

文章编号: 2095-2163(2019)04-0202-04

中图分类号: TP301.6

文献标志码: A

# 基于路网的 k 最近邻查询算法综述

陈小迪, 冯 诚

(东北林业大学 信息与计算机工程学院, 哈尔滨 150040)

**摘要:** 在互联网时代, 基于地理位置的服务越来越普遍, k 最近邻查询通过与给定位置的距离来检索 k 个最近的兴趣点 (POIs), 是一个与之高度相关的查询。与欧式空间相比, 基于路网的 k 最近邻查询的研究更具有现实的意义和价值, 同时也面临着更大的挑战, 引起了国内外学者的广泛关注。本文将对基于路网的 k 最近邻查询算法的研究进展进行综述, 并展望该问题未来的研究方向和重点。

**关键词:** 基于地理位置的服务; 路网; k 最近邻; 最短路径; 路网距离

## Survey of k-nearest neighbor query algorithm on road networks

CHEN Xiaodi, FENG Cheng

(College of Information and Computer Engineering, Northeast Forestry University, Harbin 150040, China)

**[Abstract]** In this era of the Internet, Location-Based Services are becoming more common, and a k-nearest neighbor query which retrieves k nearest points of interest (POIs) by distance from a given location, is a highly relevant query. Compared with Euclidean space, the research of k-nearest neighbor query on road networks has more practical significance and value, and also faces greater challenges, so it has attracted extensive attention by scholars at home and abroad. This paper reviews the research achievements of the k-nearest neighbor query algorithm on road networks, and looks forward to the future research directions and key points for this problem.

**[Key words]** Location-Based Service; road network; k-nearest neighbor; shortest path; road network distance

## 0 引言

近年来, 随着移动互联网、全球定位系统和地理信息等技术的迅猛发展, 以及使用移动设备的人数的爆炸式增长, 基于地理位置的服务 (Location-Based Services, LBS) 也变得越来越普遍。k 最近邻查询作为基于地理位置服务中十分重要的支持性技术之一, 成为学术界的研究热点。

k 最近邻查询, 即在给定的空间数据集中, 返回距离查询顶点最近的 k 个数据对象, 主要包括基于欧式空间和基于路网的 2 类。传统的 k 最近邻技术都是基于欧式空间的<sup>[2-4]</sup>。例如, 用户使用美团时, 若选择以“离我最近”的方式显示查询结果, 就是根据用户与查询目标之间的欧式距离即直线距离进行排序的。因为 2 点之间的欧式距离只与坐标有关, 比较容易计算, 所以传统的 k 最近邻查询技术目前已经趋于成熟。但是人们在日常生活中, 位置和运动往往都会受到道路网络建设的约束, 不可能完全按直线移动, 所以, 基于路网的 k 最近邻查询技术越

来越受到国内外学者的关注。

基于路网的 k 最近邻查询, 相对基于欧式空间的 k 最近邻查询面临着更大的挑战。由于道路网络的数据规模相当庞大且数据结构复杂多样, 要求查询算法必须具有良好的存储结构以及可扩展性。而且, 基于路网的 k 最近邻查询在很大程度上与最短路径 (或路网距离) 有关, 不如欧式距离容易计算, 要求算法设计出能够快速确定 2 点间最短路径的方法。同时, 查询对象的总数通常比 k 大得多, 如果通过计算所有对象到查询位置的最短路径距离, 来获取 k 最近邻结果是不高效的, 这就要求算法设计出高效的剪枝策略, 能够尽可能多的忽略不是 k 最近邻的对象和与查询对象无关的路网转换。

目前, 基于路网的 k 最近邻查询技术已经得到了很大的发展, 本文将对有价值的研究成果进行综述。文献[1]对比较典型的查询算法进行了实现和比较 INE<sup>[6]</sup>、IER<sup>[6]</sup>、DisBrw<sup>[7]</sup>、ROAD<sup>[8-9]</sup> 和 G-tree<sup>[10-11]</sup> 方法。文献[5]分别对基于 Dijkstra 式搜索模式、基于启发式扩展模式、基于相邻区域迭代式

**基金项目:** 国家级大学生创新创业项目 (201710225131)。

**作者简介:** 陈小迪 (1997-), 女, 本科生, 主要研究方向: 计算机网络; 冯 诚 (1986-), 女, 博士, 讲师, 主要研究方向: 物联网。

**通讯作者:** 冯 诚 Email: fengcheng@nefu.edu.cn

**收稿日期:** 2019-04-22

扩展模式的 k 最近邻查询算法进行了分析和总结。

### 1 问题定义

因为图论的思想比较简单,所以目前国内外关于路网空间的研究大多是将路网建模为图,然后对具体问题求解。本文将路网表示为无向有权图  $G(V, E)$ ,其中: $V$  是顶点集合, $E$  是边集合,每条边  $e(u, v) \in E$  上都有一个正的权值  $w(u, v)$ ,表示距离、行驶时间等度量。定义任何 2 个顶点  $u, v \in V$  之间的最短路径为  $SP(u, v)$ ,最短路径距离为  $SPDist(u, v)$ ,即其是连接  $u$  和  $v$  之间的路径中权重总和的最小值。为了概念简单,假设每个用户兴趣点 (Points of Interest, POIs) 和查询点都位于  $V$  中的某个顶点上。例如,图 1 表示一个道路网络的图模型, $e(v_1, v_2)$  为 6, $SP(v_4, v_9)$  为  $v_4v_3v_2v_6v_7v_8v_9$ ,且  $SPDist(v_4, v_9) = 15$ 。

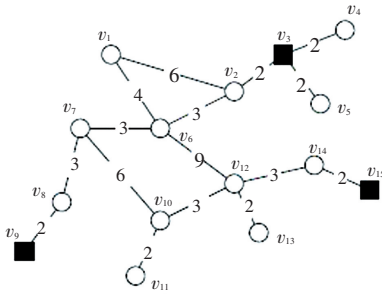


图 1 路网模型

Fig. 1 Road network model

k 最近邻查询 (k Nearest-Neighbor Query, kNN): 指在图  $G$  中,给定查询点  $q$ ,查询对象的集合  $O$  (目标 POIs),返回包含  $k$  个对象的结果集  $R$ ,满足:

- (1)  $R \subseteq O$ ;
- (2)  $\forall v \in R, u \in O - R$  有:  $SPDist(q, v) \leq SPDist(q, u)$ 。

连续 k 最近邻查询 (Continuous k Nearest Neighbor Query, CkNN): 指在图  $G$  中,查询点  $q$  和查询对象至少一方是移动的 k 最近邻查询。

### 2 基于路网的 k 最近邻查询算法

针对基于路网的 k 最近邻查询问题,已经涌现出来很多经典的算法。根据算法所用的查询方式主要分为 2 类:基于扩展的,例如本节中介绍的 INE<sup>[6]</sup>、Island<sup>[13]</sup>、S-GRID<sup>[14]</sup>、ROAD<sup>[8-9]</sup> 算法等;基于最佳优先遍历的,如本节介绍的 IER<sup>[6]</sup>、DisBrw<sup>[7]</sup>、G-tree<sup>[10-11]</sup>、DS-tree<sup>[15]</sup> 算法等。

在文献[6]中,Papadias 等人提出了 INE 和 IER 2 种方法。其中,INE 是扩展的 Dijkstra 算法,从查询位置逐步向邻近点扩展,直到获得 k 最近邻结果。IER 利用空间修剪技术改进了 INE,即通过欧式距离作为启发式从  $O$  中检索候选对象,因为它是任意连通 2 点之间距离的下界。

文献[12]中,Kolahdouzan 和 Shahabi 提出了基于网络 Voronoi 图的方法,用于将空间网络划分为 NVP,每个数据对应一个 NVP。并用空间访问法对 NVPs 进行索引,将问题简化为欧式空间中的点定位问题。并通过网络 vps 的预计算最小化在线网络距离。

文献[13]使用了 Island 方法解决 k 最近邻查询问题,以每个数据对象为中心,给定的  $r$  为半径形成 Island,Island 所覆盖的点都与其中心相关联。在该方法中,同时使用了预先计算的 Island 和从查询点开始的受限网络扩展。

文献[14]中引入了 S-GRID 算法,将空间网络划分为不相交的子网,并预先计算每对边界点的最短路径。为了找到 k 最近邻,首先在子网内执行网络扩展,然后利用预先计算的信息进行边界点之间的外部扩展。

在文献[7]中,Samet 等人提出了 DisBrw 方法,预先计算所有顶点对之间的最短路径,并使用基于二叉树的编码进行存储,通过将欧几里德距离存储为最短路径距离边界,然后具体找到 k 最近邻。

文献[8-9]提出了 ROAD 算法,通过建立路网索引和目标索引来实现对搜索空间的修剪,使整个算法更快地查询到最近的对象。构建路网索引时,将其划分为若干子图,对每个子图保存所有边界点之间的最短路径距离,如图 2 所示。在进行 k 最近邻查询时,根据目标索引若发现某个子图中没有查询对象,则直接利用这些快捷方式跳过对该子图的遍历。

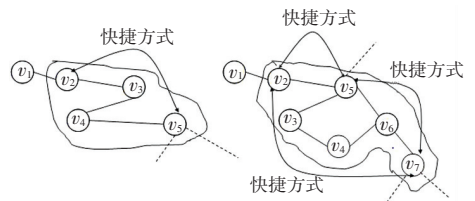


图 2 ROAD 的快捷方式  
Fig. 2 Shortcut of ROAD

文献[10-11]中提出了 G-tree 算法,通过 G-tree 索引以及最佳优先遍历的方法能够快速获得 k 最近邻查询结果,这也是目前最先进的算法之一。

在 G-tree 索引中,每个树节点都存储了一个距离矩阵,用于查询过程中快速计算  $q$  与查询对象之间的最短路径距离(使用基于拼接的方法)。同时,该算法构建目标索引发生列表,使那些没有目标对象的 G-tree 节点被过滤掉。

文献[15]中,采用新的索引技术 DS-tree 解决了 G-tree 算法在  $LCA(u, v)$  中计算最短路径可能出现错误的局部性的问题。DS-tree 中除根节点之外的所有节点都存储的是其子节点集合的收缩图,能够快速计算 2 点的最短路径距离。

### 3 连续 k 最近邻查询技术

连续 k 最近邻查询技术在日常生活中有广泛的应用,例如用户在某一位置查询距离自己最近的空闲出租车。出租车是处于运动状态的,且随着出租车的移动,返回的结果也会发生变化。

文献[16]中, Kyriakos Mouratidis 等人提出了 IMA 和 GMA 2 种实现技术。IMA 用扩展树进行存储,只有扩展树中的对象和边更新时,才能改变  $q$  的最近邻结果,从而忽略了无关更新。当最近邻结果更新或  $q$  移动到新位置时,IMA 保持扩展树中的有效部分,这样能很快查询到新的 k 最近邻结果。GMA 则使用 IMA 监视交叉点的 k 最近邻结果有效地计算路径中所有查询结果。

文献[17]中, Long Guo、Jie Shao 等人提出了 2 种以递增方式遍历网络的方法,即以查询为中心的算法(QCA)和以对象为中心的算法(OCA),这 2 种方法都维护一个树结构减少网络边缘的重复遍历,以获得更好的性能。

文献[18]提出了一种基于对象模型运动状态的对象候选处理(OCP)算法。在修剪阶段,算法修剪在给定时间间隔内不会是 k 最近邻查询结果的对象;在精炼阶段,确定给定时间间隔中能获得 k 最近邻查询结果的子间隔。

文献[19]提出了一种新的索引 V-Tree,有 2 个显著的特征:一是平衡的搜索树,可以支持高效的连续 k 最近邻查询;此外,还可以支持移动对象的动态更新。同时,在查询时还使用边界来有效地计算 k 最近邻。

文献[20]中, Kyoungso 提出了一种基于模式的方法,利用距离关系模式(DRP)来有效地处理连续 k 最近邻查询的方法。DRP 是按照单元格中的点与其它单元格之间的距离按升序排序的相对坐标列

表,以便通过顺序访问单元格。

### 4 结束语

基于路网的 k 最近邻查询技术目前已经取得了很多有价值的研究成果,给人们的生活带来了极大的便利。当然,随着路网日趋复杂以及人们的查询需求日趋多样,基于路网的 k 最近邻查询技术在深度和广度上都还有很大的发展空间。例如,从深度上讲,研究出速度更快、准确率更高的查询技术;从广度上讲,对更多的 k 最近邻变体技术进行研究,例如基于关键字的 k 最近邻查询、偏好 k 最近邻查询等。

### 参考文献

- [1] ABEYWICKRAMA T, CHEEMA M A, TANIAR D. K-nearest neighbors on road networks: A journey in experimentation and in-memory implementation [J]. Proc. of the VLDB Endowment, 2016,9(6):492-503.
- [2] SEIDL T, KRIEGEL H P. Optimal multi-step k-nearest neighbor search [J]. ACM SIGMOD Record, 1998, 27(2):154-165.
- [3] SHARIFZADEH M, SHAHABI C. Vor-Tree: R-trees with voronoi diagrams for efficient processing of spatial nearest neighbor queries [J]. Proceedings of the VLDB Endowment, 2010,3(1-2):1231-1242.
- [4] JI Shengyue, LI Chen. Location-based instant search [C]// BAYARD C J, FRENCH J, BOWERS S. Scientific and Statistical Database Management. SSDBM 2011. Lecture Notes in Computer Science. Berlin/Heidelberg: Springer, 2011,6809:17-36.
- [5] 鲍金玲,王斌,杨晓春,等. 路网环境下的最近邻查询技术 [J]. 软件学报,2018,29(3):642-662.
- [6] PAPADIAS D, ZHANG Jun, MAMOULIS N, et al. Query processing in spatial network databases [C]// Proceedings 2003 VLDB Conference. Berlin, Germany: dblp, 2008:802-813.
- [7] SAMET H, SANKARANARAYANAN J, ALBORZI H. Scalable network distance browsing in spatial databases [C]// Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD 2008). Vancouver, BC, Canada: IEEE, 2008:43-54.
- [8] LEE K C K, LEE W C, ZHENG Baihua, et al. Road: A new spatial object search framework for road networks [J]. IEEE Transactions on Knowledge and Data Engineering 2012, 24(3):547-560.
- [9] LEE K C K, LEE W C, ZHENG Baihua. Fast object search on road networks [C]// 12<sup>th</sup> International Conference on Extending Database Technology (EDBT 2009). Saint Petersburg, Russia: ACM, 2009:1018-1029.
- [10] ZHONG Ruicheng, LI Guoliang, TAN K L, et al. Gong. G-tree: An efficient and scalable index for spatial search on road networks [J]. IEEE Transactions on Knowledge and Data Engineering, 2015,27(8):2175-2189.

(下转第 208 页)