据模型和抽象定义,使用类似 Dijkstra 算法的思想来计算从查询点到目标对象间的最短距离,进而完成 C*k*NN 的查询.

Mouratids 等人[53]探讨了移动对象的 C*k*NN 监测问题,给出了一种增量监测算法 IMA(incremental monitoring algorithm).该算法在移动对象发生更新时,从前面的时间戳结果中增量获得查询结果用来重复评估当前的查询,使得重复查询产生的开销降低.但是位置更新的自然离散性使两个连续更新之间查询对象的 *k*NN 是未知的,造成 IMA 算法在两个连续更新的时间戳之间无法返回有效结果.

Huang 等人[54]提出了一种路网中移动对象 C*k*NN 的监测算法,该算法大体上可分为剪枝阶段和提炼阶段两个部分.方法中,由于剪枝策略中设置的剪枝距离过大,不能有效地去除一些不合格的对象,导致了提炼阶段的性能不佳.Demiryurek 等人[55]提出 ER-C*k*NN(Euclidian restriction based C*k*NN)查询处理方法,用于解决移动查询对象在动态变化的移动空间网络数据中的 CNN 查询.Liao 等人[56]给出了一种路网有向图模型,并且引入单向网络距离度量和双向网络距离度量方式,提出了单向网络扩展(UNE)算法和双向网络扩展(BNE)算法的 C*k*NN 查询处理方法.

Chen 等人[57]提出了 *k*-PNN(*k*-path nearest neighbor)问题,就是已知用户的当前位置和用户的目标位置,在当前位置到目标位置的最短路径上为用户提供满足其要求的最近邻对象列表,因为用户的当前位置是随时变化的,所以到目标位置的最短路径也是动态的.为了解决这一问题,文献[57]中提出了一种包括查找、验证和监视等 3 个阶段的 BNE(best-first network expansion)算法.

Li 等人[58−60]研究了解决路网中移动对象的 C*k*NN 查询的方法,处理速度不确定时移动对象的 *k*NN 查询问题.Shen 等人[61]提出了 V-Tree 索引结构,并基于这一索引结构处理路网上的移动对象的 *k*NN 查询.文献[61,62]中,利用 V-Tree 索引结构有效地降低了移动对象位置变化时索引更新的代价,在查询时可以有效过滤掉无关的节点,提高了移动对象的 *k*NN 查询性能,是目前最为高效的移动对象 *k*NN 查询方法.

3.2 聚集最近邻查询技术

聚集最近邻的定义是:已知一组数据点集合 P 和查询点集合 Q,从数据点集合 P 中查找到达查询点集合 Q 的聚集距离最小的点 p.比如几个朋友位于不同的地方,计划找一个到他们所在地方距离和最小的餐厅聚会.ANN 查询的关键就是如何减少不必要的距离计算和聚集距离计算.目前,主要的方法有利用欧式距离剪枝的改进的 IER 方法、利用平均距离作为剪枝下限的 TA(the threshold algorithm)算法和基于 Voronoi 的迭代扩展方法.

Yiu 等人[63]为了有效地解决 ANN 问题分别提出了 IER(incremental Euclidean restriction)算法、改进的 IER 算法、TA 算法和 CE(concurrent expansion)算法.其中的 IER 算法是利用欧式距离进行剪枝,利用改进的 A*算法查找 ANN 对象点,而 TA 算法和 CE 算法就是基于路网上的平均距离阈值作为剪枝条件,直接扩展查询实现的,CE 算法是对 TA 算法的改进.

文献[64]也在改进的 IER 框架基础上,采用欧式距离作为剪枝策略,提出了一种单源的多目标 A*算法处理 ANN 查询问题,并在验证阶段进行了改进.

Zhu 等人[65,66]提出利用 Voronoi 图的相邻性和预处理技术来解决 ANN 查询问题,文中将查询分为查找阶段和剪枝阶段.在查找阶段,分别从每个查询点开始寻找下一个最近邻目标对象,直至某个目标对象被所有查询点都扩展到,从而获得一个 ANN 候选集合;剪枝阶段就是去除 ANN 候选集合中不可能为 ANN 结果的目标对象,直到该候选集合中只有一个目标对象.文献[67]中又提出了一种基于 Voronoi 图的解决 *k*ANN 问题的方法,该方法首先通过 R 树索引查找出每一个查询点的 1NN,接着,按照某种策略不断计算某个查询 q 的下一个最近邻,然后检测这个最近邻对象点是否被所有的查询点扩展到:如果是,那么它就是所求的 1ANN;如果不是,那么更新它的当前聚集距离,再继续查找其他查询点的下一个最近邻.依此类推,最终可以找到所求的 *k*ANN 对象点.

还有的学者研究了连续的 ANN 查询问题.Qin 等人[68]提出了增量的 BUA(bidirectional updating algorithm)算法解决连续 ANN 查询问题,也就是从移动的数据点集中查找到多个查询点的聚合距离最小的 top- k 个数据点.Moumtidis 等人[69]研究了动态路网中的 ANN 查询.文献[69]中提出了 IMA 算法和 GMA 算法,IMA 算法存储最短路径扩展树并提出了 affecting edges 的概念,通过对目标节点更新、查询点更新和边权值更新这 3 种情况

分别讨论,并分别提出了相应的处理方法.在 GMA 算法中提出了 sequence 的概念,对处于同一 sequence 中的查询点利用 IMA 算法并行处理.

4 总结与展望

本文对路网环境下的最近邻查询技术进行了分类,从索引技术和查询方法两方面介绍了路网环境下的最近邻查询技术,并从网络规模、*k* 值的大小和对象分布密度 3 个因素探讨了最近邻查询技术之间的区别.目前的研究已经取得了阶段性的成果,但是还存在值得进一步研究的问题.

(1) 路网动态更新.路网数据是动态变化的:一是路网中的节点、路段和兴趣点是动态变化的;其次,路网中道路的权值是动态变化的;另外,查询点和查询对象都可能是动态变化的.在上述情况下,如何适应路网数据的动态更新、高效完成最近邻查询处理,是路网环境下最近邻查询技术需要面临的重大挑战.

(2) 考虑用户的个性化查询.随着基于位置服务的广泛应用,用户往往会有很多个性化的位置查询需求,这些查询大多都是以最近邻查询为基础的,比如带有时间约束或者包含关键字的最近邻查询问题.

(3) 可伸缩性.不同区域的路网规模不同、对象分布密度不同,而且同一区域路网的规模和对象分布密度也在不断增长,查询方法的可伸缩性就显得非常重要.目前,最近邻技术可伸缩性还有待进一步提高.

(4) 实验评测.文献[25]对路网环境下的主流技术进行了实验比较,但是评测还可以考虑更多的技术,同时也需要考虑各种技术在路网更新的情况下的适应性.

References:

[1] Donald K. The Art of Computer Programming (3). Indianapolis: Addison-Wesley Professional, 1973.

[2] Roussopoulos N, Kelley S, Vincent F. Nearest neighbor queries. ACM Sigmod Record, 1995,24(2):71−79. [doi: 10.1145/223784.223794]

[3] Papadopoulos AN, Yannis M. Performance of nearest neighbor queries in R-trees. In: Proc. of the Int’l Conf. on Database Theory. 1997. 394−408. [doi: 10.1007/3-540-62222-5_59]

[4] Hjaltason GR, Samet H. Distance browsing in spatial databases. ACM Trans. on Database Systems (TODS), 1999,24(2):265−318. [doi: 10.1145/320248.320255]

[5] Korn F, Sidiropoulos ND, Faloutsos C, Siegel EL, Protopapas Z. Fast nearest neighbor search in medical image databases. In: Proc. of the VLDB. 1996. 215−226.

[6] Seidl T, Kriegel H. Optimal multi-step *k*-nearest neighbor search. ACM SIGMOD Record, 1998,27(2):154−165. [doi: 10.1145/276304.276319]

[7] Berchtold S, Ertl B, Keim DA, Kriegel H, Seidl T. Fast nearest neighbor search in high-dimensional space. In: Proc. of the ICDE. 1998. 209−218.

[8] Lin K, Yang C. The ANN-tree: An index for efficient approximate nearest neighbor search. In: Proc. of the Database Systems for Advanced Applications. 2001. 174−181. [doi: 10.1109/DASFAA.2001.916376]

[9] Belussi A, Bertino E, Catania B. Using spatial data access structures for filtering nearest neighbor queries. Data & Knowledge Engineering, 2002,40(1):1−31. [doi: 10.1016/S0169-023X(01)00033-7]

[10] Hjaltason GR, Samet H. Ranking in Spatial Databases. 1995. 83−95. [doi: 10.1007/3-540-60159-7_6]

[11] Henrich A. A distance scan algorithm for spatial access structures. Int’l Journal of Geographical Information Science, 1994. 136−143.

[12] Sharifzadeh M, Shahabi C. Vor-Tree: R-trees with voronoi diagrams for efficient processing of spatial nearest neighbor queries. In: Proc. of the VLDB. 2010. 1231−1242. [doi: 10.14778/1920841.1920994]

[13] Zhu H, Yang X, Wang B, *et al*. Range-Based obstructed nearest neighbor queries. In: Proc. of the SIGMOD. 2016. 2053−2068. [doi: 10.1145/2882903.2915234]

[14] Kolahdouzan MR, Shahabi C. Voronoi-Based *k* nearest neighbor search for spatial network databases. In: Proc. of the VLDB. 2004. 840−851. [doi: 10.1016/B978-012088469-8.50074-7]

[15] Sankaranarayanan J, Alborzi H, Samet H. Efficient query processing on spatial networks. In: Proc. of the GIS. 2005. 200−209. [doi: 10.1145/1097064.1097093]

[16] Samet H, Sankaranarayanan J, Alborzi H. Scalable network distance browsing in spatial databases. In: Proc. of the SIGMOD. 2008. 43−54. [doi: 10.1145/1376616.1376623]

[17] Lee KCK, Lee WC, Zheng BH. Fast object search on road networks. In: Proc. of the EDBT. 2009. 1018−1029. [doi: 10.1145/1516360.1516476]

[18] Huang XG, Jensen CS, Šaltenis S. The Islands approach to nearest neighbor querying in spatial networks. In: Proc. of the Int'l Symp. on Spatial and Temporal Databases (SSTD). 2005. 73−90. [doi: 10.1007/11535331_5]

[19] Huang XG, Jensen CS, Šaltenis S. Multiple *k* nearest neighbor query processing in spatial network databases. In: Proc. of the Advances in Databases and Information Systems. Berlin, Heidelberg: Springer-Verlag, 2006. 266−281. [doi: 10.1007/11827252_21]

[20] Huang XG, Jensen CS, Liu H, Šaltenis S. S-GRID: A versatile approach to efficient query processing in spatial networks. In: Proc. of the SSTD. 2007. 93−111. [doi: 10.1007/978-3-540-73540-3_6]

[21] Papadias D, Zhang J, Mamoulis N, Tao YF. Query processing in spatial network databases. In: Proc. of the VLDB. 2003. 802−813. [doi: 10.1016/B978-012722442-8/50076-8]

[22] Hu HB, Lee DL, Lee VCS. Distance indexing on road networks. In: Proc. of the VLDB. 2006. 894−905.

[23] Hu HB, Lee DL, Xu JL. Fast nearest neighbor search on road networks. In: Proc. of the EDBT. 2006. 186−203. [doi: 10.1007/11687238_14]

[24] Zhong RC, Li GL, Tan KL, Zhou LZ. G-Tree: An efficient index for *k*nn search on road networks. In: Proc. of the CIKM. 2013. 39−48. [doi: 10.1145/2505515.2505749]

[25] Abeywickrama T, Cheema MA, Taniar D. *K*-Nearest neighbors on road networks: A journey in experimentation and in-memory implementation. Proc. of the VLDB Endowment, 2016,9(6):492−503. [doi: 10.14778/2904121.2904125]

[26] Gaede V, Günther O. Multidimensional access method. ACM Computing Surveys, 1998,30(2):170−231. [doi: 10.1145/280277.280279]

[27] Dijkstra EW. A note on two problems in connexion with graphs. In: Proc. of the Numerische Mathematik. 1959. 269−271. [doi: 10.1007/BF01386390]

[28] Sun Y. Design and implementation of nearst neighbor query in spatial network database. Computer Science, 2008,35(3):73−75 (in Chinese with English abstract). [doi: 10.3969/j.issn.1002-137X.2008.03.021]

[29] De AVT, Güting RH. Using Dijkstra's algorithm to incrementally find the *k*-nearest neighbors in spatial network databases. In: Proc. of the Symp. on Applied Computing. 2006. 58−62. [doi: 10.1145/1141277.1141291]

[30] Hou SJ, Liu GH, Yu J, Chu BY. An algorighm for *k* nearest neighbors queries in spatial network databases. Computer Science, 2006,s32(8):360−362 (in Chinese with English abstract).

[31] Safar M. *K* nearest neighbor search in navigation systems. Mobile Information Systems, 2005,1(3):207−224. [doi: 10.1155/2005/692568]

[32] Cho HJ, Chung CW. An efficient and scalable approach to CNN queries in a road network. In: Proc. of the VLDB. 2005. 865−876.

[33] Lee KCK, Lee WC, Zheng BH, Tian Y. Road: A new spatial object search framework for road networks. TKDE, 2012,24(3):547−560. [doi: 10.1109/TKDE.2010.243]

[34] Zhong RC, Li GL, Tan KL, Zhou LZ, Gong ZG. G-Tree: An efficient and scalable index for spatial search on road networks. TKDE, 2015,27(8):2175−2189. [doi: 10.1109/TKDE.2015.2399306]

[35] Wang H. Pre-Computing-Based algorithm for nearest neighbors leaping query over road network. Journal of Tianjin University of Technology, 2011,27(2):38−42 (in Chinese with English abstract). [doi: 10.3969/j.issn.1673-095X.2011.02.009]

[36] Zheng Z, Zhang SZ, Guo L, Shi BL. An approach to continuous *k* nearest neighbor query in road network. Journal of Computer Research and Development, 2007,44(s3):398−401 (in Chinese with English abstract).

[37] Wang B, Yang XC, Wang GR, Yu G, Zang WY, Yu M. Energy efficient approximate self-adaptive data collection in wireless sensor networks. Frontiers of Computer Science in China, 2016,10(5):936−950. [doi: 10.1007/s11704-016-4525-7]

[38] Feng J, Watanabe T. Search of continuous nearest target objects along route on large hierarchical road network. In: Proc. of the Data Engineering Workshop (DEWS). 2003. 45−50.

[39] Shekhar S, Yoo JS. Processing in-route nearest neighbor queries: A comparison of alternative approaches. In: Proc. of the Workshop Geographic Information Systems. 2003. 9−16. [doi: 10.1145/956676.956678]

[40] Yoo JS, Shekhar S. In-Route nearest neighbor queries. Geoinformatica, 2005,9(2):117−137. [doi: 10.1007/s10707-005-6671-1]

[41] Kolahdouzan MR, Shahabi C. Continuous *k*-nearest neighbor queries in spatial network databases. In: Proc. of the STDBM. 2004. 33−40.

[42] Kolahdouzan MR, Shahabi C. Alternative solutions for continuous *k* nearest neighbor queries in spatial network databases. GeoInformatica, 2005,9(4):321−341. [doi: 10.1007/s10707-005-4575-8]

[43] Guan YY, Xiao YY, Li YK. Continuous *k* nearest neighbor queries in road networks. Journal of Tianjin University of Technology, 2012,28(6):31−33, 43 (in Chinese with English abstract). [doi: 10.3969/j.issn.1673-095X.2012.06.008]

[44] Li XL, He YB. Continuous *k* nearest neighbor queries in road network based on network Voronoi diagrams. Information Technology, 2007,31(12):103−104, 108 (in Chinese with English abstract). [doi: 10.3969/j.issn.1009-2552.2007.12.031]

[45] Feng HY, Guo JF. Continuous nearest neighbor queries in road network. Computer Engineering, 2010,36(8):79−82 (in Chinese with English abstract). [doi: 10.3969/j.issn.1000-3428.2010.08.028]

[46] Li YH, Li JJ, Shu LC, Li Q, Li GH, Yang FM. Searching continuous nearest neighbors in road networks on the air. In: Proc. of the Information Systems. 2014. 177−194. [doi: 10.1016/j.is.2014.01.003]

[47] Shahabi C, Kolahdouzan MR, Sharifzadeh M. A road network embedding technique for *k*-nearest neighbor search in moving object databases. Int'l Journal of Geographical Information Science, 2002. 94−100. [doi: 10.1145/585147.585167]

[48] Shahabi C, Kolahdouzan MR, Sharifzadeh M. A road network embedding technique for *k*-nearest neighbor search in moving object databases. Geoinformatica, 2003,7(3):255−273. [doi: 10.1023/A:1025153016110]

[49] Kriegel H, Kroger P, Kunath P, Renz M. Proximity queries in large traffic networks. Int'l Journal of Geographical Information Science, 2007. 21−28. [doi: 10.1145/1341012.1341040]

[50] Wang B, Zhu R, Yang XC, Wang GR. Top-*k* representative documents query over geo-textual data stream. World Wide Web-Internet & Web Information Systems, 2017,20(8). [doi: 10.1007/s11280-017-0470-0]

[51] Kriegel H, Kroger P, Renz M, Schmidt T. Hierarchical graph embedding for efficient query processing in very large traffic networks. In: Proc. of the SSDBM. 2008. 150−167. [doi: 10.1007/978-3-540-69497-7_12]

[52] Jensen CS, Kolařvr J, Pedersen TB, Timko I. Nearest neighbor queries in road networks. In: Proc. of the Int'l Symp. on Advances in Geographic Information Systems. 2003. 1−8. [doi: 10.1145/956676.956677]

[53] Mouratidis K, Yiu ML, Papadias D, Mamoulis N. Continuous nearest neighbor monitoring in road networks. In: Proc. of the VLDB. 2006. 43−54.

[54] Huang YK, Chen ZW, Lee C. Continuous *k*-nearest neighbor query over moving objects in road networks. In: Proc. of the Advances in Data and Web Management. 2009. 27−38. [doi: 10.1007/978-3-642-00672-2_5]

[55] Demiryurek U, Banaeikashani F, Shahabi C. Efficient continuous nearest neighbor query in spatial networks using euclidean restriction. In: Proc. of the SSTD. 2009. 25−43. [doi: 10.1007/978-3-642-02982-0_5]

[56] Liao W, Zhang Q, Wu XP, Zhong ZN. Research on continuous *k* nearest neighbor queries in road networks. Journal of Chinese Computer Systems, 2010,3l(4):666−671 (in Chinese with English abstract).

[57] Chen ZB, Shen HT, Zhou XF, Yu JX. Monitoring path nearest neighbor in road networks. In: Proc. of the SIGMOD. 2009. 591−602. [doi: 10.1145/1559845.1559907]

[58] Li GH, Fan P, Li YH, Du JQ. An efficient technique for continuous *k*-nearest neighbor query processing on moving objects in a road network. In: Proc. of the Computer and Information Technology (CIT). 2010. 627−634. [doi: DOI10.1109/CIT.2010.127]

[59] Li GH, Li YH, Shu LY, Fan P. C*k*NN query processing over moving objects with uncertain speed in road networks. In: Proc. of the APWeb. LNCS 6612. 2011. 65−76. [doi: 10.1007/978-3-642-20291-9_9]

[60] Fan P, Li GH, Yuan L, Li YH. Vague continuous *k*-nearest neighbor queries over moving objects with uncertain velocity in road networks. Information Systems, 2012,37(1):13−32. [doi: 10.1016/j.is.2011.08.002]

[61] Shen BL, Zhao Y, Li GL, Zheng W, Qin Y, Yuan B, Rao Y. V-Tree: Efficient *k*NN search on moving objects with road-network constraints. In: Proc. of the ICDE. 2017. 609–620. [doi: 10.1109/ICDE.2017.115]

[62] Yang X, Wang B, Yang K, Liu C, Zheng B. A novel representation and compression for queries on trajectories in road networks. IEEE Trans. of Data Engineering (TKDE), 2018. [doi: 10.1109/TKDE.2017.2776927]

[63] Yiu ML, Mamoulis N, Papadias D. Aggregate nearest neighbor queries in road networks. TKDE, 2005,17(6):1–14. [doi: 10.1109/TKDE.2005.87]

[64] Htoo H, Ohsawa Y, Sonehara N, Sakauchi M. Aggregate nearest neighbor search methods using SSMTA* algorithm on road-network. In: Proc. of the Advances in Databases and Information Systems. 2012. 181–194. [doi: 10.1007/978-3-642-33074-2_14]

[65] Zhu L, Jing Y, Sun W, Mao DD, Liu P. Voronoi-Based aggregate nearest neighbor query processing in road networks. In: Proc. of the SIGSPATIAL Int'l Conf. on Advances in Geographic Information Systems. 2010. 518–521. [doi: 10.1145/1869790.1869876]

[66] Sun WW, Chen CN, Zhu L, Gao YJ, Jing YN, Li Q. On efficient aggregate nearest neighbor query processing in road networks. Journal of Computer Science and Technology, 2015,30(4):781–798. [doi: 10.1007/s11390-015-1560-z]

[67] Zhu L, Sun WW, Jing YN, Du JF. Voronoi-Based *k*-aggregate nearest neighbor query processing in road networks. Journal of Computer Research and Development, 2011,s48(10):155–162. [doi: 10.1145/1869790.1869876]

[68] Qin L, Ding B. Monitoring aggregate *k*-NN objects in road networks. In: Proc. of the Statistical and Scientific Database Management. 2008. 168–186. [doi: 10.1007/978-3-540-69497-7_13]

[69] Mouratidis K, Yiu ML, Papadias D, Mamoulis N. Continuous nearest neighbor monitoring in road networks. In: Proc. of the VLDB. 2006. 43–54.

附中文参考文献:

[28] 孙亚.空间网络数据库中最近邻查询的设计与实现.计算机科学,2008,35(3):73–75. [doi: 10.3969/j.issn.1002-137X.2008.03.021]

[30] 侯士江,刘国华,佘靖,褚兵义.空间网络数据库中的 *k* 个最近邻查询算法.计算机科学,2006,s32(8):360–362.

[35] 王恒.路网中基于预计算的跳跃式查询最近邻的算法.天津理工大学学报,2011,27(2):38–42. [doi: 10.3969/j.issn.1673-095X.2011.02.009]

[36] 郑铮,张守志,郭立,施伯乐.一种解决道路空间中连续 *k* 最近邻居查询的方法.计算机研究与发展,2007,44(s3):398–401.

[43] 管莹莹,肖迎元,李玉坤.基于路网的连续 *k* 最近邻查询.天津理工大学学报,2012,28(6):31–33,43. [doi: 10.3969/j.issn.1673-095X.2012.06.008]

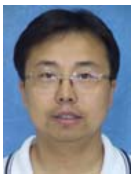
[44] 李晓丽,何云斌.基于网络 Voronoi 图的道路网络连续 *k* 近邻查询.信息技术,2007,31(12):103–104,108. [doi: 10.3969/j.issn.1009-2552.2007.12.031]

[45] 冯惠妍,郭俊风.道路网络中的连续最近邻查询.计算机工程,2010,36(8):79–82. [doi: 10.3969/j.issn.1000-3428.2010.08.028]

[56] 廖巍,张琪,吴晓平,钟志农.道路网络环境下的连续 *k* 近邻查询处理研究.小型微型计算机系统,2010,3l(4):666–671.

鲍金玲(1977－),女,吉林白城人,博士生,副教授,主要研究领域为空间数据库.

杨晓春(1973－),女,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,大数据管理与分析,数据隐私保护.



王斌(1972－),男,博士,副教授,主要研究领域为数据科学,数据流管理,数据质量管理.

朱怀杰(1988－),男,博士生,主要研究领域为空间数据库,隐私保护.