技术、移动定位技术和物联网的不断发展,移动设备可以广泛应用在智能交通系统、自助旅游服务、数字化战场等应用中,提供各类基于位置的服务(LBS).由于时空信息数据量大,如何高速度低代价地处理时空查询,一直是时空数据管理领域的一个研究热点.在实际的情况下,铁路、公路、水运航线等空间网络环境限定了对象只能沿着预先设定的方向和路线移动.特别地,连续的范围查询(也称移动范围查询)是较常见的查询类型之一.具体地,网络环境下的移动范围查询定义为:查询点 q 沿着预先定义的路径移动,在每个查询周期,从位置已知且固定的对象集合 O 中找出与 q 的距离小于查询半径 r 的对象子集.需要说明的是,虽然查询点

* 基金项目: 国家自然科学基金(61003058, 61033007); 国家重点基础研究发展计划(973)(2012CB316201); 中央高校基本科研业务费专项资金(N110404006, N100704001)

收稿时间: 2010-08-11; 修改时间: 2012-04-17; 定稿时间: 2012-09-12

和目标对象必须按照路网行进,但是范围查询的距离计算可以按照欧式空间和网络空间两种方式进行,不同的距离计算方式可以对应不同的应用场景,衍生出两类不同的移动范围查询类型。

- 欧式距离移动范围查询

查询点 q 沿着预先定义的路径移动,在每个查询周期,从位置已知且固定的对象集合 O 中找出在以 q 所在位置 Q 点为中心、半径长度等于查询半径 r 的圆内的对象子集,查询范围用 $O(Q)$ 表示.该查询的查询区域是规则的圆形区域,但查询点的移动要受到路径的限制.例如,火车经过时可能产生噪音,要周期性查询可能影响到哪些居民区;再如,出租车司机组成互助小组,每个司机可以将自己周围的路况、出租车需求量等信息通过对讲机实时地传达给组内成员,而对讲人员要连续查询有哪些对讲人员在通信范围内.如图 1(a)所示,假设某查询周期查询点在 Q 处,各段路径长度在图中标出,若查询 4 公里范围的对象,EDMR 查询区域即为虚线圆包围的部分.可见,只有 O_1 和 O_2 在该查询区域内,所以 O_3 和 O_4 不属于查询结果.

- 网络距离移动范围查询

查询点 q 沿着预先定义的路径移动,在每个查询周期,从位置已知且固定的对象集合 O 中找出与 q 的网络距离小于查询半径 r 的对象子集.该查询的查询区域是由一些线段组成的,不再是闭合规则的形状.这些线段的起点在查询点处重合,终点是在从查询点出发沿着各种可能路径所能到达的最远点.例如,车载导航系统可以提供这样一种服务:在行驶的过程中,查询 4 公里范围内的加油站有哪些.显然,这种情况下,要计算到达目的地的网络最短路径.如图 1(b)所示,如果某查询周期查询点在 Q 处,各段路径长度在图中标出,若查询 4 公里范围内的对象,查询区域是与路径重合的虚线部分,斜线杠标识出了距查询点 4 公里的最远点,可见所有对象(O_1, O_2, O_3 和 O_4)都不在该查询区域内,因此返回空集.

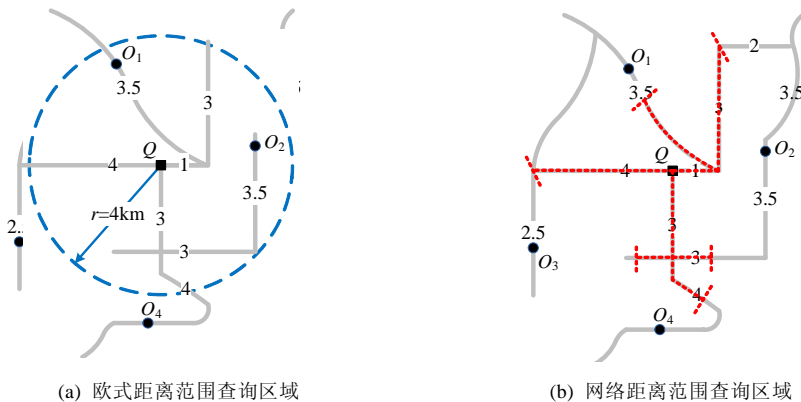


Fig.1 Examples of the two kinds of range queries

图 1 两种范围查询举例

特别地,由于定位技术本身以及设备精度、网络延迟、环境干扰、隐私保护等因素会导致我们所获得的位置信息带有不确定性,因此近些年来,越来越多的工作开始关注基于位置不确定性的空间查询技术.本文针对这两种范围查询,考虑网络空间环境下位置信息不确定的移动范围查询的特点,提出了基于分割区间概念的增量处理方法.利用分割区间将查询点的可能路径划分成多个有效区间,从而高效地过滤剪枝和增量的维护查询结果;如果查询点进入一个新的有效区间,查询不需要重新计算,只需要对查询结果进行增量维护.此外,在解决查询点和目标对象都存在不确定性的复杂情况时,考虑采用概率分布近似中心(PDAC)作为目标对象的估计位置,以较小的精度损失换取效率上的极大提高.

本文是在假设位置不确定性的前提下,对路网空间内的连续(移动)范围查询技术展开的研究,分别提出了基于欧式距离和网络距离的范围查询模型和增量处理优化算法.

第 1 节介绍相关工作.第 2 节给出基本的定义.第 3 节提出针对欧式距离范围查询的模型和算法.第 4 节进一步分析基于网络距离范围查询处理的区别和一些算法优化.第 5 节通过真实数据实验对相关的模型和算法

进行评估.最后一节总结全文.

1 相关工作

针对路网的空间查询处理技术得到了越来越多的关注,例如:文献[1]提出了一种基于 Voronoi 图的路网分割方法,该方法可以潜在地支持路网上的连续 k NN 查询;Cho 等人^[2]针对移动对象上路网空间内的最近邻查询和连续最近邻查询提出一种高效的可扩展的算法,将预计算的最近邻列表与 Dijkstra 算法相结合;文献[3]针对前期算法无法兼顾效率和存储的缺点,提出了对网络拓扑进行索引的网络压缩技术,将网络简化为互连的树形结构,并对其建立索引;Liu 等人^[4]讨论了空间网络环境下的动态范围查询,并给出了优化方法.以上工作都假设位置信息是确定的.

近年来,越来越多的工作在假设对象位置不确定的前提下研究各类空间查询.文献[5]研究了移动对象上基于不确定模型的概率范围查询的处理算法,用一种概率的形式表示对象可能的运动模型.考虑到大量对象可能带来非常高的计算代价,将不确定运动模型映射到二维空间以进行索引,然后用一种近似的方法决定最终结果.Kriegel 等人^[6]研究了如何处理不确定对象的概率最近邻问题.Trajcevski 等人^[7]讨论了基于不确定移动轨迹的连续最近邻查询的情况.Cheng 等人在文献[8,9]中分别讨论了不确定数据上的限制最近邻和阈值最近邻查询问题的概率处理算法.此外,Huang 等人在文献[10]中考虑了速度不确定的移动对象上的连续 k 最近邻查询算法.文献[11,12]在不确定数据上的反最近邻查询方面作出了贡献.Trajcevski 等人^[13]较早地研究了移动对象数据库中的不确定性问题,并提供了移动对象不确定性的一种概率模型,将轨迹用统一的随机过程向量形式表示,并高效地支持各类静态的时空范围查询.以上这些假设位置不确定的工作主要是针对欧式空间开展的,相关的技术无法直接应用到路网空间下的移动范围查询场景.

此外,有一些工作考虑由于低采样导致的轨迹不连续,通过推演来填补可能的中间位置.例如,在 Zheng 等人的工作中^[14],针对这种轨迹不确定性,对路网上的概率范围查询技术展开了深入的研究,并提出了一种新的索引方法.此外,有一些工作还关注了网络拓朴本身具有的不确定性.作为代表,文献[15]考虑了路网环境下的基于网络边权重不确定假设下的空间查询问题.需要说明的是,这些轨迹不确定性和网络拓朴不确定性与本文考虑的由于隐私保护和探测精度造成的位置不确定性有着不同的语义、模型和应用背景.

2 基本定义

表示位置的不确定性信息有很多方式,最具一般性的方法是给出目标可能出现的位置范围 S ,以及出现在 S 中对应位置可能性的概率密度函数(PDF) $f(s)$.例如,出于隐私保护的考虑,对于某一个体的位置信息只能透露在某个方形区域内,其在该区域内的任何一点的可能性相等;再如,GPS 定位技术只能每隔一段时间周期地报告目标位置,如果一个移动的机器人在上一次报告中的位置是 (x,y) ,那么在下一个周期之前的位置可以根据机器人速度变化区间 $[v_{\max},v_{\min}]$ 确定到某个范围.需要注意的是,在路网环境下,位置分布只能在路径上,因此需要用一维的 PDF 表示带不确定性的位置信息.

定义 1(可能位置区域(probable location region)). 在路网环境中,一个对象 o_i 可能的位置分布区域是由一条或多条连续的路段组成,用 PLR_i 表示.对象 o_i 的真实位置可能且仅可能是 PLR_i 内的任何一点.用 $PLR[j]$ 可以访问到第 j 个段(任意顺序), $|PLR[j]|$ 表示第 j 段长度.简称为 PLR.

定义 2(可能位置分布(probable location distribution)). 一个对象 o_i 可能的位置分布是由 PLR_i 中的段和对段上 PDF 组成的二元组集合,可以形式化地表示为

$$PLD_i = \left\{ \left\langle [s_j, e_j]_{eid}, p_j(x) \right\rangle \mid P = \sum_{j=1}^{j \leq n} P_j = \sum_{j=1}^{j \leq n} \int_0^{d(s_j, e_j)} p_j(x) dx = 1, 1 \leq j \leq n \right\},$$

其中, $[s_j, e_j]$ 表示边 e_{eid} 上的以 s_j 为起点、以 e_j 为终点的路段; $p_j(x)$ 表示在该段上的概率密度函数; x 取值范围是 $[0, d(s_j, e_j)]$, 表示距 s_j 的距离; 需要满足在所有路段上的概率分布之和为 1. 简称为 PLD.

定义 3(可能位置区域边缘点(edge point)). 可能位置区域边缘点是 PLR 中的最外点,是进入或者离开 PLR

的必经点.边缘点集合可定义为 $\{s_j\} \cup \{e_j\} - S'$, 其中, S' 表示任意两个路段 $[s_j, e_j]_{eid}$ 存在端点重合的那些端点集合. PLR 可以用 EP 的运算符形式获得其边缘点集合.

PLR 的边缘点可能多于两个, 如图 2 所示, sn_1 和 sn_2 附近的 q .PLR 有两个边缘点, sn_3 附近的 q .PLR 有 4 个边缘点 $\{EP5, EP6, EP7, EP8\}$, sn_4 附近的 q .PLR 有 3 个边缘点 $\{EP9, EP10, EP11\}$.

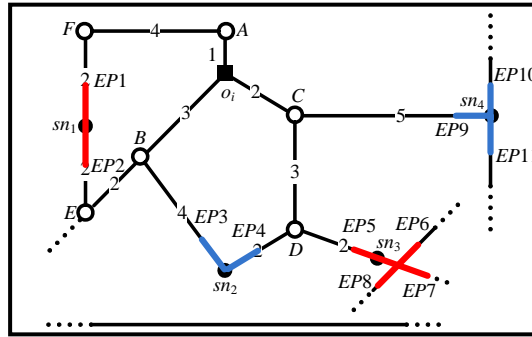


Fig.2 Illustration of PLR

图 2 PLR 示例

位置的不确定性必然造成某些目标对象不能确定是否在查询范围内(可能在,可能不在),在查询处理过程中,我们用二元组 $\langle o_i, p_i \rangle$ 表示中间结果,表示 o_i 在查询范围内的概率为 p_i , 当 p_i 取不同值时,有如下的不同意义:

$$p_i = \begin{cases} 0, & o_i \text{一定不在查询范围内} \\ \rho (0 < \rho < 1), & o_i \text{可能在查询范围内} \\ 1, & o_i \text{一定在查询范围内} \end{cases}$$

那么,根据不同的距离计算方法来确定查询范围,定义两种常用的基于不确定位置的移动范围查询. 设 d_E 和 d_N 分别表示取欧式距离和网络距离:

定义 4(概率网络距离移动范围查询(probabilistic network distance moving range query,简称 NDMR)).

给定概率阈值 θ , 查询点 q 在路网中移动, 在每个查询周期, 从对象集合 O 中找出 $\{o_i | o_i \in O, P(d_N(q, o_i) \leq r) \geq \theta\}$ 作为查询结果.

定义 5(概率欧式距离移动范围查询(probabilistic euclidean distance moving range query,简称 EDMR)).

给定概率阈值 θ , 查询点 q 在路网中移动, 在每个查询周期, 从对象集合 O 中找出 $\{o_i | o_i \in O, P(d_E(q, o_i) \leq r) \geq \theta\}$ 作为查询结果.

3 欧式距离范围查询不确定性处理

在某些应用中, 查询点和目标对象的位置不确定性不是同时存在的. 例如, 某些服务提供商希望得到最大的经济效益, 会尽可能地让顾客知道其详细地址, 而顾客要考虑自己的安全问题、确保正常生活不被打扰, 会要求位置信息隐私保护, 那么这时只有查询点存在位置不确定性; 再比如在紧急情况下, 查询者查询某些私人地址(要求保证隐私安全)需要尽可能准确的位置信息, 所以不考虑自己的隐私问题, 那么这时只有目标对象存在位置不确定性. 下面我们先分别考虑查询点和目标对象的位置不确定性, 再考虑同时存在时的复杂情况.

3.1 查询点位置不确定性处理

对于连续的查询, 避免每个周期都重新计算查询结果, 本文提出了一种增量处理方法: 利用将查询点的可能路径划分成多个有效区间, 即空间网络中的一个连续段, 只要查询点在有效区间内, 其查询结果保持不变; 如果查询点进入一个新的有效区间, 查询不需要重新计算, 只需要对查询结果进行增量维护. 在介绍算法之前, 先给出一些关键的定义.

定义 6(分割点(split node)). 若给定查询半径,分割点是路网图中位于边或顶点上的一个标志性位置,其位置是根据目标对象的位置计算得出的,如果查询点经过这个位置,那么该查询的结果就可能发生改变.用 sn 表示分割点,函数 $o(sn)$ 返回该分割点的关联对象,即可能影响查询结果的对象.

定义 7(有效区间(valid interval)). 对于查询 q 的分割点集合 SN ,在查询点可能的查询路线上的每对相邻分割点 sn_i 和 $sn_j(sn_i, sn_j \in SN)$ 之间的路段区间 $[sn_i, sn_j]$ 都是该 q 的有效区间,在每个这样的区间内 q 的查询结果保持不变.

根据查询点在经过分割点时与该分割点关联对象距离的变化情况,将分割点分为极大值分割点、极小值分割点、非极值分割点这 3 类.极值分割点是一种不会影响查询结果的分割点.若查询点从某分割点出发沿着各个方向移动,与目标点的最短路径距离都增大,这样的分割点为极小值分割点;反之,若查询点从某分割点出发沿着各个方向移动,与目标点的最短路径距离都减小,这样的分割点为极大值分割点;既不是极大值分割点也不是极小值分割点的分割点为非极值分割点.例如在图 3 中,空心圆点表示顶点,实心圆点表示分割点,实心方形点为 o_i 所在的位置.若此时查询半径为 5km,根据分割点的定义,可以找出关于 o_i 的 5 个分割点 sn_5, sn_6, sn_7, sn_8 和 sn_9 ,若查询点从 sn_5 和 sn_7 出发沿着各个方向移动,与 o_i 的最短路径距离都增大,为极小值分割点;若查询点从 sn_6 出发沿着各个方向移动,与 o_i 的最短路径距离都减小,因此为极大值分割点;而 sn_8 和 sn_9 没有极值分割点的特性,是非极值分割点.

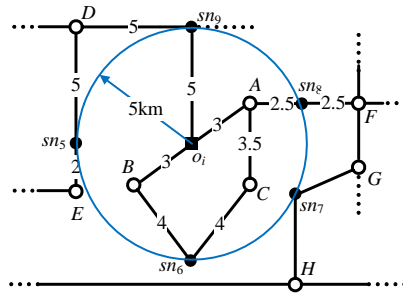


Fig.3 Examples of the extreme points in EDMR
图 3 EDMR 查询的极值点示例

每个分割点需要存储一些必要的信息,所以分布在边上的分割点表示为四元组 $esn(oid, flag, direction \downarrow, position)$,分布在顶点处的分割点表示为四元组 $vsn(oid, flag, set \downarrow, set \uparrow, position)$,其中, oid 表示该分割点所关联的对象标识, $id, flag$ 表示该点是否是极值点(0:不是极值点;-1:是极小值点;1:是极大值点), $direction \downarrow$ 指出 $d(q, o_{oid})$ 减小的方向(从 vl 到 vr 或者从 vr 到 vl , 分别用表示 $vl \rightarrow vr$ 和 $vr \rightarrow vl$ 表示), $set \downarrow$ 和 $set \uparrow$ 分别是 $d(q, o_{oid})$ 减小和增大方向的相邻顶点集合, $position$ 用来存储相对位置信息.

目标对象的位置是准确的,那么可以用几何方法得到分割点集合. SN_{oid} 是关联 o_{oid} 的分割点, $q.PLR$ 表示查询点 q 可能的位置分布区域, $SN_{oid} \Theta q.PLR$ 表示 SN_{oid} 在 $q.PLR$ 中的分割点集, 函数 $Num(SN_{oid} \Theta q.PLR)$ 返回 $SN_{oid} \Theta q.PLR$ 中元素个数. 根据 $Num(SN_{oid} \Theta q.PLR)$, 分 4 种情况进行处理:

情况 1. $Num(SN_{oid} \Theta q.PLR) = 0$

此类情况不能确定 o_{oid} 是否在查询范围内,但是能确定如果在则概率一定为 1, 否则一定不在 ($p_{oid} = 0$). 需要根据在本周期以前和上一周期之后经过了哪些分割点以及上一周期 $p_{oid} > 0$ 的 q 点位置集分布情况来判断. 在两周期之间经过的极值分割点不影响查询结果, 但会给出重要提示信息, 关联同一对象的分割点中最后经过的是起决定作用的分割点, 用 d_{sn} 表示. 对于极小值分割点, 如果 q 在他们周围(不存在其他关联相同对象 o_i 的分割点的范围)的任何一个位置, 都会有 $d(q, o_i) > r$ 成立; 对于极大值分割点, 如果 q 在他们周围的任何一个位置, 都会有 $d(q, o_i) < r$ 成立. 设 q 点经过 sn 的方向为 $q.dirac(sn)$, 根据上一周期的结果信息以及周期间经过的分割点集推断在新周期内的某对象的存在概率总结见表 1.

Table 1 Existence probability judgements for Num($SN_{oid} \Theta q.PLR$)=0

表 1 Num($SN_{oid} \Theta q.PLR$)=0 时的存在概率判定表

上个查询周期结果 经过的分割点集 P_{oid} 值	无 极小值 分割点 极大值 分割点			$dsn.direction \downarrow$ 与 $q.direction(dsn)$ 相同	$dsn.direction \downarrow$ 与 $q.direction(dsn)$ 相反
	无	极小值 分割点	极大值 分割点		
$p_{oid}=0$	0	0	1	1	0
$p_{oid}=1$	1	0	1	1	0
$0 < p_{oid} < 1 \wedge \neg eep \Theta LR(p_{oid} > 0)$	0	0	1	1	0
$0 < p_{oid} < 1 \wedge eep \Theta LR(p_{oid} > 0)$	1	0	1	1	0

情况 2. Num($SN_{oid} \Theta q.PLR$)=1, 设 $sn \in SN_{oid} \Theta q.PLR$, 且 $sn \in E$

当 $sn.flag=-1$ 时, o_{oid} 直接产生中间结果 $\langle o_{oid}, 0 \rangle$. 因为 sn 是极小值点, 那么, 对于 $q.PLR$ 上任意一点 q_i 都有 $d(q_i, o_{oid}) \geq r$ 成立, 则 o_{oid} 一定不在查询范围内. 当 $sn.flag=0$ 时, 则 o_{oid} 可能在查询范围内, 需要计算其存在概率 p 值. 因为 sn 不是极值点, 那么对于 PLR 上任意一个与 sn 不重合的点 q_i : 若与 $sn.set \downarrow$ 中的点在 sn 的同侧, 则有 $d(q_i, o_{oid}) < r$ 成立; 否则, $d(q_i, o_{oid}) > r$. 假设当前时刻 $q.PLR$ 共有 n 段, sn 所在段为 $PLR[m+1]$, $sn.direction \downarrow = s_{m+1} \rightarrow e_{m+1}$, 在 s_{m+1} 端的段为 $PLR[1 \dots m]$, 在 e_{m+1} 端的段为 $PLR[m+2 \dots n]$, 如图 4, 则

$$p(o_{oid}) = \int_{d(s_{m+1}, sn)}^{d(s_{m+1}, e_{m+1})} p_{m+1}(x) dx + \sum_{i=m+2}^{i \leq n} \int_0^{d(s_i, e_i)} p_i(x) dx.$$

特别地, 当 q 在 PLR 内的位置在内均匀分布时, o_{oid} 存在概率等于 $q.PLR$ 在 o_{oid} 的作用区间内的部分所占长

度的百分比, 即 $p = \frac{|e_{m+1}, sn| + \sum_{i=1}^{m+2} |PLR[i]|}{\sum_{i=1}^n |PLR[i]|}$.

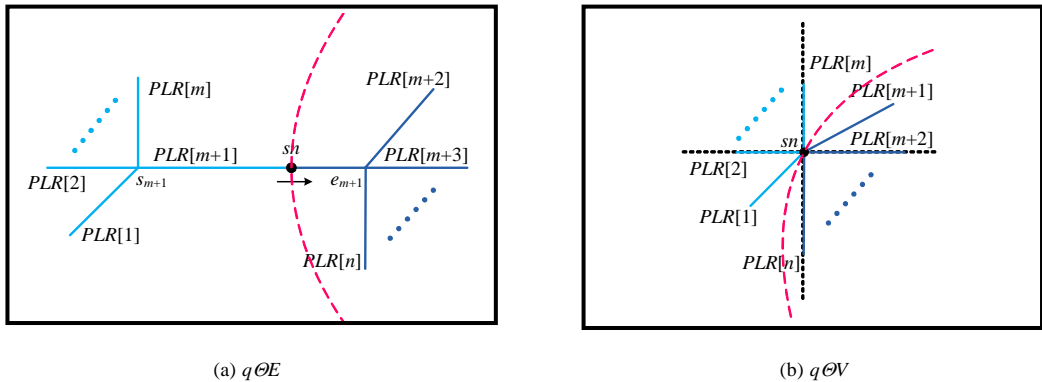


Fig.4 Examples of the location uncertainty in query point

图 4 查询点位置不确定性的例子

情况 3. Num($SN_{oid} \Theta q.PLR$)=1, 设 $sn \in SN_{oid} \Theta q.PLR$, 且 $sn \in V$

当 $sn.flag=1$ 时, o_{oid} 直接产生中间结果 $\langle o_{oid}, 1 \rangle$. 因为 sn 是极大值点, 那么对于 $q.PLR$ 上任意一点 q_i 都有 $d(q_i, o_{oid}) \leq r$ 成立, 则 o_{oid} 一定在查询范围内. 当 $sn.flag=-1$ 时, o_{oid} 直接产生中间结果 $\langle o_{oid}, 0 \rangle$. 因为 sn 是极小值点, 那么对于 $q.PLR$ 上任意一点 q_i 都有 $d(q_i, o_{oid}) \geq r$ 成立, 则 o_{oid} 一定不在查询范围内. 当 $sn.flag=0$ 时, 则 o_{oid} 可能在查询范围内, 需要计算其存在概率 p 值. 因为 sn 不是极值点, 那么对于 $q.PLR$ 上任意一个与 sn 不重合的点 q_i : 若与 $sn.set \downarrow$ 中的点在 sn 的同侧, 则有 $d(q_i, o_{oid}) < r$ 成立; 否则, $d(q_i, o_{oid}) > r$. 假设当前时刻 $q.PLR$ 共有 n 段, sn 所在段为 $PLR[m+1]$, $sn.direction \downarrow = s_{m+1} \rightarrow e_{m+1}$, 与 $sn.set \uparrow$ 中的点在 sn 同侧的段为 $PLR[1 \dots m]$, 与 $sn.set \downarrow$ 中的点在 sn 同侧的

段为 $PLR[m+1\dots n]$, 则 $p(o(sn)) = \sum_{i=m+1}^{i \leq n} \int_0^{d(s_i, e_i)} p_i(x) dx$. 特别地, 当 q 的位置在 PLR 内均匀分布时, o_{oid} 的存在概率等

$$\text{于 } q.PLR \text{ 在 } o_{oid} \text{ 的作用区间内的部分长度百分比, } p = \frac{\sum_1^{m+1} |PLR[i]|}{\sum_1^n |PLR[i]|}.$$

情况 4. $\text{Num}(SN_{oid} \Theta q.PLR) > 1$, 设极值点集合为 ESN_{oid} :

若 $\text{Num}(SN_{oid} \Theta q.PLR - ESN_{oid}) = 1$ 且 $sn \Theta E$, 则忽略 ESN_{oid} , 用情况 2 的方法处理;

若 $\text{Num}(SN_{oid} \Theta q.PLR - ESN_{oid}) = 1$ 且 $sn \Theta V$, 则忽略 ESN_{oid} , 用情况 3 的方法处理;

若 $\text{Num}(SN_{oid} \Theta q.PLR - ESN_{oid}) > 1$, 在集合 $(SN_{oid} \Theta q.PLR - ESN_{oid}) \cup q.PLR.EP$ 中, 找出满足下列条件的二元组 (α, β) 集合:

- 1) α, β 可以是单个的分割点或边缘点, 也可以是由多个分割点或边缘点组成的向量;
- 2) α 和 β 之间包含的区域必须是连续的路段, 不包含其他分割点;
- 3) α 和 β 中是分割点的元素, 其 $direction \downarrow$ 的方向指向对方.

这样确定出的区间是对象 o_{oid} 可能成为查询结果 ($p_{oid} > 0$) 的 q 点位置集, 那么 p_{oid} 的值就是 q 点在该位置集中的概率和, 即各段的分布概率积分之和.

3.2 目标对象位置不确定性处理

目标对象的位置带有不确定性, 同样需要用 PLD 代替点坐标来表示. 在基于精确位置的处理算法中, 一个核心的组成部分——分割点位置的确定是根据可能在查询范围内的对象的位置以及查询半径确定出来的, 由于目标对象精确的位置信息是一个点, 所以确定出来的每个与其相关联的分割点位置也是一个点. 那么用 PLD 确定出来的分割点位置形式也会发生变化.

定义 8(分割区间(split interval)). 分隔区间即分割点可能的位置分布情况, 根据所关联的对象 PLD, 确定分割点的 PLD.

NDMR 查询的距离计算方式是沿着路径延伸的, 所以 NDMR 查询分割区间的计算可以看做是关联对象的 PLD 沿着路径的逻辑平移和叠加(在下一节中详细讨论). 而 EDMR 查询的距离计算方式在空间中是自由的, 与 NDMR 查询的方法不同, 需要更复杂的过程.

3.2.1 确定 PLR

o_i 可能的作用范围是以 $o_i.PLR$ 中的每个点为圆心、查询半径 r 的长度为半径所构造出的所有圆覆盖的范围. 按照对象 o_i 的 PLR 是否跨顶点的性质分为两种情况:

- 1) 若 $o_i.PLR$ 在同一条边上, 则可能的作用范围取极限的位置有 2 个, 即圆心的位置与边缘点重合, 如图 5(a) 所示, 可能的作用范围形状是由两个半圆和夹在中间的一个矩形组合而成;
- 2) 若 $o_i.PLR$ 跨过某个顶点, 那么极限位置(影响最终覆盖范围的位置)有 2 个或以上, 圆心的位置与边缘点或者拐点重合, 如图 5(b) 所示, 可能的作用范围的形状是由多个情况 1) 中的形状带重叠的组合.

定义 9(可能作用范围的最大/最小边界(maximal/minimal boundary of active region)). 一定处于 o_i 作用范围内的点集合是极限位置处的可能作用范围的重叠区域 $\min R, \min R$ 的边界是可能作用范围的最小边界(min-boundary). 所有可能处于 o_i 作用范围内的点的集合是 $o_i.PLR$ 中任何一个位置处的作用范围所覆盖的区域 $\max R, \max R$ 的边界是可能作用范围的最大边界(max-boundary).

如图 5 所示虚线标识的闭合曲线中处于外部的为 Max-Boundary, 处于内部的为 Min-Boundary. 可以看出, 可能作用区域中每个点的概率等于多个圆形区域的概率在此点上的叠加, 颜色越深表示概率密度越大.

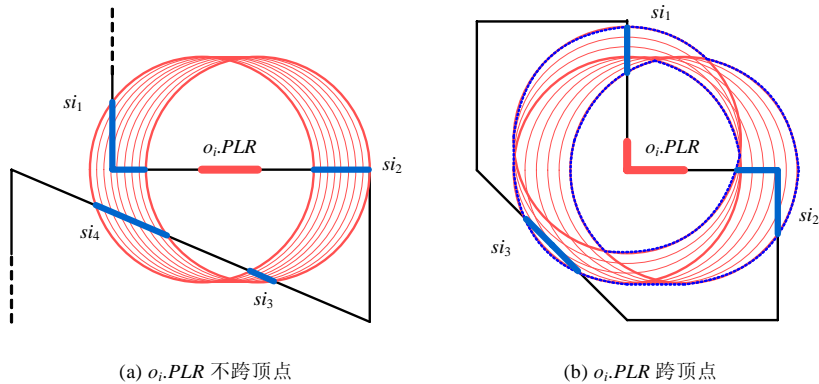


Fig.5 Split interval examples

图 5 分割区间示例

根据分割区间对查询结果的影响程度将分割区间分为两类:

- 1) 确定性的分割区间,分别以可能作用范围的 Max-Boundary 和 Min-Boundary 与路网图中边的交点为边缘点的连续区间;
 - 2) 非确定性的分割区间,以可能作用范围的 Max-Boundary 与路网图中边的交点为边缘点的连续区间.
- 每个对象的分割区间要存储必要的信息,可以形式化表示为四元组的形式 $si(oid,flag,direction\downarrow/dp,PLD)$.

其中,

- oid 表示该分割区间所关联的对象标识 id;
- $flag$ 表示该点是否为确定性分割区间(0:否;1:是);
- $flag=0,direction\downarrow$ 指出 $d(q,o_{oid})$ 减小的方向(从一个边缘点指向另一个边缘点),具有唯一性;若 $flag=1$,此项存储进入和离开的分割点分界点 dp ;
- PLD 用来存储位置分布信息,每个分割区间可能包含多个段,用 $PLD=\{([s_i,e_i],p_i(x))|1\leq i\leq n\}$ 来表示.

3.2.2 确定 PDF

若查询周期即当前时刻 q 不在 o_i 的任何分割区间内时,需要回顾周期期间经过的分割区间.确定性的分割区间在增量维护过程中会对结果产生影响,若为 q 最新经过的关联对象的分割区间,与其 $direction\downarrow$ 相同则将关联的对象插入到结果集中,与其 $direction\downarrow$ 相反从结果集中删除.非确定性分割区间没有方向性,完全经过的情况下不影响查询结果,但会给出重要提示信息.

若查询周期 q 处于 o_i 的某个分割区间内时, o_i 的存在概率计算需要基于分割区间的 PDF.下面讨论如何确定分割区间的 PDF.

(1) 关联对象的 PLR 不跨顶点时,确定性分割区间的 PDF 确定.

若分割点在确定性分割区间内的某点 v 处,设决定该分割点的作用范围是 O_r,O_i 的圆心在 $o_i.PLR$ 内的位置可以确定出来,即与 v 点的距离等于查询半径的点(如图 6(a)所示)是一一对应的,那么分割点在 v 点的概率等于 o_i 在 O_i 的圆心处的概率.可见,两个 PDF 的变化趋势是完全相同的,但是变化区间的长度不等,而且 PDF 是与距边缘点的距离成正比的,所以只需将变化参数的值域扩大或缩小相应倍数即可,如图 6(a)所示.

$$o_i.PLD = \{([s_1,e_1], p_1(x))\}, si_2.PLD = \left\{ \left[[s_2,e_2], p_1 \left(\frac{|si.PLR|}{|o_i.PLR|} x \right) \right] \right\},$$

而 si_3 的积分区间端点顺序与 o_i 的相反:

$$si_3.PLD = \left\{ \left[[s_2,e_2], p_1 \left(\frac{|si.PLR|}{|o_i.PLR|} x \right) \right], \left[[s_3,e_3], 1 - p_1 \left(\frac{|si.PLR|}{|o_i.PLR|} x \right) \right] \right\}.$$

(2) 关联对象的 PLR 跨顶点时,确定性分割区间的 PDF 确定.

$o_i.PLD$ 含有两个段 $[s_1,m_1]$ 和 $[m_1,e_1]$,设以 m_1 点为圆心、 r 为半径的圆与 si_2 的交点为 m_2 .设 o_i 在 $o_i.PLR$ 的 $[s_1,m_1]$ ($[m_1,e_1]$)区间内的某点 $u_1(v_1)$ 处的作用范围与 $si_2.PLD$ 的交点为 $u_2(v_2)$,那么 $u_2(v_2)$ 一定在 $si_2.PLD$ 的区间 $[s_2,m_2]$ ($[m_2,e_2]$)上.那么分割点在 $u_2(v_2)$ 点的概率等于 o_i 在 $u_1(v_1)$ 处的概率.由于 si_1 是确定性分割区间,所以仍然有一一对应的性质:变化趋势是完全相同的,变化区间长度不等.如图 6(b)所示, $o_i.PLD$ 含有两个段,对应段内的两个 PDF 的变化趋势相同,那么 $si_2.PLD$ 也分两段,只需将对应段的 PDF 变化参数的值域扩大或缩小相应倍数.若 $o_i.PLD = \{ \langle [s_1,m_1], p_1(x) \rangle, \langle [m_1,e_1], p_2(x) \rangle \}$, 则

$$si_2.PLD = \left\{ \left\langle [s_2,m_2], p_1 \left(\frac{|si.PLDR[1]|}{|o_i.PLDR[1]|} x \right) \right\rangle, \left\langle [m_2,e_2], p_2 \left(\frac{|si.PLDR[2]|}{|o_i.PLDR[2]|} x \right) \right\rangle \right\}.$$

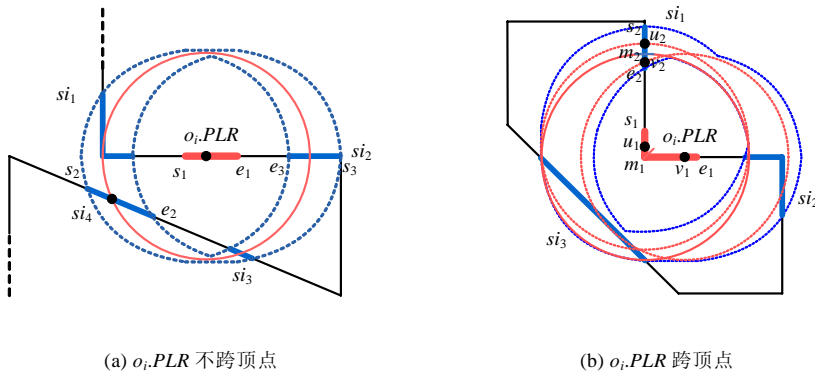


Fig.6 Confirmation process of PDF for the determined split interval
图 6 确定性分割区间的 PDF 确定过程

对于任意一个对象 o_i ,其非确定性分割区间内总能找到这样一点 M 或区间 I :它的两边分别分布的是进入和离开 o_i 作用区间的分割点的可能位置.该点或区间不一定是分割区间的中点(或区间仅在分割区间与 $o_i.PLR$ 平行或垂直时在中心位置),其位置与 $o_i.PLR$ 的夹角有关.由于非确定性分割区间仅是 Max-Boundary 与路网图中中边的交点,与 Min-Boundary 不相交,所以 $o_i.PLR$ 内可能会存在一部分位置的 o_i 作用范围 O_i 与该分割区间无交点.分界位置 N 可能是当 O_i 与分割区间的切点(唯一交点)为 M 时 O_i 的圆心位置.以 N 点为界,距离分割区间较近一侧的所有位置参与构成分割区间.对于跨顶点的分割区间,关联对象在一个段上可能不会影响整个分割区间,影响区域间隔区间为 I .下面将详细距离说明.

(3) 关联对象的 PLR 不跨顶点时,非确定性分割区间的 PDF 确定.

如图 7(a)所示, $o_i.PLD$ 中以 m_1 和 m_3 为圆心、 r 为半径的圆与 si_2 和 si_3 分别相切于 m_2 和 m_4 .若分割点在区间 $[s_2,m_2]$ ($[m_2,e_2]$)内的任意一点 v 处,设决定该分割点的作用范围是 O_i , O_i 的圆心在 $o_i.PLR$ 内的位置可以确定出来(在区间 $[s_1,m_1]$ 内),即与 v 点的距离等于查询半径的点.由于连续分布所以是一一对应的,那么分割点在 v 点的概率等于 o_i 在 O_i 圆心处的概率.同理,只需将变化参数的值域扩大或缩小相应倍数.若 $o_i.PLD = \{ \langle [s_1,e_1], p_1(x) \rangle \}$:

$$si_2.PLD = \left\{ \left\langle [s_2,m_2], p_1 \left(\frac{|s_1,m_1|}{|s_2,m_2|} x \right) \right\rangle, \left\langle [m_2,e_2], 1 - p_1 \left(\frac{|s_1,m_1|}{|m_2,e_2|} x \right) \right\rangle \right\},$$

$$si_3.PLD = \left\{ \left\langle [s_3,m_4], 1 - p_1 \left(\frac{|m_3,e_1|}{|s_3,m_4|} (x + |s_1,m_3|) \right) \right\rangle, \left\langle [m_4,e_3], p_1 \left(\frac{|m_3,e_1|}{|m_4,e_3|} (x + |s_1,m_3|) \right) \right\rangle \right\}.$$

(4) 关联对象的 PLR 跨顶点时,非确定性分割区间的 PDF 确定.

与步骤(2)相似,对于跨顶点的非确定性分割区间需要分段进行分析. $o_i.PLD$ 含有两个段 $[s_1,m_1]$ 和 $[m_1,e_1]$,设以 m_1 点为圆心、 r 为半径的圆 O_1 与 si_3 的交点为 m_3 ,以 s_1 点为圆心、 r 为半径的圆 O_2 与 si_3 的

边缘点为与 O_1 和 O_2 构成的 Max-Boundary 的交点,所以如图 7(b)所示,若 o_i 若分布在 $[s_1, m_1]$,那么分割点可能的位置区间为 $[s_2, m_3]$ 和 $[m_4, e_2]$. 若 $\langle [s_1, m_1], p_1(x) \rangle \in o_i.PLD$, 则 $[s_2, m_3]$ 和 $[m_4, e_2]$ 上的关于 o_i 分布于 $[s_1, m_1]$ 内的 $si_3.PLD$ 分别为 $\left\langle [s_2, m_3], p_1 \left(\frac{|o_i.PLR[1]|}{|s_2, m_3|} x \right) \right\rangle$ 和 $\left\langle [m_4, e_2], p_2 \left(\frac{|o_i.PLR[1]|}{|m_4, e_2|} x \right) \right\rangle$. 若 o_i 若分布在 $[m_1, e_1]$, 那么分割点可能的位置区间为 $[m_3, e_2]$. 以 m_2 为圆心、以 r 为半径的圆与 si_3 相切于 m_5 , 若 $\langle [m_1, e_1], p_2(x) \rangle \in o_i.PLD$, 则关于 o_i 分布于 $[m_1, e_1]$ 内的 $si_3.PLD = \left\{ \left\langle [m_3, m_5], p_1 \left(\frac{|o_i.PLR[2]|}{|m_3, m_5|} x \right) \right\rangle, \left\langle [m_5, e_2], 1 - p_2 \left(\frac{|o_i.PLR[2]|}{|m_5, e_2|} x \right) \right\rangle \right\}$. 将 o_i 分布于 $o_i.PLD$ 内各段的 $si_3.PLD$ 合并, 得到最终

$$si_3.PLD = \left\{ \left\langle [s_2, m_3], p_1 \left(\frac{|o_i.PLR[1]|}{|s_2, m_3|} x \right) \right\rangle, \left\langle [m_3, m_5], p_1 \left(\frac{|o_i.PLR[2]|}{|m_3, m_5|} x \right) \right\rangle, \left\langle [m_4, m_5], p_2 \left(\frac{|o_i.PLR[1]|}{|m_4, e_2|} x \right) \right\rangle, \left\langle [m_5, e_2], 1 - p_2 \left(\frac{|o_i.PLR[2]|}{|m_5, e_2|} x \right) + p_2 \left(\frac{|o_i.PLR[1]|}{|m_4, e_2|} (x + |m_4, e_2|) \right) \right\rangle \right\}.$$

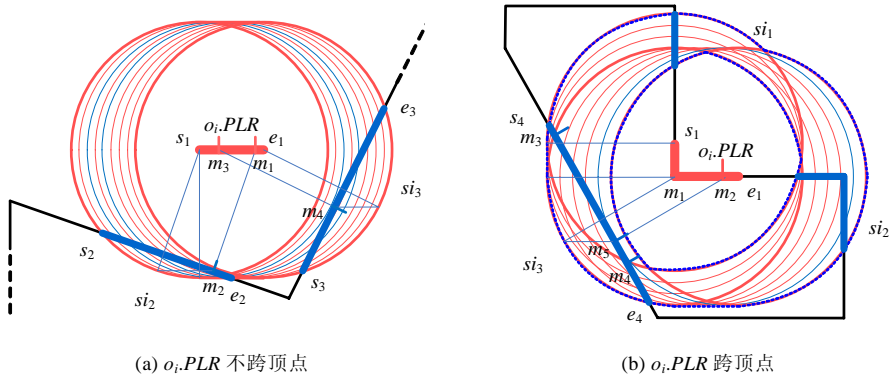


Fig.7 Confirmation process of PDF for the uncertain split interval

图 7 非确定性分割区间的 PDF 确定过程

若已知分割区间的 PLD, 当查询点在分割区间内分布时, 计算关联对象 o_i 在查询结果中的存在概率分两种情况:

- 情况 1. q 在确定性分割区间内

设 $si.PLD = \{ \langle [s_1, e_1], p_1(x) \rangle \}$: 若 q 的移动方向与 $si.direction \downarrow$ 相同, 如图 8(a) 中的情形(1), o_i 在查询结果中的存在概率 $p(o_i) = \int_{d(s_1, s_1)}^{d(s_1, q)} p_1(x) dx$; 若 q 的移动方向与 $si.direction \downarrow$ 相反, 如图 8(a) 中的情形(2), o_i 在查询结果中的存在概率 $p(o_i) = 1 - \int_{d(s_1, s_1)}^{d(s_1, q)} p_1(x) dx$.

- 情况 2. q 在非确定性分割区间内

若关联对象的 PLR 不跨顶点时, 设 $si.dp = D, si.PLD = \{ \langle [s_1, D], p_1(x) \rangle, \langle [D, e_1], p_2(x) \rangle \}$, q 的移动方向如图 8(b) 中情形(1)的箭头标识. 当 q 在区间 $[s_1, D]$ 内某一点 u 处, $p(o_i) = \int_{d(s_1, s_1)}^{d(s_1, u)} p_1(x) dx$; 当 q 在区间 $[D, e_1]$ 内某一点 v 处,

$$p(o_i) = 1 - \int_{d(D, D)}^{d(D, v)} p_2(x) dx.$$

若关联对象的 PLR 跨顶点时, 设 $si.dp = D, si.PLD = \{ \langle [s_1, m_1], p_0(x) \rangle, \langle [m_1, m_2], p_1(x) \rangle, \langle [m_j, D], p_j(x) \rangle, \langle [D, m_{j+1}], p_{j+1}(x) \rangle, \langle [m_{n-1}, m_n], p_n(x) \rangle \}$, q 的移动方向如图 8(b) 中情形(2)的箭头标识. 当 q 在区间 $[m_i, m_{i+1}] (i < j)$ 内某一点 u 处,

$$p(o_i) = \int_{d(s_1, s_1)}^{d(s_1, m_1)} p_1(x) dx + \sum_{k=1}^{i-1} \int_{d(m_k, m_k)}^{d(m_k, m_{k+1})} p_k(x) dx + \int_{d(m_i, m_i)}^{d(m_i, u)} p_i(x) dx.$$

当 q 在区间 $[m_i, m_{i+1}] (i > j)$ 内某一点 v 处, $p(o_i) = 1 - \int_{d(m_i, m_i)}^{d(m_i, v)} p_i(x) dx + \sum_{k=i+1}^{k=n} \int_{d(m_k, m_k)}^{d(m_k, m_{k+1})} p_k(x) dx.$

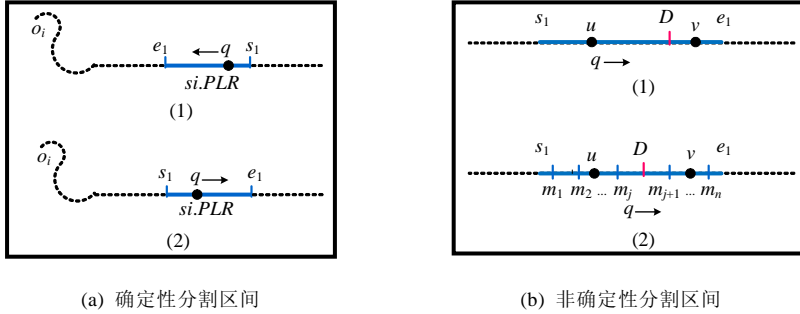


Fig.8 Computation of o_i 's presence probability when q is in the split interval

图 8 q 在分割区间内分布时 o_i 的存在概率计算

3.3 普遍位置不确定性处理

查询点和目标对象的位置都带有不确定信息的情况,我们称为普遍位置不确定性问题.与前两种单独存在的不确定性问题相比,这类问题在对查询结果的增量处理方法方面没有差别,但存在概率的计算较为复杂,要根据查询点的位置区域 $q.PLR$ 和分割区间 $si.PLR$ 的位置关系分情况进行讨论.

假设 $q.PLR$ 和关联目标对象 o_i 的某个 $si.PLR$ 是两条连续的线段 $[s_1, e_1], [s_2, e_2]$ (为了表述方便,我们用最简单的情况来进行说明,与其他情况的原理相同):

情况 1. $q.PLR$ 与 $si.PLR$ 相离

此类情况与第 3.2 节中的情况 1 属于同一类情况,可以采用相同的处理方法.唯一一点不同是用了“完全经过某个分割区间”替换“经过某一个分割点”的概念,不影响判定流程.这里不再进行重复性表述.

情况 2. $q.PLR$ 与 $si.PLR$ 交叠

不失一般性,假设 s_1 和 s_2 分别是 $q.PLR$ 和 $si.PLR$ 的距离 o_i 较近的边缘点, $q.PLR$ 和 $si.PLR$ 上的概率分布函数分别为 $p_1(x)$ 和 $p_2(x)$, 那么交叠的部分只能为 $[s_2, e_1], [s_1, e_2]$ (如图 9(a)所示).

1) 若是 $[s_2, e_1]$, 设点 v 是 $[s_1, e_1]$ 上的一点, 那么当 $v \in [s_1, s_2] = q.PLR - si.PLR$, 则 $p_i = 1$; 当 $v \in [s_2, e_1] = q.PLR \cap si.PLR$, 则 $p_i = \int_{d(s_2, s_2)}^{d(s_2, v)} [1 - p_2(x)] dx$. 那么, o_i 的平均存在概率为

$$\begin{aligned} p &= \int_{d(s_1, s_1)}^{d(s_1, e_1)} p_1(x) \cdot p_i(x) dx \\ &= \int_{d(s_1, s_1)}^{d(s_1, s_2)} p_1(x) \cdot 1 dx + \int_{d(s_1, s_2)}^{d(s_1, e_1)} p_1(x) \cdot \left\{ \int_{d(s_2, s_2)}^{d(s_1, s_2)} [1 - p_2(y)] dy \right\} dx \\ &= \int_{d(s_1, s_1)}^{d(s_1, s_2)} p_1(x) dx + \int_{d(s_1, s_2)}^{d(s_1, e_1)} p_1(x) \cdot \left\{ \int_{d(s_2, s_2)}^{x-d(s_1, s_2)} [1 - p_2(y)] dy \right\} dx. \end{aligned}$$

2) 若是 $[s_1, e_2]$, 设点 v 是 $[s_1, e_1]$ 上的一点, 那么当 $v \in [s_1, e_2] = q.PLR \cap si.PLR$, 则 $p_i = \int_{d(s_1, s_1)}^{d(s_1, v)} p_2(x) dx$; 当 $v \in [e_2, e_1] = si.PLR - q.PLR$, 则 $p_i = 0$. 那么, o_i 在查询结果中的平均存在概率为

$$\begin{aligned} p &= \int_{d(s_1, s_1)}^{d(s_1, e_1)} p_1(x) \cdot p_i(x) dx \\ &= \int_{d(s_1, s_1)}^{d(s_1, e_2)} p_1(x) \cdot \left[\int_{d(s_2, s_1)}^{x+d(s_2, s_1)} p_2(y) dy \right] dx + \int_{d(s_1, e_2)}^{d(s_1, e_1)} p_1(x) \cdot 0 dx \\ &= \int_{d(s_1, s_1)}^{d(s_1, e_2)} p_1(x) \cdot \left[\int_{d(s_2, s_1)}^{x+d(s_2, s_1)} p_2(y) dy \right] dx. \end{aligned}$$

情况 3. $q.PLR$ 包含 $si.PLR$

不失一般性,假设 s_1 和 s_2 分别是 $q.PLR$ 和 $si.PLR$ 的距离 o_i 较近的边缘点, $q.PLR$ 和 $si.PLR$ 上的概率分布函数分别为 $p_1(x)$ 和 $p_2(x)$,那么,交叠的部分只能为 $[s_2, e_2]$ (如图 9(b)中情形(1)所示).

设点 v 是 $[s_1, e_1]$ 上的一点,那么当 $v \in [s_1, s_2]$, 则 $p_i=1$; 当 $v \in [s_2, e_2]$, 则 $p_i = \int_{d(s_2, s_2)}^{d(s_2, v)} [1 - p_2(x)] dx$; 当 $v \in [e_2, e_1]$, 则 $p_i=0$. 那么, o_i 在查询结果中的平均存在概率为

$$\begin{aligned} p &= \int_{d(s_1, s_1)}^{d(s_1, e_1)} p_1(x) \cdot p_i(x) dx \\ &= \int_{d(s_1, s_2)}^{d(s_1, s_2)} p_1(x) \cdot 0 dx + \int_{d(s_1, s_2)}^{d(s_1, e_2)} p_1(x) \cdot \left\{ \int_{d(s_2, s_2)}^{x-d(s_1, s_2)} [1 - p_2(y)] dy \right\} dx + \int_{d(s_1, e_2)}^{d(s_1, e_1)} p_1(x) \cdot 1 dx \\ &= \int_{d(s_1, s_2)}^{d(s_1, e_2)} p_1(x) \cdot \left\{ \int_{d(s_2, s_2)}^{x-d(s_1, s_2)} [1 - p_2(y)] dy \right\} dx + \int_{d(s_1, e_2)}^{d(s_1, e_1)} p_1(x) dx. \end{aligned}$$

情况 4. $q.PLR$ 被 $si.PLR$ 包含

不失一般性,假设 s_1 和 s_2 分别是 $q.PLR$ 和 $si.PLR$ 的距离 o_i 较近的边缘点, $q.PLR$ 和 $si.PLR$ 上的概率分布函数分别为 $p_1(x)$ 和 $p_2(x)$,那么交叠的部分只能为 $[s_1, e_1]$ (如图 9(b)中的情形(2)所示).

设点 v 是 $[s_1, e_1]$ 上的一点,则 $p_i = \int_{d(s_2, s_1)}^{d(s_2, v)} [1 - p_2(x)] dx$. 那么, o_i 在结果中的平均存在概率为

$$p = \int_{d(s_1, s_1)}^{d(s_1, e_1)} p_1(x) \cdot p_i(x) dx = \int_{d(s_1, s_1)}^{d(s_1, e_1)} p_1(x) \cdot \left\{ \int_{d(s_2, s_2)}^{x-d(s_2, s_1)} [1 - p_2(y)] dy \right\} dx.$$

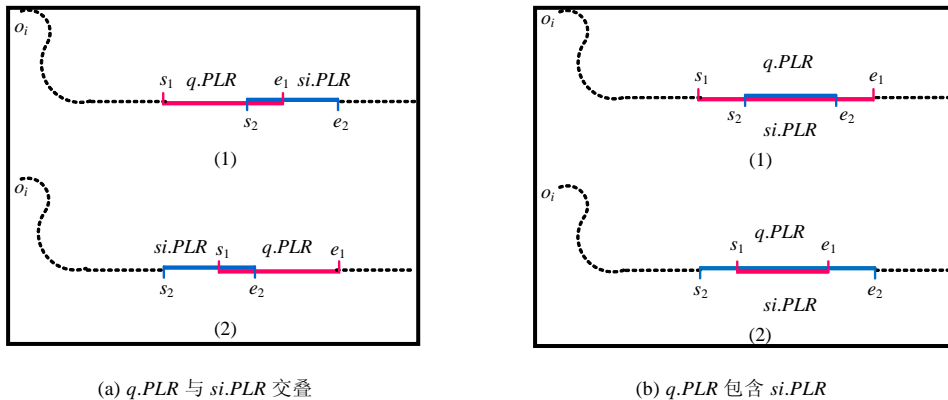


Fig.9 Instances of $q.PLR$ and $si.PLR$'s position relationships

图 9 $q.PLR$ 与 $si.PLR$ 位置关系示例图

4 网络距离范围查询不确定性处理

4.1 网络距离与欧式距离范围查询不确定性处理差异

根据定义 4 和定义 5 可以看出,两类查询除了在查询点与目标对象之间的距离计算方法上存在差异,其他特性均相同.所以,同样可以采用基于有效区间划分的增量处理算法来解决网络距离范围查询的不确定性问题.

在查询点和目标对象位置不确定的处理过程中,查询点与目标对象之间的距离计算方法的不同,决定了对于相同的目标对象集合,两类查询会计算出不同的分割点和分割区间集合.对于 NDMR,分割点的位置意义是以对象的位置为起点,沿着路径所能到达的最远位置.那么分布在一定位置区间内的对象,能够到达的最远位置一定是以各个方向上区间的端点为起点出发的,所以采用网络扩展算法确定分割点和分割区间集合.

普遍位置不确定性问题中,目标对象的位置不确定性导致分割区间没有绝对极值的情况,使 PEMDR 查询与 NDMR 查询的区别更模糊,所以在对查询结果的增量处理的原理和方法方面没有差别,只是在分割区间的计

算细节存在差异;对于计算存在概率的公式也相同,根据查询点的 PLR 和分割区间的 PLR 关系分情况讨论,这里不再重复说明.关于分割区间的分布计算属于积分计算,而存在概率的计算是基于分割区间的概率分布函数的二重积分.这样,无论是对于目标对象分布区间的存储空间还是目标对象在结果中的存在概率,计算效率上都是—种挑战.可采用近似方法来解决目标对象的位置带有不确定性的问题,以较小的精度损失换取效率上的极大提高.

4.2 普遍位置不确定性问题的处理优化

将目标对象 o_i 的 PLR 用一个点来代替,这类点的位置应该尽量在多数情况下反映对象实际的精确位置,所以应该根据 PDF 的取值估计位置.PLR 是由一个或多个连续线段组成,可将其看作是有重量的绳子,PDF 就是相应的质量分布函数,那么其质心点即为满足以上要求的点.PLR 可能含有多个段,两种查询的分布都被限制在路网上,在 PLR 内,任意两点之间的距离等价于最短路径距离,所以都要按路径距离找“质心”,忽略他们之间的夹角.而且根据要求,我们要找的这个点必须在路网上而不是无限制的空间内任何位置处一点,所以这种特殊的“质心”不一定存在,我们可以找到一个最接近的点来近似.根据质心的性质,我们给出概率分布近似中心的定义以及所需的两个基础概念:目标对象 PLR 中某个区间的概率分布中心和概率中心距.

定义 10(概率分布中心(probability distribution center)). 对于某个 $\langle [s_j, e_j], p_j \rangle \in o_i.PLD$,在区间 $[s_j, e_j]$ 内、与 s_j 的距离为 $d = \frac{\int_{d(s_j, s_j)}^{d(s_j, e_j)} x \cdot p_j(x) dx}{\int_{d(s_j, s_j)}^{d(s_j, e_j)} p_j(x) dx}$ 的点即为 o_i 在 $[s_j, e_j]$ 段的概率分布中心,简称为 PDC.

定义 11(概率中心矩(probability moments)). 对某个 $\langle [s_j, e_j], p_j(x) \rangle \in o_i.PLD$,假设它的 PDC 为点 c_j 相对于路网中任意一点 v 的概率中心距为 $PM(o_i.PLR[j], v) = d(c_j, v) \cdot \int_{d(s_j, s_j)}^{d(s_j, e_j)} p_j(x) dx$, 简称为 PM.

定义 12(概率分布近似中心(approximative center of probability distribution)). 对于 $o_i.PLD$ 中某一个点 c ,能够使各段相对于点 c 的 PM 值之间的差距尽可能小,那么,这个点 c 称为 o_i 的概率分布近似中心,简称为 PDAC.

如图 10 所示:

- 对于情况(1),设 $o_1.PLD = \{ \langle [s_1, e_1], p_1(x) \rangle \}$,只有一个段,那么, o_1 在区间 $[s_1, e_1]$ 上的 PDC 就是 o_1 的 PDAC,即为区间 $[s_1, e_1]$ 上的与 s_1 的距离为 $d = \frac{\int_{d(s_1, s_1)}^{d(s_1, e_1)} x \cdot p_1(x) dx}{\int_{d(s_1, s_1)}^{d(s_1, e_1)} p_1(x) dx}$ 的点;
- 对于情况(2),设 $o_2.PLD = \{ \langle [s_2, m_2], p_1(x) \rangle, \langle [m_2, e_2], p_2(x) \rangle \}$,有两个段,但是可以看做是一个整体,那么 o_2 的 PDAC 为区间 $[s_2, e_2]$ 上的与 s_2 的距离为 $d = \frac{\int_{d(s_1, s_1)}^{d(s_1, m_1)} x \cdot p_1(x) dx + \int_{d(m_1, m_1)}^{d(m_1, e_1)} (x + |o_2.PLR[1]|) \cdot p_2(x) dx}{\int_{d(s_1, s_1)}^{d(s_1, m_1)} p_1(x) dx + \int_{d(m_1, m_1)}^{d(m_1, e_1)} p_2(x) dx}$ 的点;
- 情况(3)和情况(4)的 PDAC 不能简单地按照质心的坐标公式求得

$$O_i.PLD = \{ \langle [s_2, m], p_1(x) \rangle, \dots, \langle [s_n, m], p_n(x) \rangle, \langle [m, e], p_{n+1}(x) \rangle \} (n > 1).$$

根据情况(1)的计算公式,可以求出 $o_i.PLR$ 中各段上 PDC 的集合 $CS = \{ c_1, c_2, \dots, c_{n+1} \}$,与 m 的距离分别为 d_1, d_2, \dots, d_{n+1} ,所占概率比为 p_1, p_2, \dots, p_{n+1} ,则 $PM(o_i.PLR[j], m) = d_j \cdot p_j$. 设 $\max \{ d_j \cdot p_j | 1 \leq j \leq n+1 \} = d_k \cdot p_k$ (若 k 值多于一个且概率比相等,则计算其中一个即可;否则,取概率比大者进行计算),依据 PDAC 的定义, o_i 的 PDAC 点 c 应该在 $o_i.PLR[k]$ 上,设 c 与 m 的距离为 $x(x > 0)$ 的点,应该满足:除第 k 段以外的其他段中找到最小 x 值的条件,最先达到其概率中心距 $PM(o_i.PLR[j], c) (j \neq k)$ 与 $PM(o_i.PLR[k], c)$ 相等,即

$$x = \min \{ \min(x_j) | m_j(d_j + x_j) \leq m_k(d_k - x_j), 1 \leq j \leq n+1 \wedge j \neq k \}.$$

- 若求出的 $x \leq |o_i.PLR[k]|$,则结果就是 $o_i.PLR[k]$ 上与 m 的距离为 x 的 c 点,若求出的 x 值(相等同为最小)多于 1 个,则都作为结果保留;

- 若 $x > |o_i.PLR[k]|$, 则 c 点的位置与 $o_i.PLR[k]$ 的边缘点重合.

如图 10 所示:

- 对于情况 (3), 设 $o_3.PLD = \{ \langle [s_{13}, m_3], 0.1 \rangle, \langle [s_{23}, m_3], 0.2 \rangle, \langle [s_{33}, m_3], 0.3 \rangle, \langle [m_3, e_3], 0.4 \rangle \}$, $|o_3.PLR[1]| = 2$,

$|o_3.PLR[2]| = 2, |o_3.PLR[4]| = 2, |o_3.PLR[4]| = 4$. 那么,

$$\max \{ d_j p_j | 1 \leq j \leq 4 \} = \max \{ 0.1 \times 1, 0.2 \times 1, 0.3 \times 1, 0.4 \times 2 \} = \max \{ 0.1, 0.2, 0.3, 0.8 \} = 0.8 = d_4 \cdot p_4$$

$$x = \min \{ 1.4, 1, 0.714 \} = 0.714,$$

则 o_3 的 PDAC 是区间 $[m_3, e_3]$ 内与 m_3 距离为 0.714 的点.

- 对于情况 (4), 设 $o_3.PLD = \{ \langle [s_4, m_4], 0.2 \rangle, \langle [m_4, e_{14}], 0.4 \rangle, \langle [s_{33}, m_3], 0.3 \rangle, \langle [m_4, e_{24}], 0.4 \rangle \}$, $|o_3.PLR[1]| = 2$,

$|o_4.PLR[2]| = |o_4.PLR[3]| = 4$. 那么, $\max \{ d_j p_j | 1 \leq j \leq 3 \} = \max \{ 0.2, 0.8, 0.8 \} = 0.8 = d_2 \cdot p_2 = d_3 \cdot p_3$, $x = \min \{ 1 \} = 1$, 则

o_4 的 PDAC 有两个, 是区间 $[m_4, e_{14}]$ 和 $[m_4, e_{24}]$ 内的与 m_4 的距离为 1 的点.

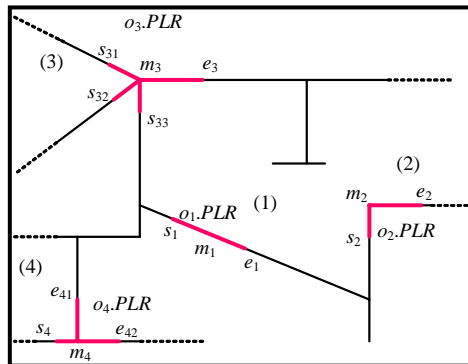


Fig.10 Examples of locating PDAC

图 10 PDAC 的定位示例

5 实验

在这节中,我们用实验结果分析本文提出的模型和方法的性能及其有效性.实验的所有算法使用 C++ 语言实现,执行环境是 Intel Core 2 Duo 2.00 GHz 处理器,2.0GB 内存,120GB SATA 硬盘,Windows XP 操作系统.

5.1 实验设置

实验数据集采用真实路网和模拟的目标对象分布合成的数据集.其中,真实的路网数据集使用的是德国奥尔登堡市(OL)(包括 7 034 条边、6 104 路网节点)和美国加利福尼亚州的圣华金市(TG)(包括 23 873 条边、18 262 路网节点)两个城市的路网拓扑数据,均可以在 <http://www.maproom.psu.edu/dew> 网站上下载到.目标对象在路网上的位置是根据均匀分布和齐夫分布模拟出来的(如图 11(a)、图 11(b)所示).我们将路网和目标对象的分布映射到 10000×10000 的二维空间范围中,并用 quadtree 进行索引(如图 11(c)、图 11(d)所示).我们假定查询点在路网内做匀速运动(单位时间移动 5 个单位),每次实验中的出发点和路径的选择都是随机的(图 11 中,连续线段表示查询点的路线).不失一般性,所有参数的值都是根据物理世界的实际情况设置的.实验参数的基本设置总结见表 2.在考察每个变量对结果的影响时,其他参数都设置为默认值.

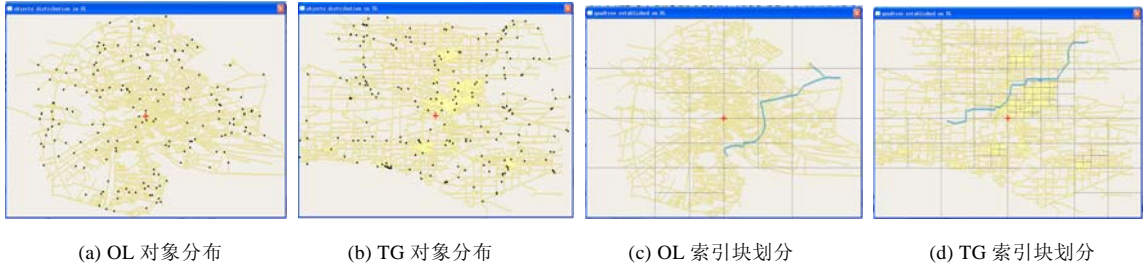


Fig.11 Quadtree index and object distribution in real datasets for road networks

图 11 真实路网 quadtree 索引块划分和模拟的对象分布

Table 2 Default parameter settings for the experiments

表 2 实验参数设置

变量	取值范围	默认值	意义
D_{object}	Uniform, zipf	Uniform	目标对象分布
D_{edge}	OL, TG	OL	路网分布
$R(\times 1000)$	0.5, 1, 1.5, 2, 2.5	1	查询半径
N_{object}	100~10000	200	目标对象数量
T	5, 10, 15, 20, 25, 30	20	查询周期
$L(\times 1000)$	5, 10, 15, 20, 25, 30	20	查询点路径长度
$Ambig_d(\%)$	0.1, 0.2, 0.3, 0.4, 0.5	0.3	模糊度(表示不确定位置的 PLR 与整个二维空间的边长之比)
θ	0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9	0.7	结果概率阈值

5.2 实验结果分析

每组实验得到的结果数据都是各种算法在相同实验设置下连续执行 100 次查询得到的平均值,即用每个查询结果的平均响应时间来评估查询效率,用结果的平均准确率来评估查询准确性.准确率包括查全率(recall)和查准率(precision).设根据模糊位置算得的查询结果为 $ambR$,根据精确位置算得的查询结果为 $accR$,则查全率为 $ambR$ 中包含的准确结果对象个数占 $accR$ 结果对象个数的比率 $\frac{|ambR \cap accR|}{|accR|}$,查准率为 $ambR$ 中包含的准确结果对象个数占 $ambR$ 结果对象个数的比率 $\frac{|ambR \cap accR|}{|ambR|}$.

由于没有相关工作研究此类问题,我们仅对比扩展后的算法在处理查询点位置不确定性、目标对象位置不确定性、普遍位置不确定性问题的平均响应时间和查询结果的准确率.分别用缩写符号 QPLU,QOLU,ULU 表示 EDMR 和 NDMR 这 3 类不确定性问题.每类实验中的准确性用查全率和查准率来评估.

(1) R 的影响

图 12 显示了当 R 在区间 $[0.5 \times 10^3, 2.5 \times 10^3]$ 内变化时,扩展算法在处理 QOLU 时平均响应时间和查全率的变化情况.从图 12(a)中可以看出,算法的平均响应时间会随着 R 的增加而增长,增长速度比较缓慢.这是因为扩展的方法采用分割区间代替分割点来构造有效区间,分割区间(PDF)的计算复杂度要高于分割点的计算,因 R 的增加而包含的新候选目标对象的分割区间的计算是影响平均响应时间的主要因素.从图 12(b)中可以看出,扩展算法的查全率会随着 R 的增加而提高.这是因为处于查询范围边缘的对象才需要用积分计算其存在概率,而处于查询范围较为中心位置的对象的存在概率一定为 1,不需要复杂计算. R 增加使查询范围变大,那么在其他参数值不变的情况下,存在概率一定为 1 的目标对象占结果的比例变大,因此查全率会提高.当 R 在区间 $[0.5 \times 10^3, 2.5 \times 10^3]$ 内变化时,处理 NDMR 和 EDMR 的 QOLU 时准确率都能达到 85%以上.

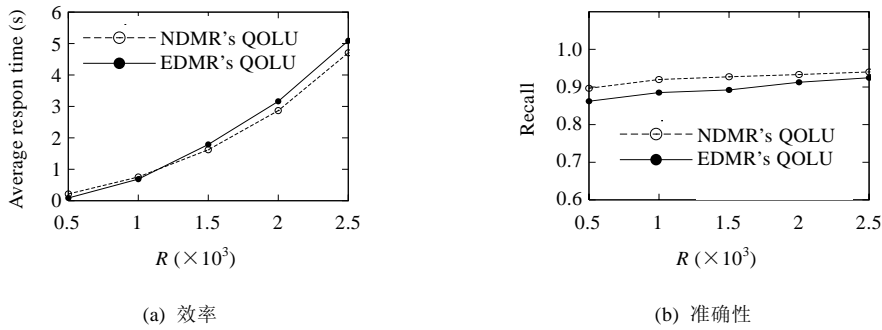


Fig.12 Comparison of efficiency with different length of query radius

图 12 不同查询半径下处理 QOLU 的效率和准确性比较

(2) *Ambig_d* 的影响

图 13 反映出当 *Ambig_d* 在区间[0.1,0.5]内变化时,扩展算法处理 3 类不确定性时的平均响应时间和查全率的变化情况.从图 13(a)中可以看出,扩展算法处理 ULU 的平均响应时间最长,QOLU 的处理要比处理 QPLU 更耗时.这是因为扩展的方法处理 QPLU 时仍然采用分割点来构造对象的有效区间,而处理 QOLU 和 ULU 时采用分割区间代替分割点来构造有效区间,分割区间(PDF)的计算复杂度要高于分割点的计算.从图 13(b)中可以看出,扩展算法的查全率会随着 *Ambig_d* 的增加而降低.这是因为处于查询范围边缘的对象才需要用积分计算其存在概率,而处于查询范围较为中心位置的对象的存在概率一定为 1,不需要复杂计算.*Ambig_d* 增加使概率变大,那么在其他参数值不变的情况下,存在概率一定为 1 的目标对象占结果的比例变小,因此查全率会下降.

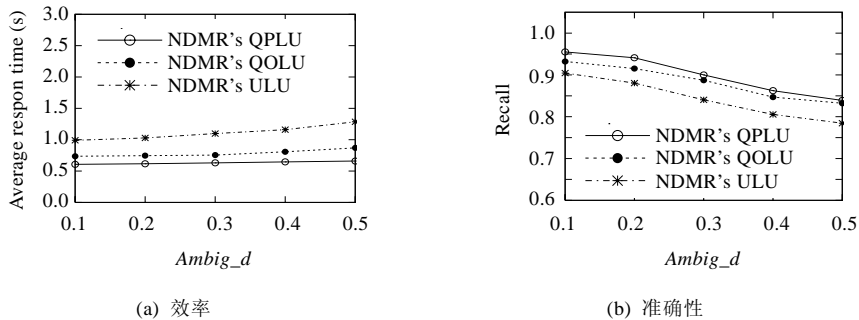


Fig.13 Comparison of efficiency and accuracy with different ambiguity degree

图 13 不同模糊度下的效率和准确性比较

(3) θ 的影响

图 14 反映了当 θ 在区间[0.3,0.9]内变化时,扩展算法在处理 3 类不确定性问题时的平均响应时间和查全率的变化情况.从图 14(a)中可以看出,扩展算法的平均响应时间不受 θ 的变化影响.这是因为在搜索分割点时不考虑 θ 限制, θ 条件的过滤是在将所有候选结果对象的存在概率计算之后进行过滤的.从图 15(b)中可以看出,扩展算法的查全率会随着 θ 的增加下降一些.这是因为计算得出的一些真阳性结果的存在概率会小一些,随着 θ 的增加,会过滤掉一些真阳性结果.当 θ 取 0.7 左右时,查全率接近极大值.

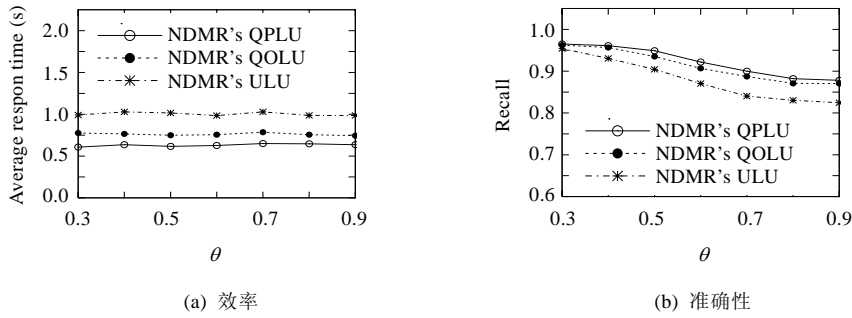


Fig.14 Comparison of efficiency and accuracy with different result probability threshold

图 14 不同结果概率阈值下的效率和准确性比较

(4) 目标对象分布的影响

图 15 显示了当 N_{object} 在区间 $[0.5 \times 10^3, 2.5 \times 10^3]$ 内变化时,扩展算法在处理 QOLU 的平均响应时间和查准率的变化情况.从图 15(a)和图 15(b)中可以看出,扩展算法平均响应时间受查询目标对象分布的影响较小,只显示出较小的波动差别,但是不会影响到总体的对比关系.这是因为齐夫分布中不同位置的查询目标数量差异较大,而每次选择的路线出发点和走向都是随机的,导致响应时间会有波动, N_{object} 越大,波动得越强烈,所以均匀分布的变化曲线较为平滑一些.查准率会随着 N_{object} 的增加而体现出微小幅度的提高,这是因为处于查询范围边缘的对象才需要用积分计算其存在概率,而处于查询范围较为中心位置的对象的存在概率一定为 1,不需要复杂计算. N_{object} 的增加,使查询范围存在概率一定为 1 的目标对象增量较多,因此查准率会提高.当 R 在区间 $[0.5 \times 10^3, 2.5 \times 10^3]$ 内变化时,处理两种分布下 EDMR 的 QOLU 时,准确率都能达到 85% 以上.

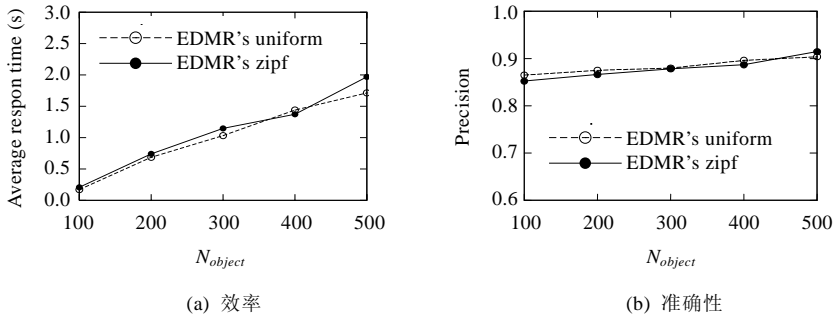


Fig.15 Comparison of efficiency and accuracy with different object distribution

图 15 不同对象分布下的效率和准确性比较

(5) 路网分布的影响

图 16 显示了当 N_{object} 在区间 $[0.5 \times 10^3, 2.5 \times 10^3]$ 内变化时,扩展算法在处理 QOLU 平均响应时间和查准率的变化情况.从图 16(a)中可以看出,在其他参数值相同的情况下,处理 TG 图的平均响应时间要高于处理 OL 图的平均响应时间.由于将两个城市的路网都映射到相同大小的二维空间中,TG 图包含的边和顶点的数量分别大约是 OL 图的 3 倍,所以密度大.查询范围相同的情况下,边和顶点分布密度大的图,其 I/O 读数量会较大,并且在搜索和处理分割区间时的工作量也较多.从图 16(b)中可以看出,扩展算法的查准率几乎不受路网分布的影响.但由于多次重复性实验为查询点随机选择的不同路线的出发点和走向,而且在 OL 和 TG 中,不同位置的路的长度和数量差异较大,导致响应时间会有波动.从图 11 中可以看出,OL 的分布比 TG 的分布更均匀一些,所以 OL 的曲线较为平滑一些.

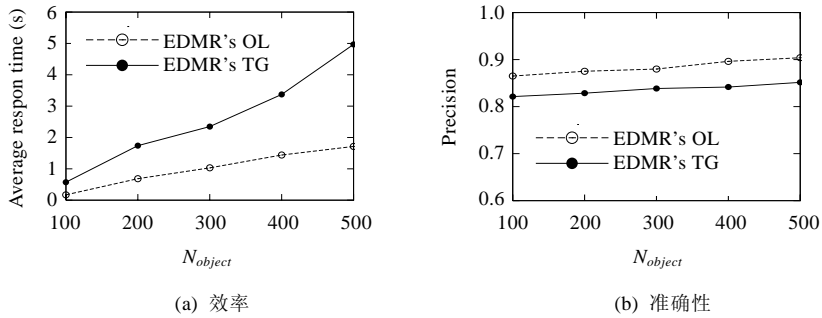


Fig.16 Comparison of efficiency and accuracy with different road network distribution

图 16 不同路网分布下的效率和准确性比较

(6) 目标对象数目的影响

图 17 反映了当目标对象较为密集时,QOLU 平均响应时间和查准率的变化情况.从图 17(a)中可以看出,算法的平均响应时间会随着 N_{object} 的增加而增长.这是因为分割区间是根据目标对象的位置计算得出的,目标对象越多,分割区间的集合越庞大,每个查询周期需要考虑的增量维护的工作量就会增加,但是没有因查询半径变大所增长的速度快.由于我们的查询是针对满足某种条件的静止目标对象进行连续范围查询,例如周期性查询离目标对象最近的加油站或者宾馆等等,在一般的应用场景中,这类路网地图上的范围很大,即使相关目标对象较多,密度也相对较小.从图 17(b)中可以看出,查准率会随着 N_{object} 的增加而体现出微小幅度的提高,总体趋于稳定.这是因为处于查询范围边缘的对象才需要用积分计算其存在概率,而处于查询范围较为中心位置的对象的存在概率一定为 1,不需要复杂计算. N_{object} 增加使查询范围存在概率一定为 1 的目标对象增量较多,因此查准率会提高.图 17(c)显示了结果中目标对象的数量及其变化.显然,对象结果的数目跟目标对象密度、查询范围和结果概率阈值有关.

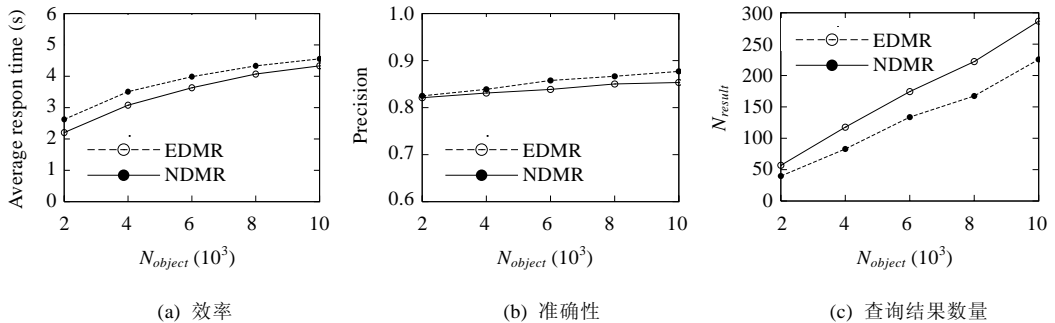


Fig.17 Comparison of efficiency、accuracy and query result number with maissive object number

图 17 大规模目标对象环境下的效率、准确性和查询结果数量比较

(7) 采用 PDAC 近似处理算法的性能评估

将未用 PDAC 的算法和采用 PDAC 近似处理算法分别简单记为基本扩展算法(BEA)和 PDAC 近似算法(PDAC).图 18 显示了当 R 在区间 $[0.5 \times 10^3, 2.5 \times 10^3]$ 内变化时,采用 PDAC 近似处理算法的平均响应时间和查准率的变化情况.

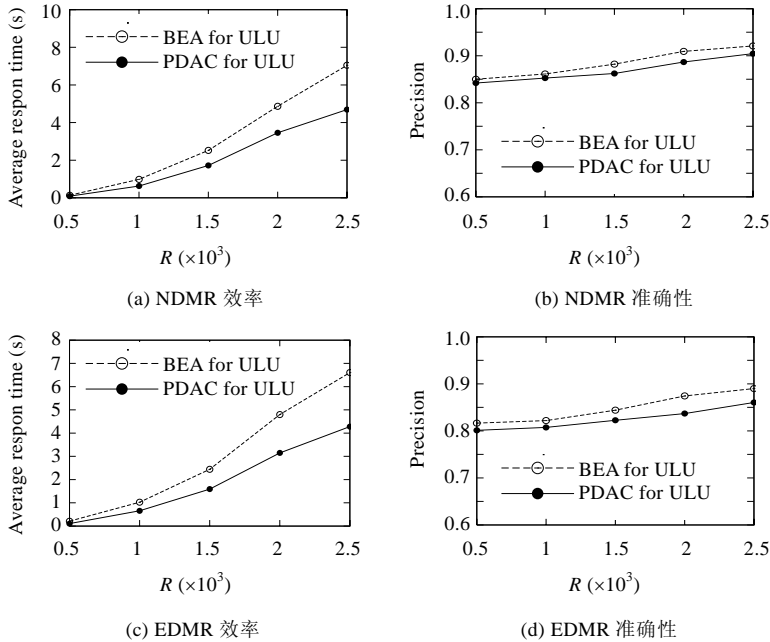


Fig.18 Comparison of efficiency and accuracy between the methods with and without PDAC for ULU

图 18 不同数量对象下是否采用 PDAC 处理 ULU 的效率和准确性比较

从图 18(a)、图 18(c)中可以看出,该类算法的平均响应时间都有所降低,而处理 EDMR 查询的平均响应时间降低的幅度最大.这是因为采用 PDAC 近似对象的位置来处理 ULU,可以将定位和计算分割区间(PDF)的工作简化为搜索分割点的过程,EDMR 查询的分割区间计算处理更复杂.从图 18(b)、图 18(d)中可以看出,采用 PDAC 近似处理算法会导致查准率有所下降,但是幅度较小.这是因为 PDAC 是根据目标对象的可能位置分布情况计算出来的,在一定程度上反映了目标对象的真实位置.

综上,算法在分别处理 EDMR 和 NDMR 两种概率范围查询的 3 类不确定性问题时,处理普遍位置不确定性的平均响应时间最长,目标对象位置不确定性的处理要比处理查询点位置不确定性更耗时.当结果阈值 θ 取 0.7 左右时,结果的准确率已开始趋于稳定,均能达到 80% 以上.考察了在解决普遍位置不确定性问题时,采用概率分布近似中心(PDAC)作为目标对象的估计位置方法,在保证了一定的准确性标准的同时,其效率有了明显的提升.通过牺牲较少的准确性换来处理速度的极大提高,这种方法适用于那些对实时性要求较高的应用.

6 结论

本文首先采用最具一般性的一维 PDF 方法表示路网中的不确定位置信息.然后用分割区间将查询点的可能路径划分成支持增量处理的有效区间,针对 EDMR,NDMR 这种移动范围概率查询.根据 PLR 和分割区间的分布特性,分情况讨论目标对象存在概率的计算方法和查询结果的增量维护算法.最后,考虑采用概率分布近似中心(PDAC)作为目标对象的估计位置,近似地解决普遍位置不确定性的问题,以较小的精度损失换取效率上的极大提高.实验结果证明,本文提出的模型和算法具有良好的性能.

References:

- [1] Kolahdouzan M, Shahabi C. Voronoi-Based K nearest neighbor search for spatial network databases. In: Nascimento MA, *et al.*, eds. Proc. of the 30th Int'l Conf. on Very Large Data Bases. Toronto: Morgan Kaufmann Publishers, 2004. 840–851.
- [2] Cho HJ, Chung CW. An efficient and scalable approach to CNN queries in a road network. In: Bohm K, *et al.*, eds. Proc. of the 31st Int'l Conf. on Very Large Data Bases. Trondheim: ACM Press, 2005. 865–876.

- [3] Hu HB, Lee DL, Xu JL. Fast nearest neighbor search on road networks. In: Ioannidis YE, *et al.*, eds. Proc. of the 10th Int'l Conf. on Extending Database Technology. LNCS 2992, Munich: Springer-Verlag, 2006. 186–203. [doi: 10.1007/11687238_14]
- [4] Liu FY, Do TT, Hua KA. Dynamic range query in spatial network environments. In: Bressan S *et al.*, eds. Proc. of 17th Int'l Conf. of Database and Expert Systems Applications. LNCS 4080, Krakow: Springer-Verlag, 2006. 254–265. [doi: 10.1007/11827405_25]
- [5] Chung BSE, Lee WC, Chen ALP. Processing probabilistic spatio-temporal range queries over moving objects with uncertainty. In: Kersten ML, *et al.*, eds. Proc. of the 12th Int'l Conf. on Extending Database Technology. Saint-Petersburg: ACM Press, 2009. 60–71. [doi: 10.1145/1516360.1516369]
- [6] Kriegel HP, Kunath P, Renz M. Probabilistic nearest-neighbor query on uncertain objects. In: Kotagiri Ramamohanarao K, *et al.*, eds. Proc. of 12th Int'l Conf. on Database Systems for Advanced Applications. LNCS 443, Bangkok: Springer-Verlag, 2007. 337–348. [doi: 10.1007/978-3-540-71703-4_30]
- [7] Trajcevski G, Tamassia R, Ding H, Scheuermann P, Cruz IF. Continuous probabilistic nearest-neighbor queries for uncertain trajectories. In: Kersten ML, *et al.*, eds. Proc. of the 12th Int'l Conf. on Extending Database Technology. Saint-Petersburg: ACM Press, 2009. 874–885. [doi: 10.1145/1516360.1516460]
- [8] Cheng R, Chen JC, Mokbel M, Chow CY. Probabilistic verifiers: Evaluating constrained nearest-neighbor queries over uncertain data. In: Alonso G, *et al.*, eds. Proc. of the 24th Int'l Conf. on Data Engineering. Cancun: IEEE Computer Society, 2008. 973–982. [doi: 10.1109/ICDE.2008.4497506]
- [9] Cheng R, Chen L, Chen JC, Xie XK. Evaluating probability threshold k -nearest-neighbor queries over uncertain data. In: Kersten ML, *et al.*, eds. Proc. of the 12th Int'l Conf. on Extending Database Technology. Saint-Petersburg: ACM Press, 2009. 672–683. [doi: 10.1145/1516360.1516438]
- [10] Huang YK, Chen CC, Lee C. Continuous K -nearest neighbor query for moving objects with uncertain velocity. *Geoinformatica*, 2009,13(1):1–25. [doi: 10.1007/s10707-007-0041-0]
- [11] Lian X, Chen L. Efficient processing of probabilistic reverse nearest neighbor queries over uncertain data. *VLDB Journal*, 2009, 18(3):787–808. [doi: 10.1007/s00778-008-0123-0]
- [12] Cheema MA, Lin XM, Wang W, Zhang WJ, Pei J. Probabilistic reverse nearest neighbor queries on uncertain data. *IEEE Trans. on Knowledge and Data Engineering*, 2010,22(4):550–564. [doi: 10.1109/TKDE.2009.108]
- [13] Trajcevski G, Wolfson O, Hinrichs K, Chamberlain S. Managing uncertainty in moving objects databases. *ACM Trans. on Database Systems*, 2004,29(3):463–507. [doi: 10.1145/1016028.1016030]
- [14] Zheng K, Trajcevski G, Zhou XF, Scheuermann P. Probabilistic range queries for uncertain trajectories on road networks. In: Ailamaki A, *et al.*, eds. Proc. of the 14th Int'l Conf. on Extending Database Technology. Uppsala: ACM Press, 2011. 283–294. [doi: 10.1145/1951365.1951400]
- [15] Hua M, Pei J. Probabilistic path queries in road networks: Traffic uncertainty aware path selection. In: Manolescu I, *et al.*, eds. Proc. of the 13th Int'l Conf. on Extending Database Technology. Lausanne: ACM Press, 2010. 347–358. [doi: 10.1145/1739041.1739084]



谷峪(1981—),男,辽宁鞍山人,博士,副教授,CCF 会员,主要研究领域为空间数据管理,图数据管理.

E-mail: guyu@ise.neu.edu.cn



于戈(1962—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为数据库理论与技术.

E-mail: yuge@ise.neu.edu.cn



郭娜(1985—),女,硕士,主要研究领域为空间数据管理.

E-mail: guona_neu@163.com