

# 基于位置范围的道路网 skyline 查询

施常月 秦小麟 许建秋 胡彩平

(南京航空航天大学计算机科学与技术学院 南京 210016)

**摘要** 随着无线通信和定位技术的发展,道路网 skyline 查询在基于位置的服务等方面越来越重要。考虑到现今道路网中位置隐私保护和定位设备的精度问题,用户在道路网上的位置通常用一个范围来表示。但是,已有的道路网 skyline 研究都是基于单一查询点。针对这一问题,研究了一种新的查询——基于位置范围的道路网 skyline 查询(RNS),提出了一种基于边界点替换的有效查询处理算法。另外,针对已有的道路网 skyline 查询中复杂的道路网距离计算对查询效率的影响问题,通过计算兴趣点在道路网上的有效 skyline 路段,将其与道路网信息融合,建立了道路网 skyline 模型。基于该模型设计了一种能有效支持 RNS 查询的道路网 skyline 索引 SSR-tree,提出了基于索引的 RNS 查询处理算法。通过大量实验验证了所提方法的有效性,并比较了基于索引的算法在查询效率和精度上的提高。

**关键词** 道路网, Skyline 查询, 位置范围, 查询处理

**中图分类号** TP311.13 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.09.036

## Location Range-based Skyline Query in Road Networks

SHI Chang-yue QIN Xiao-lin XU Jian-qi HU Cai-ping

(College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

**Abstract** With the development of wireless communication and positioning technology, skyline query in road networks has recently been important in LBS. For the consideration of human privacy and the limited accuracy of positioning devices, user's location is always represented as a spatial range. However, the existing researches focus on the point-based skyline query. This paper studied a new problem of the location ranges-based skyline query in road networks(RNS), and proposed an efficient query processing algorithm. In addition, because the complex distance calculation brings decrease of the query efficiency, this paper also proposed an index-based query algorithm, and then by pre-computing effective skyline road segments of POI and creating a new road network model, designed an index to support the RNS efficiently. Extensive experiments on real networks datasets verify the performance and accuracy of the proposed algorithms.

**Keywords** Road networks, Skyline query, Location ranges, Query processing

## 1 引言

skyline 查询<sup>[1]</sup>最早由 Borzanyi 等人提出,是一种典型的多目标决策问题,用于返回兴趣集中所有不被其他点支配的点的集合。给定两个兴趣对象点  $p_1, p_2$ , 其分别具有  $n$  维属性  $(a_1, a_2, \dots, a_n)$  和  $(b_1, b_2, \dots, b_n)$ 。那么,  $p_1$  支配  $p_2$  当且仅当  $p_1$  在任意维属性上都不差于  $p_2$ , 且至少在某一维上优于  $p_2$ 。传统的 skyline 查询的多维属性都是价格等非空间属性,而在实际中,很多需求都是基于用户位置等空间属性。尤其是随着无线通信技术的提升,道路网上基于位置的服务<sup>[2]</sup>也发展起来,空间属性的考虑必不可少。例如,用户查找距离当前位置近且便宜的宾馆、停车位等。已有的空间 skyline 查询以基于欧氏空间为主,而道路网 skyline 查询是一种受限空间

的 skyline 查询。相比欧氏空间的距离计算,道路距离要考虑道路的连通性而不仅是两个数据点的位置,因此原有的研究中基于欧氏环境的查询处理方法不能精确适用于道路网环境。

在移动对象数据库中,服务器通过位置更新来不断确定用户的位置。但定位设备的精度和时间延迟等问题往往会带来位置定位的不确定性<sup>[7]</sup>,只能把用户位置确定在某一个范围内。另外,在现实生活中,出于对用户的隐私保护,通常也把用户的位置模糊成一个区域范围。因此本文研究基于用户位置范围的道路网 skyline 查询。

本文研究基于位置范围的道路网 skyline(RNS)查询,提出了一种基本 RNS 查询处理算法,其能够有效处理 RNS 查询;另外,为了解决复杂的道路网距离计算对查询效率的影

到稿日期:2013-11-22 返修日期:2014-03-04 本文受国家自然科学基金项目(61373015,61300052,41301407),国家教育部高等学校博士学科点博导基金资助项目(20103218110017),江苏高校优势学科建设工程资助项目(PAPD),中央高校基本科研业务费专项项目(NP2013307)资助。

施常月(1990-),女,硕士,主要研究方向为移动对象建模与查询处理, E-mail: scynuaa@126.com; 秦小麟(1953-),男,教授,博士生导师,主要研究方向为分布式环境的数据管理与安全(通信作者); 许建秋(1982-),男,副教授,主要研究方向为空间数据库、移动对象数据库; 胡彩平(1977-),男,副教授,硕士生导师,主要研究方向为空间数据挖掘。

响,设计了一种道路网 skyline 索引 SSR-tree,并提出基于索引的 RNS 查询算法,有效提高了 RNS 的查询效率和查询精度。最后采用真实的城市道路网数据进行大量的实验,对所提方法的有效性和高效性进行验证。

## 2 相关工作

### 2.1 skyline 查询

skyline 查询最早由 Borzanyi 等人在数据库系统中提出,并提出了 BNL 和 D&C 两种基本算法<sup>[1]</sup>,之后又有很多学者致力于 skyline 查询的研究,并相继提出许多改进的高效算法,如:Bitmap<sup>[3]</sup>、NN<sup>[4]</sup>、BBS<sup>[5,6]</sup>等。

近年来随着全球定位技术与无线通信技术的发展,基于位置服务开始兴起,学者们开始将 skyline 查询扩展到更广的场景下进行研究。Sharifzadeh 等人首先提出了空间 skyline<sup>[8]</sup>的概念,随后 Deng 等人又将空间 skyline 扩展到了道路网上,提出了道路网上的多源 skyline 查询(MSQ)<sup>[10]</sup>,将兴趣点到多个查询点之间的距离作为兴趣点的多维属性,并提出了 3 种算法:协同扩张算法(Collaborative Expansion, CE),直接以每个查询点为中心递增地扩大查询范围来筛选候选集;欧氏距离约束算法(Euclidean Distance Constraint, EDC)以及下界约束算法(Lower Bound Constraint, LBC),利用欧氏距离作为网络距离的下界来进行数据集剪枝。Huang 等人针对道路网的连续 skyline 查询的问题,提出了两种 skyline 查询的变种:CD-SQ 和 CKnn-SQ<sup>[11]</sup>,排除用户可能不太感兴趣的距离较远的 skyline 点,比较符合用户需求。

### 2.2 基于范围的查询

由于隐私保护变得越来越重要,移动用户端将自己的位置模糊成一个区域提交服务器,在返回的结果中根据自己的准确位置来选择所需的结果。目前,已经有一些基于范围查询的研究,如 Ku, Xu 等人已经研究了基于范围的 KNN 查询<sup>[13,14]</sup>,并提出了相应的查询处理算法 RkNN。

Lin 等人最早开始研究基于范围的 skyline 查询,并提出了两种算法:I-SKY 和 N-SKY<sup>[9]</sup>。I-SKY 是一种基于索引的方法;N-SKY 则是将基于范围的 skyline 查询分解为基于边界线段的 skyline 查询。但是,上述算法都是基于欧氏空间,用欧氏距离来代替道路网络距离,这样会带来查询结果的误差。因此,本文针对真实的道路网环境,首次提出基于位置范围的道路网 skyline 查询,并研究相应的查询处理算法。

## 3 问题描述

基于范围的道路网 skyline 查询是 skyline 查询在道路网环境下的变种,该问题涉及 3 个组成对象:兴趣点集、查询范围以及道路网。本文假定兴趣点集是一组具有  $n$  维属性的点的集合,其中前  $n-1$  维为静态的非空间属性,如宾馆价格、星级等,第  $n$  维属性为兴趣点距离查询位置的道路距离。首先给出基于位置范围的道路网 skyline 查询的定义:

**定义 1(非空间支配)** 给定两个兴趣点  $p_1, p_2$ , 考虑所有的非空间属性,如果  $p_1$  能被  $p_2$  支配,称  $p_2$  非空间支配  $p_1$ , 记为  $p_2 \prec_{ns} p_1$ 。所有非空间支配  $p_1$  点的集合称为  $p_1$  的非空间支配集  $NDom(p_1)$ 。

**定义 2(道路网支配)** 给定一个查询点  $q$  和兴趣点  $p_1, p_2, d(p_2, q)$  表示兴趣点  $p_2$  距离查询点  $q$  的距离。当且仅当满足以下两个条件时,称  $p_2$  道路网支配  $p_1$  (下文简称“支配”),记为  $p_2 \prec_q p_1$ :

- 1)  $p_2 \prec_{ns} p_1$
- 2)  $d(p_2, q) < d(p_1, q)$ 。

**定义 3(基于点的道路网 skyline(PNS)查询)** 给定一个查询点  $q$  和一个兴趣点集 POI, 其 skyline 查询结果返回 POI 的一个子集,其中的每个对象点都不能被 POI 中的任何其他点所支配,记为  $PNS(q, POI)$ 。

**定义 4(基于范围的道路网 skyline(RNS)查询)** 给定一个查询范围  $R$  (假定  $R$  为一个矩形范围) 和一个兴趣点集 POI, RNS 查询返回范围  $R$  中所有道路网查询点的 PNS 结果的并集,记为  $RNS(R, POI)$ 。

表 1 所列是一个宾馆兴趣点分布在道路网上的查询实例,  $p_1, p_2, p_3, p_4$  是 4 个具有 4 维属性的兴趣点,其中前 3 项非空间属性如表 1 所列。图 1 给出了它们在道路网上的位置分布以及查询范围  $R$ , 其道路网 skyline 查询查找既便宜、星级高、评价好又距离用户当前位置近的宾馆。对于范围  $R$  中不同的查询点,其 skyline 结果不同,因此, RNS 查询需要返回  $R$  中所有查询点的 PNS 结果的并集。

表 1 兴趣点集的非空间属性

宾馆	价格	星级	环境评价
$p_1$	195	3	8.7
$p_2$	220	2	8.7
$p_3$	190	4	9.5
$p_4$	175	4	9.2

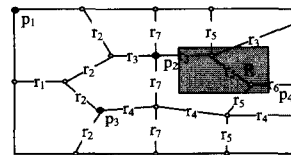


图 1 宾馆兴趣点集在道路网上的分布及查询范围

## 4 基本 RNS 查询处理算法 B-RNS

假定兴趣点前  $n-1$  维的非空间属性是固定的,那么只考虑非空间属性,其 skyline 结果也是固定的。由于兴趣点为道路网上的静态点,故其道路网 skyline 查询结果的变化是由查询位置与兴趣点之间的距离的变化引起的。

**定理 1** 如果  $NDom(p_1) = \emptyset$ , 即  $p_1$  为非空间 skyline 点,那么  $p_1$  肯定属于 RNS 结果集。

当  $NDom(p_1) \neq \emptyset$ , 无论查询点位于什么位置,  $p_1$  都不可能同时在非空间和空间属性上被某个其他点支配。因此,将  $p_1$  这类非空间 skyline 点称作 RNS 的固定 skyline 点,其余的称为候选 skyline 点。

**定理 2** 如果  $NDom(p_1) \neq \emptyset$ , 给定查询范围  $R$ , 那么,  $p_1$  属于 RNS 结果集当且仅当  $\exists q \in R$ , 对  $\forall p \in NDom(p_1)$ , 都有  $d(p_1, q) < d(p, q)$ , 否则,它将被  $NDom(p_1)$  中的某个点支配。

根据 RNS 的定义可知,  $\forall p_i \in POI$ , 如果  $p_i$  包含在查询范围  $R$  中,那么  $p_i$  一定是该查询的 skyline 点,而对于查询范

围  $R$  之外的兴趣点  $p_i$ , 我们采用边界点(查询范围  $R$  与道路网的交点)替代法, 将基于范围  $R$  的 RNS 查询转化为基于边界点的 PNS 查询。

**定理 3(边界替换定理)** 对于给定的查询范围  $R$ , 基于范围  $R$  的 RNS 查询结果等价于所有边界点 PNS 查询结果的并集, 即

$$RNS(R, POI) = \bigcup_{i=1}^k PNS(cp_i, POI)$$

其中,  $cp_1, cp_2, \dots, cp_k$  为查询范围  $R$  与道路网的交点。范围  $R$  中的任何一个可能的查询点  $q$ , 到达范围  $R$  外的任一兴趣点  $p_i$  的最短路径必定要经过某一个与道路网的交点  $cp_k$ , 有  $d(p_i, cp_k) \leq d(p_i, q)$ , 则  $PNS(q, POI) \subseteq PNS(cp_k, POI)$ 。又根据 RNS 查询的定义, 其结果返回  $R$  中所有查询点 PNS 查询结果的并集, 因此有定理 3 成立。

B-RNS 算法如算法 1 所示, 首先将兴趣点集划分为固定 skyline 集和候选 skyline 集(第 3-7 行)。根据道路网 R-tree 索引查找范围  $R$  与道路网的交点, 并对每一个交点  $cp_i$  采用并行 Dijkstra 算法, 计算其到每一个候选点的距离, 作为该候选点的距离属性, 并比较每一个候选 skyline 点与其非空间支配集中其他兴趣点的距离属性, 找出基于  $cp_i$  的 PNS 查询的 skyline 集(第 12-17 行), 最后将所有  $cp_i$  的 skyline 集进行合并, 获得最终的结果集。

#### 算法 1(基本 RNS 查询处理算法 B-RNS)

输入: 查询范围  $R$

输出: 基于查询范围  $R$  的道路网 skyline 查询结果集 resultSP

1. resultSP  $\leftarrow \emptyset$
2. fixedSP  $\leftarrow \emptyset$  and candidateSP  $\leftarrow \emptyset$
3. for each  $p \in POI$  do
4.   if  $N\text{Dom}(p) = \emptyset$  Then
5.     add  $p$  to fixedSP and resultSP
6.   else
7.     add  $p$  to candidateSP
8. compute the cross points set CPoint of  $R$  and the road network based on the top R-tree
9. add CPoint to the graph model  $G$
10. for each  $cp_i \in \text{CPoint}$  do
11.   for each  $p \in \text{candidateSP}$  do
12.     compute the shortest distance between  $cp_i$  and  $p$   $d(cp_i, p)$
13.     for each  $p_j \in N\text{Dom}(p)$  do
14.       compute the shortest distance of  $cp_i$  and  $p_j$   $d(cp_i, p_j)$
15.     if  $\forall j, s. t. d(cp_i, p) < d(cp_i, p_j)$  Then
16.       add  $p$  to resultSP
17. Return resultSP

## 5 基于索引的 RNS 查询算法 I-RNS

B-RNS 算法首先筛选出固定 skyline 点, 有效减少了 POI 的处理数量, 提高了查询效率。但是在候选 skyline 集合与其非空间支配集中点的距离属性的比较上, 有大量的道路距离计算。尤其在 POI 数量增大, 道路网更加复杂的情况下, 这会带来查询效率的降低。另外, B-RNS 返回的查询结果是查询范围  $R$  内所有查询位置点 PNS 查询结果的并集, 但并不能很明确地表示在范围  $R$  上的不同位置的 skyline 查询结果。

针对以上问题, 本节中我们将道路网和兴趣点集的属性相结合, 设计了道路网 skyline 索引 SSR-tree, 提出一种优化的 RNS 查询算法——基于索引的 RNS 查询算法 I-RNS, 其有效地提高了 RNS 查询的查询效率和精度。

### 5.1 道路网 skyline 模型

对复杂的道路网问题, 通常使用简单的数学模型对其进行抽象, 本文中我们将双向道路网抽象为一个无向加权图  $G(V, E, W)$ , 其中  $V$  为顶点集  $V = \{v_1, v_2, \dots, v_n\}$ , 表示道路网中的道路交叉路口;  $E = \{e_1, e_2, \dots, e_k\}$  为边的集合, 表示连接各交叉路口的道路段;  $W$  为边  $E$  上的权值, 对应道路的长度。给定某个兴趣点集  $POI(POI \subseteq V)$ , 可以对其建立网络 Voronoi 图, 每个兴趣点  $p_i \in POI$ , 都有一个 Voronoi 单元<sup>[12]</sup>, 记为  $V(p_i, POI)$ ,  $p_i$  对应的 Voronoi 单元内的点到  $p_i$  的道路网距离总是最短的。

如图 2(a) 给出的一个道路网示例,  $\{p_1, p_2, p_3\}$  为道路网兴趣点集, 对应的网络 Voronoi 图如图 2(b) 所示,  $d(p, p_i)$  表示道路网中一点  $p$  到兴趣点  $p_i$  的距离。对于  $POI$  中的任意两个兴趣点  $p_i, p_j$ , 定义  $p_i$  在道路网上相对  $p_j$  的距离支配域为:

$$\text{Dom}_d(p_i, p_j) = \{p | d(p, p_i) < d(p, p_j), j \neq i, p \in POI\}$$

$\text{Dom}_d(p_i, p_j)$  定义了道路网上一些连续点的集合, 这些点到  $p_i$  的距离都小于到  $p_j$  的距离。因此,  $p_i$  相对于整个兴趣点集  $POI = \{p_1, p_2, \dots, p_n\}$  的道路网 Voronoi 单元  $V(p_i, POI)$  便可以定义为:

$$V(p_i, POI) = \bigcap_{j=1}^n \text{Dom}(p_i, p_j), j \neq i$$

对  $\forall p_i \in POI, V(p_i, POI)$  中的点距离  $p_i$  总是最近的, 图 2(b) 中用 3 种不同类别的线分别表示出了  $p_1, p_2, p_3$  在该道路网上的 Voronoi 单元。

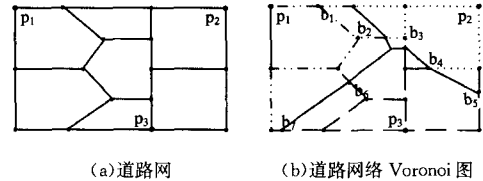


图 2 道路网络 Voronoi 图的示例

RNS 查询中, 需要比较候选 skyline 集中的点与其非空间支配集中的点到查询位置的距离, 因此, 我们引用网络 Voronoi 单元来建立道路网 skyline 模型。首先给出道路网 skyline 路段的定义:

**定义 5(道路网 skyline 路段)** 对于兴趣点  $p$ , 定义其道路网 skyline 路段  $NSS(p)$  为一些道路网路段上点的集合:  $NSS(p) = V(p, O)$ , 其中,  $O = \{p\} \cup N\text{Dom}(p)$ 。

因此, 对于  $\forall q \in NSS(p)$ , 都满足  $\forall p_i \in N\text{Dom}(p)$ , 由  $d(p, q) < d(p_i, q)$  可知, 如果查询点  $q$  落在兴趣点  $p$  的  $NSS(p)$  上, 那么,  $p$  到  $q$  的距离比  $p$  非空间支配集中的所有点都要近, 因此,  $p$  必定为该查询的 skyline 点。

对于兴趣点  $p$ , 其道路网 skyline 路段  $NSS(p)$  用一组四元组的集合  $\{(oid, rid, pos_1, pos_2)\}$  来表示, 其中  $oid$  为兴趣点的对象标志符,  $rid$  为道路标识符,  $pos_1, pos_2 \in [0, length]$ ,  $length$  为道路长度,  $pos_1, pos_2$  分别表示  $oid$  标志的兴趣点的

道路网 skyline 路段在  $rid$  道路上的始末位置到该条道路起点的距离。

对于第 3 节中的实例,其宾馆兴趣点集为  $\{p_1, p_2, p_3, p_4\}$ ,根据其非空间属性可以得到  $NDom(p_1) = NDom(p_4) = \emptyset, NDom(p_2) = \{p_1, p_4\}, NDom(p_3) = \{p_1, p_2, p_4\}$ ,  $p_1, p_4$  为非空间属性的 skyline 点集,即 RNS 查询的固定 skyline 点集,  $p_2, p_3$  的道路网 skyline 路段分别为  $p_2, p_3$  相对于  $\{p_1, p_2, p_4\}$  和  $\{p_1, p_2, p_3, p_4\}$  的 Voronoi 单元,如图 3 所示。

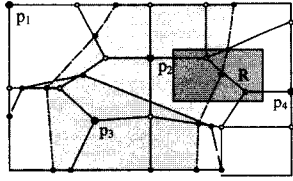


图 3  $p_2, p_3$  在道路网中的 skyline 路段

给定一个查询范围  $R$ ,由于  $R$  只与  $p_2$  的道路网 skyline 路段有交,则  $p_2$  为该 RNS 查询的动态 skyline 点。所以,该 RNS 查询的 skyline 点集为  $\{p_1, p_2, p_4\}$ 。查询返回结果为查询范围内的 skyline 点集以及它们的有效路段,如表 2 所列。

表 2 基于范围的道路网 skyline 查询结果集示例

道路路段	结果集
$(r_3, 405.2, 577.41)$	$\{p_1, p_2, p_4\}$
$(r_5, 337.64, 520.1)$	$\{p_1, p_2, p_4\}$
$(r_5, 520.1, 847.02)$	$\{p_1, p_4\}$
$(r_6, 0, 324.9)$	$\{p_1, p_4\}$

## 5.2 SSR-tree 索引

兴趣点的道路网 skyline 路段由于其非空间属性的支配集和它的空间位置有关,而与查询点的位置无关,因此是可预计算的。所有兴趣点的  $NSS(p)$  信息是按照其兴趣点  $oid$  进行组织的,如果将所有兴趣点的  $NSS$  信息合并,按照其  $rid$  进行数据重组,可以得到一组如  $roadSP(rid, pos_1, pos_2, SPs)$  四元组形式的以道路  $rid$  为单位的道路网 skyline 路段信息。本节中针对道路网 skyline 模型提出了一种基于道路网的 skyline 路段的索引 SSR-tree。

SSR-tree 索引是基于道路网索引的,从逻辑上分为两层,如图 4 所示。上层为 2D R-tree 结构,即  $SSR\_top$ ,用来对道路网信息进行索引,采用最小包容矩形 MBR 进行索引;中间节点包含  $\langle mbb, childpt \rangle$ ,  $mbb$  为包含所有子节点的 MBR;叶节点包含  $\langle mbb, polypt, treept \rangle$ ,  $mbb$  为道路折线的 MBR,  $polypt$  指向道路折线的详细信息,  $treept$  指向与该条道路相对应的下层 B-tree。

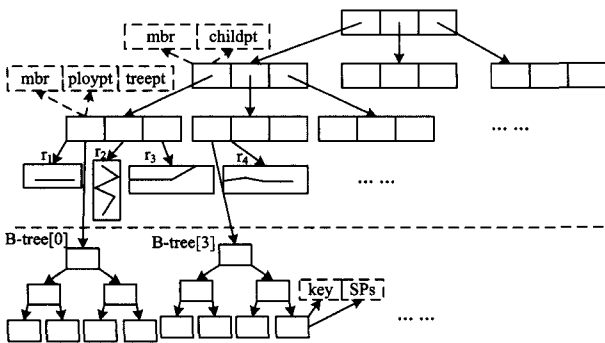


图 4 SSR-tree 索引结构

SSR-tree 下层为 1D B-tree 集,每一个上层叶子节点对应一个下层 B-tree,即  $SSR\_bottom$ ,索引在该条道路上的 skyline 点集(即 skyline 路段落在该道路上的兴趣点集)及其有效范围。下层 B-tree 按与该道路起点的距离进行索引,每个节点信息包含  $\langle key, sp \rangle$ ,  $key$  为节点关键字,对应该道路上的 skyline 路段分界点的  $pos$  值,  $sp$  为从该节点前继节点  $key$  值到该节点  $key$  值之间的路段所对应的 skyline 点集。其建立算法如算法 2 所示。

### 算法 2(SSR-tree 的建立算法 CreateSSR-tree)

输入:道路网信息及所有兴趣点在道路网上的 skyline 路段信息

输出:用于 RNS 查询的道路网 skyline 索引 SSR-tree

1. create  $SSR\_top$  based on roads information
2. initialize each  $SSR\_bottom \leftarrow NULL$
3. for each  $p \in candidateSP$  do
4. compute  $NSS(p) = V(p, \{p\}, \cup Dom(p))$
5. combine all  $NSS_i = (oid, rid, pos_1, pos_2)$  to create the skyline points set for each road  $roadSP[rid] = \{(rid, pos_1, pos_2, SPs)\}$
6. for each  $rid$  do
7. add fixedSP to SPs inroadSP[rid]
8. create the corresponding  $SSR\_bottom$  of road  $rid$  based on roadSP[rid]

## 5.3 基于 SSR-tree 的 RNS 查询算法

结合兴趣点的非空间属性和在道路网上的位置信息,基于道路网 skyline 模型建立的 SSR-tree,能够有效地支持 RNS 查询。查询步骤如算法 3 所示,首先查找  $SSR\_top$ ,找到道路网上被  $R$  覆盖的具体路段(第 2 行),进而对每一个覆盖路段,查找其  $SSR\_bottom$ ,找到该覆盖路段起始点的位置,中序遍历  $SSR\_bottom$  直至到达该路段终止点位置,将起始点到终止点之间的路段及其对应的 skyline 点集添加到结果集中(第 4—7 行)。

### 算法 3(基于索引的 RNS 查询处理算法 I-RNS)

输入:查询范围  $R$

输出:基于查询范围  $R$  的道路网 skyline 查询结果集  $resultSP = \{(rid, pos_1, pos_2, SPs)\}$

1.  $resultSP \leftarrow \emptyset$
2. compute RoadSegments based on  $SSR\_top$  where  $R$  overlap the road network
3. for each  $(rid, pos_1, pos_2) \in RoadSegments$  do
4. search  $SSR\_bottom$  of  $rid$  and find the position of  $pos_1$  (equal to the key a node or between the keys of two nodes)
5. inorder traversal the  $SSR\_bottom$  from  $pos_1$  to  $pos_2$
6. for each tempnode the traversal passed do
7. Add the road segment and its SPs to resultSP
8. Return resultSP

相比 B-RNS 算法, I-RNS 算法将复杂的道路网距离计算在预处理中一次性完成。对每一次查询处理,只需要查找 SSR-tree,直接搜索查询结果,在查询效率和查询结果的精度上都有很大的提高。

## 6 实验结果与分析

为了验证本文所提方法的有效性,我们对其进行了实验分析。实验中的道路网数据采用 Berlin 城市的真实道路数

据<sup>[15]</sup>,共 3250 条道路信息(约 15000 个道路交叉点),兴趣点集则是在道路网上随机生成的数据点集,其非空间属性值在 [0,500]上随机生成。按照本文的查询处理思路,需要首先处理兴趣点集的非空间属性,本文采用基于 R-tree 的方法计算各兴趣点的非空间支配集。本文提出 B-RNS 和 I-RNS 算法将非空间属性与空间属性结合处理,并比较它们在查询效率上的优劣性,以及各相关参数,如道路条数(默认 2000)、POI 数量的大小(默认为 2000)、POI 非空间属性维数(默认为 3)、查询范围 R 的大小(默认为 200m \* 200m)对查询性能的影响。

本文的实验环境为:3.30GHz CPU,2.0GB 内存,Windows 7 操作系统。主要通过测试查询响应时间来比较两种算法的查询处理效率。对于 I-RNS,给定道路网信息,可以预计算兴趣点集的道路网 skyline 路段并建立 SSR-tree 索引,这一部分过程我们把它作为预处理建立 SSR-tree 索引部分。表 4 给出了实验中随兴趣点 POI 数量的变化,I-RNS 算法中 SSR-tree 建立时间的大小。

表 3 SSR-tree 建立时间

POI 数量(k)	0.5	1	1.5	2	2.5	3
预处理时间(mins)	1.96	4.15	10.89	20.86	37.46	65.37

### 6.1 道路条数对查询性能的影响

图 5 示出道路条数对查询性能的影响。在道路条数较少时,由于道路条数越少,等数量 POI 的分布就越密集,因此每条道路网上的 skyline 路段相对较多,I-RNS 算法的查询响应时间会随着道路条数的增多而减少;随着道路条数的增多,道路网变得复杂,查询范围所覆盖的道路网路段也增多,因此查询响应时间也会渐渐增大。但在总体上,I-RNS 算法的查询效率远远优于 B-RNS 算法,查询响应时间接近  $1:10^4$ 。

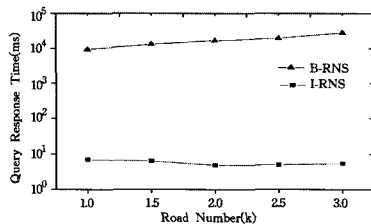


图 5 道路条数对查询性能的影响

### 6.2 POI 的数量对查询性能的影响

如图 6 所示,兴趣点集 POI 的大小从 500 变化到 3000,两种算法的查询响应时间均随 POI 数量的增大而增大,且 I-RNS 算法远远优于 B-RNS 算法。POI 数量增大时,B-RNS 算法中有大量的兴趣点距离计算,而基于索引的 I-RNS 算法在查询过程中只需要对道路网 skyline 索引 SSR-tree 进行查找,查找索引的效率要明显优于计算复杂的道路距离。

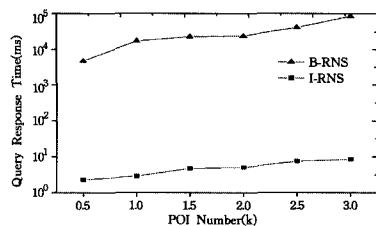


图 6 POI 的数量对查询性能的影响

### 6.3 POI 的非空间属性维数对查询性能的影响

实验中我们测试了 POI 的非空间属性维数从 2 维增加到 7 维的情况下两种查询算法的查询性能,实验结果如图 7 所示,I-RNS 算法的平均查询响应时间明显小于 B-RNS 算法,但随着非空间属性维数的增大,I-RNS 算法查询响应时间不断增大,而 B-RNS 算法的查询响应时间却越来越小。

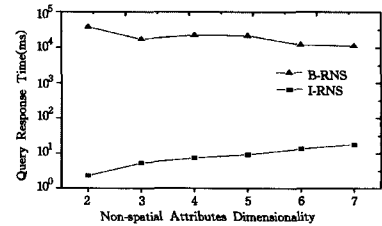


图 7 POI 的非空间属性维数对查询性能的影响

为了简便,本文中我们首先考虑了 POI 的非空间属性,筛选出 POI 中只考虑非空间属性的 skyline 点(即固定 skyline 点集),随着 POI 非空间属性维数的增大,其非空间属性之间相互支配的可能性会降低,从而其固定 skyline 点的数量会有所增大。剪枝掉这些固定 skyline 点,那么 B-RNS 算法需要处理的候选兴趣点的数量便减少,其查询处理时间也相应减少;而对于 I-RNS,尽管其固定 skyline 点增多,但我们也将它们结合到了道路网上,融合到道路网 skyline 索引中,因此在查询中对 SSR-tree 的查找并不会随固定 skyline 点的增多而变快。

### 6.4 查询范围 R 的大小对查询性能的影响

假设查询范围 R 为一个矩形区域,B-RNS 算法将 RNS 查询等价替换为 R 边界点的 PNS 查询。随着查询范围 R 面积的增大,B-RNS 算法中 R 与道路网交点的数量增大,I-RNS 算法中范围 R 中所包含的道路路段也增多。因此,如图 8 所示,随着查询范围 R 面积的增大,B-RNS 和 I-RNS 两种算法的查询响应时间都有所增大,但 I-RNS 的查询效率还是远远优于 B-RNS。

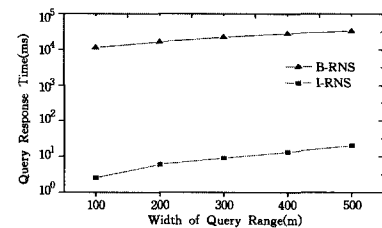


图 8 查询范围 R 的大小对查询性能的影响

### 6.5 查询精度的比较

对于 B-RNS 算法,根据定理 3,其 RNS 的查询结果为所有边界点 PNS 查询 skyline 集合的并集,结果集中包含了范围 R 中任意点作为查询点的 PNS 查询结果。因此,B-RNS 结果集中可能存在某些点并不是用户真实位置点的 skyline 点,并且用户也无法辨认。

而对于 I-RNS 算法,将 POI 的 skyline 路段与道路网相结合,查询结果中包含范围 R 中的每一个道路段以及在各道路段上查询点的 skyline 点集。用户可以根据自己当前的准备位置获得精确的 skyline 查询结果。经比较,I-RNS 算法在查询精度上较 B-RNS 算法更优。

**结束语** 本文针对道路网环境,考虑用户隐私保护和定位设备的精度问题,研究了基于范围的道路网 skyline 查询,提出了两种有效的查询算法,并通过实验对其查询性能进行比较分析。B-RNS 算法通过边界点替换法将 RNS 查询转化为基本的 PNS 查询,降低了复杂性,但仍有大量复杂的道路距离计算。针对这一问题,提出了一种道路网 skyline 模型,并基于模型设计了能有效地支持 RNS 查询的道路网 skyline 索引 SSR-tree,提出了基于索引的查询算法 I-RNS,经过实验验证,其提高了查询效率和查询结果的精度。

## 参 考 文 献

[1] Borzanyi S, Kossmann D, Stocker K. The Skyline Operator [C] // Proceedings of the 17th International Conference on Data Engineering (ICDE), 2001:421-430

[2] 周傲英,杨彬,金澈清,等. 基于位置的服务:架构与进展 [J]. 计算机学报,2011,34(7):1155-1171

[3] Tan K-L, Eng P-K, Ooi B-C. Efficient progressive skyline computation [C] // Proceedings of the 27th International Conference on Very Large Data Bases (VLDB), 2001:301-310

[4] Kossmann D, Ramsak F, Rost S. Shooting Stars in the Sky: An Online Algorithm for Skyline Queries [C] // Proceedings of the 28th International Conference on Very Large Data Bases (VLDB), 2002:275-296

[5] Papadias D, Tao Y-F, Fu G, et al. An Optimal and Progressive Algorithm for Skyline Queries [C] // Proceedings of SIGMOD, 2003:467-478

[6] Papadias D, Tao Y-F, Fu G, et al. Progressive Skyline Computation in Database System [J]. ACM Transactions on Database System (TODS), 2005, 30(1):41-82

[7] 付世昌,董一鸿,陈华辉,等. 基于道路网不确定移动对象的连续概率 skyline 查询 [J]. 计算机科学,2011,38(7):152-156

[8] Sharifzadeh M, Shahabi C. The Spatial Skyline Queries [C] // Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB), 2006:751-762

[9] Lin X, Xu J-L, Hu H-B. Range-Based Skyline Queries in Mobile Environments [J]. IEEE Transactions on Knowledge and Data Engineering (TKDE), 2013, 25(4):835-849

[10] Deng K, Zhou X-F, Shen H-T. Multi-source Skyline Query Processing in Road Networks [C] // Proceedings of the 23rd International Conference on Data Engineering (ICDE), 2007:796-805

[11] Huang Y-K, Chang C-H, et al. Continuous Distance-based Skyline Queries in Road Networks [J]. Information Systems (IS), 2012, 37(7):611-633

[12] Hu L, Jing Y, Ku W-S, et al. Enforcing K-Nearest Neighbor Query Integrity on Road Networks [C] // Proceedings of the 20th ACM SIGSPATIAL/GIS, 2012:422-425

[13] Ku W-S, Zimmermann R, Peng W-C, et al. Privacy Protected Query Processing on Spatial Networks [C] // Proceedings of ICDE Workshop Privacy Data Management, 2007:215-220

[14] Xu J, Tang X, Hu H, et al. Privacy-Conscious Location-Based Queries in Mobile Environments [J]. IEEE Transactions on Parallel and Distributed Systems, 2010, 21(3):313-326

[15] <http://www.bbbike.de/cgi-bin/bbbike.cgi>

(上接第 189 页)

[2] Wang T D, Parsia B, Hendler J, et al. A Survey of the Web Ontology Landscape [J]. Lecture Notes in Computer Science, 2006, 4273:682-694

[3] Ma Li, Yu He-long, Chen Gui-fen, et al. Research on Construction and SWRL Reasoning of Ontology of Maize Diseases [J]. Computer and Computing Technologies in Agriculture VI IFIP Advances in Information and Communication Technology, 2013, 393:386-23

[4] Guerrero A, Villagra V A, de Vergara J E L, et al. Ontology-Based Policy Refinement Using SWRL Rules for Management Information Definitions in OWL [J]. Large Scale Management of Distributed Systems Lecture Notes in Computer Science, 2006, 4269:227-232

[5] Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions [J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17:734-749

[6] Hurrell E, Smeaton A F. A Conversational Collaborative Filtering Approach to Recommendation [J]. Advances in Visual Informatics Lecture Notes in Computer Science, 2013, 8237:13-24

[7] Martin G H, Schockaert S, Cornelis C, et al. An Exploratory Study on Content-Based Filtering of Call for Papers [J]. Multidisciplinary Information Retrieval Lecture Notes in Computer Science, 2013, 8201:58-69

[8] Liu Xiao-qin, Mo Xiu-liang, Wang Chun-dong, et al. A Rule-Based Ontology for Context-Aware Computing [C] // International Conference on Biomedical Engineering and Informatics, 2009. Tianjin; IEEE Computer Society, 2009:1-4

[9] Lemdani R, Bennacer N, Polaillon G, et al. A collaborative and semantic-based approach for recommender systems [C] // International Conference on Intelligent Systems Design and Applications (ISDA), 2010. Cairo; IEEE Computer Society, 2010:469-476

[10] 王立才,孟祥武,张玉洁. 上下文感知推荐系统 [J]. 软件学报, 2012, 23(1):1-20

[11] Blanco-Fernandez Y, Pazos-arias J, Gil-Solla A, et al. Providing entertainment by content-based filtering and semantic reasoning in intelligent recommender systems [J]. IEEE Transactions on Consumer Electronics, 2008, 54:727-735

[12] 支丽平,王恒山,张楠. 专利领域本体的构建方法研究 [J]. 图书情报工作, 2010, 8:116-119