Keyword K-Nearest Neighbor Query of Road Network under Wireless Broadcast Environment

Chen Wen

School of Mathematics and Computer Science, Tongling College, Tongling, P.R.China
Email:tlxychenwen@163.com

Abstract

As a new branch of location-based query, space keyword query has gained extensive attention from researchers and become a new hotspot of database filed. Able to comprehensively consider the distance between query object and query point as well as similarity degree of keywords, such query can better satisfy specific requirements of location-based services. In this paper, a space keyword index structure is designed, so as to effectively organize keyword information, location information and index information of objects on various sides in road network. Besides, keyword k-nearest neighbor query of road network under wireless broadcast environment is proposed. Finally, performance of the proposed index structure and algorithm is verified through simulation experiment.

Keywords: *K-Nearest Neighbor Query, Spatial Database, Wireless Broadcast Environment, Road Network*

1. Introduction

In modern society, people often use mobile terminals like notebook computer and mobile phone to share network data and walk out from the limitations of fixed wired network gradually. The omnipresent information access has come true little by little. In practice, "location" is an important factor for the omnipresent information access. The demand for location-based information has promoted the emergence and development of a large number of location-based services (LBSs). As a supportive technique of location-based services, location-based query processing has already become a research hotspot.

In recent ten years, researchers have made an in-depth study on location-based query especially nearest neighbor query (NN) and reverse nearest neighbor query (RNN), and proposed many effective query processing methods. However, major forms of location-based query processing are based on traditional on-demand access patterns. On-demand access adopts a pull-based access data acquisition mode. Many mobile clients send out query request to the server and the server will conduct query processing in a concentrated way. Then the query results are fed back to mobile clients that send out the query in a point-to-point way. The data acquisition mode of on-demand access has a high requirement for network bandwidth and processing capacity of the server, so it is only suitable for light load system. When the number of query requests increases sharply, processing capacity of the server and network bandwidth will become the bottleneck of the entire processing system. Thus the query response time and timeliness will be affected and the system expandability can also be influenced. Periodic broadcast mode requires the server to actively broadcast information to the customers through a distribution wireless channel. The client side will monitor the communication channel to acquire information required by query processing, to automatically execute query processing algorithm to compete the execution process of query, and to acquire query

results finally. Such method allows various clients to visit data at the same time. Therefore, it is especially suitable for heavy load system. Owing to the good expandability, data broadcast mode has been extensively applied in mobile communication system and become a major data acquisition mode under mobile communication environment. Information access through wireless technology can be divided into two basic modes: on-demand access and periodic broadcast. With huge expandability and characteristics of satisfying a large number of clients, broadcast technology has become a very attractive and efficient technique in asymmetric system. In recent years, people have started to study location-based query processing under data broadcast mode, but the existing researches are based on Euclidean space, so they cannot be applied to location-based query processing under road network environment. In real life, the mobile objects move under road network environments like highway network and railway network in most cases. Therefore, it is very urgent and necessary to study location query processing algorithm in road network under data broadcast environment.

2. Relevant Work

In recent decades, R tree is extensively studied and applied to many services including kNN query. Its variants and other index structures are also applied to location-based query processing. Researchers will index data objects at first, so as to accelerate the query process. Under wireless broadcast environment, people have effectively processed location-based query under broadcast environment through continuous broadcast data items and index structures [1-6]. These index structures allow mobile clients to modulate to the data channel and receive relevant data only when related data they are interested in arrive. In this way, the modulation time can be reduced greatly and energy consumption will be decreased; this is very important for mobile clients. Other researchers dynamically adjust data broadcast according to some indexes such as the difference of access frequency [7-11] or access pattern [12-16]. Zheng et al. proposed grid-partition index [17] and D-tree index [18] on the basis of establishing Voronoi diagram, and processed location-based query including NN query under broadcast environment by utilizing these two index structures. However, their methods cannot be extended to kNN query processing. Some researchers transfer objects in two-dimensional or multi-dimensional spaces into one-dimensional objects of linear arrangement by observing the data distribution in the space and utilizing some space-filling curves such as Hilbert curve, and put forward corresponding index structures and index modes, so as to accelerate various kinds of query processing. B. Gedikde et al. [19] pointed out that effective kNN query could be conducted through broadcast R-tree on the basis of traditional kNN query algorithm [20]. However, this method only aims to reduce the modulation time, and does not consider the access delay and usage amount of internal storage. Chuan-Ming Liu [21] et al. added some useful information in R-tree structure and supported kNN query through broadcast modified R-tree index structure. This method has not only reduced the modulation time but also shortened the access time. Zheng et al. [22] proposed a distributed spatial index structure (DSI) to support location-based data access under wireless broadcast environment. DSI is a linear and fully distributed structure, and different query paths can share the same index link, so it is efficient. Meanwhile, DSI is also suitable for wireless broadcast environment where errors can be made easily. When wrong broadcast is received, DSI can recover query soon. This method supports typical location-based query, such as window query and kNN query (including snapshot kNN and continuous kNN). Kwangjin Park [23] et al. proposed a new space-time query processing algorithm to support continuous kNN query under wireless broadcast environment. This algorithm does not require any traditional space-time indexes in query processing and it can be applied to dynamic and static objects. Moreover, it does

not need other additional information except maximum speed and location of the objects. Lien-Fa Lin et al. [24] studied how to effectively organize location-based information and answer RNN query questions under wireless broadcast environment for the first time. Besides, they proposed an index structure called Jump-Rdnn tree that could adapt to linear access and effectively save electric power as well as corresponding RNN query algorithms.

However, the above algorithms only study location-based query under wireless broadcast environment in Euclidean space. In Euclidean space, the distance between object and query as well as the distance between object and object only depends on their coordinate values. In real life, the mobile objects move under road network environments like highway network and railway network in most cases, and the distance between objects is decided by the connectivity of road network. Due to the different distance calculation methods between Euclidean space and road network space, location-based query algorithm in Euclidean space cannot be applied to road network space through simple modification. The biggest difference between location query in road network and location query in Euclidean space is the difference of distance calculation methods between objects. In Euclidean space, the distance between objects is decided by relative location of the objects; however, in road network space, the distance between objects is decided by the connectivity of road network where the object is located, and two seemingly adjacent objects might have a huge road network distance. During location query processing, plentiful objects should be considered and the final query results can be determined only by calculating their road network distance with query point. Under broadcast environment, road network information is broadcasted by the server side through a group of ordering broadcast packets, and the client side receives some broadcast packets to acquire road network information. In order to improve the query processing efficiency, the road network structure should be reasonably divided and organized, so as to make the structure after division beneficial to communication in the form of broadcast packets. Besides, adjacent sides and nodes are located in broadcast packets near each other, so it is convenient to gain the required information at the client side. During division, the actual size of broadcast data packet should be considered. When the road network structure after division is organized, road network connectivity information should be retained.

3. Road Network Data and Index Structure

Generally speaking, source data of road network information are expressed with point set and side set, and common diagram storage modes include adjacent matrix and adjacent list. If such information is directly broadcasted, users need to acquire the final query results through a large amount of calculation, which is unsuitable under data broadcast environment. Data in data broadcast are often query results that can be directly gained by users or results that can be obtained by users through simple processing and calculation. Therefore, source data of road network will be re-organized at the server side, so as to make the final data structure form and organization form provide convenience for user query. In the following, definitions of some common symbols used in articles are given:

Table1. Symbol Definition

Symbol	definition
$G(v,e)$	Road network G: v is a point in G and e is a side in G
vr	Intersection set in G
vg	Query point set in G
vs	Break point set of Voronoi cell in G

As a regional division method, Voronoi diagram has a big advantage that NN to be queried can be gained immediately after the Voronoi cell of the searcher is judged. Division of Voronoi cells requires a lot of work, but the corresponding query method is simple, so such mode is especially suitable for the idea of thin client in data broadcast. Literature [25-26] studied Voronoi cell division based on road network, which meant to find out corresponding vr and vs sets for each vgi; Voronoi cell after the final division was a sub-diagram of the original diagram.

Under data broadcast environment, the key issue is to effectively dispatch broadcast data and reduce users' access time. Therefore, an effective index structure is proposed in the following.

A tree structure is established with information about two intersections of various sides in road network. As shown in Fig. 1, vr1 and vr4, vr2 and vr4, vr2 and vr5, vr3 and vr5 constitute multiple sides in road network. Each leaf node contains an object Ri located on the side; for instance, the leaf node of vr4 is the object R1 located on side (vr1, vr4).
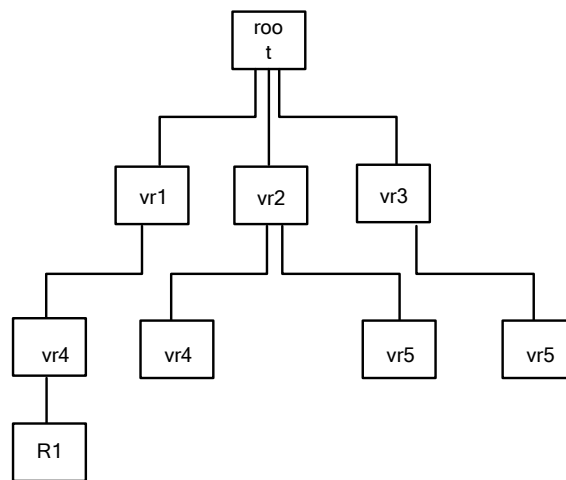

Figure 1. Tree Structure of Road Network Information

When space keyword query is processed, the object Ri not only contains location information but also includes a group of keyword information. Each node of the tree shown in Fig. 1 must contain a pointer pointing to relevant keyword information. Here inverted file is not used to store keyword information of the objects. Instead, inverted bitmap (InvB) that can save more space is adopted. Table 2 presents keyword situations of various objects in the system as well as inverted bitmaps corresponding to various nodes.

Table 2. Keyword Set

Object	keywords
R1	a,b
R2	a,c
R3	a,d
R4	e,f
R5	a,b

(b) Inverted Bitmap	
Side	*Keywords*
vr1	*a:11,b:10*
vr2	*a:11,c:11*
vr3	*a:00,d:10*
Vr4	*a:00,d:11*
Vr5	*a:10,d:11*

By taking vr1 node in Table 2(b) as an example, a:11 means that the left and right sub-trees contain objects that have the keyword "a": b:10 means that the left sub-tree contains objects that have the keyword "b", while the right sub-tree has no object that contains the keyword "b".

(1, m) index building mechanism is adopted, this index tree is repeated m times at first, and the data segment is evenly divided into m sections. Then the index is interlaced with data segment to form the final broadcast period. The value of m is taken by referring to the conclusion of literature [27], so as to reach an optimum balance between tuning time and access delay. The server broadcasts the built broadcast period to customers through wireless broadcast channels. After the client side initiates k-nearest neighbor query of space keywords, it will be tuned to broadcast channel, so as to receive index information. By analyzing the index information, regions that might contain query result objects can be gained. The pointer of index leaf node stores the arrival time of information required by query processing, and the client side will revive at the predetermined arrival time to acquire such information. In this way, query processing is conducted. Algorithm description is given in the following.

Keyword K-nearest neighbor query algorithm of road network under wireless broadcast environment

Input: Query point q

Output: K-nearest neighbor result set of space keywords of query q

Heap H=∅,list L,Candset=∅；

Monitor the broadcast channel to acquire the first index block;

Put the root node of index tree into the heap H;

dk＝∞；

while (H is non-empty)

{The heap tope element P is taken out from the heap;

if (P is non-leaf node), then

　　　{for (the child node Ri of P)

　　　if (InvB index corresponding to Ri indicates that the child node contains all keywords of query q ＆＆MinD(Ri,q)<dk)

　　　　　Insert Ri into H according to the ascending order of MinD(Ri,q);}

　　　else

　{if (Min(Ri,q)＜dk＆＆|L|<k)

　　　{Examine the InvB list of P; if it has objects containing all keywords of q, insert P into L according to the ascending order of MinD(P,q);

　　　If(dk>MaxD(Lk,q))

　　　Dk= MaxD(Lk,q)}}}

　for (Ri in various regions of L)

　　{Gain all objects o containing all keywords of q in Ri; put o into the candidate set according to the ascending order of distance value d(q,o) }

The previous k objects in the candidate set constitute the result set of query q.

After the client side gains the index structure, a heap H is initialized at first, and the root node of index tree is added into the heap H. In order to accelerate query processing, the algorithm establishes a candidate region set L and distance dk; the distance value dk and keyword set of query q are utilized to eliminate regions where result object cannot exist. Then the query client side receives predetermined information of various regions, and InvB contains relevant information about the region Ri containing all keywords of query q. Finally, objects o containing all keywords of q in these regions are added into the candidate result set according to the ascending order of d(q,o). When all regions that meet the conditions are processed, the previous k elements in the candidate set constitute the final query result set.

4. Experimental Analysis

Performance of the proposed algorithm is assessed by measuring the average access time and average tuning time required by processing for k-nearest neighbor query of keywords under wireless broadcast environment. In the experiment, the size of broadcast packet is 128byte, and the access time and tuning time are expressed with the number of broadcast packets received. The bandwidth of broadcast channel is 2Mb/s which is the commo n bandwidth of 3G network. In the experiment, performance of the proposed algorithm is tested via number of objects adjusted, number of k-nearest neighbors that should be acquired (i.e. k value), and number of keywords. The algorithm of this paper is realized with C++, and it runs in Windows XP with E3400 2.59GHz processor and 2 G internal storage.

Table 3. The Influence of Number of Objects on Tuning Time

The number of objects/1000	1	2	5	10
the number of broadcast packets	1612	2351	2614	2966

(1) Table 3 shows the influence of number of objects on tuning time of the algorithm. Tuning time rises with the increase of number of objects. When number of objects rises, density of objects increases and more information should be received during query processing.

Table 4. The Influence of K Value on Tuning Time

the value of k	10	20	50	100
the number of broadcast packets	2462	2573	2614	3054

(2) Table 4 shows the influence of k value on tuning time of the algorithm. When k value increases from 10 to 100, tuning time of the algorithm rises accordingly. Obviously, when k value rises, the size of both query space and result set will increase.

Table 5. The Influence of Number of Keywords on Tuning Time

The number of keywords	2	4	6	10
the number of broadcast packets	912	2351	2614	2966

(3) Table 5 shows the influence of number of keywords on tuning time of the algorithm. Tuning time of the algorithm rises with the increase of number of keywords. When number of keywords rises, the possible range of candidate objects will increase, thus information of more regions and objects should be downloaded.

Table 6. The Influence of Number of Objects on Access Time

The number of objects/1000	1	2	5	10
the number of broadcast packets	12.345	13.514	16.210	17.854

(4) Table 6 shows the influence of number of objects on access time of the algorithm. Access time of the algorithm rises with the increase of number of objects.

Table 7. The Influence of K Value on Access Time

the value of k	10	20	50	100
the number of broadcast packets	11.114	12.258	16.210	18.355

(5) Table 7 shows the influence of k value on access time of the algorithm. Access time of the algorithm rises with the increase of k value. As for the reason, with the rise of k value, number of objects involved in query processing also increases, so information of more objects should be acquired.

Table 8. The Influence of Number of Keywords on Access Time

The number of keywords	2	4	6	10
the number of broadcast packets	10.249	12.158	16.210	20.224

(6) Table 8 shows the influence of number of keywords on access time of the algorithm. When number of keywords rises, access time of the algorithm increases accordingly. As for the reason, the rise of number of keywords means that the possible range of candidate objects will increase, thus information of more regions and objects should be downloaded.

5. Conclusion

As a new branch of location-based query, k-nearest neighbor query of road network keywords under wireless broadcast environment has gained extensive attention from researchers. In this paper, an effective space keyword index structure of road network information is proposed, and corresponding k-nearest neighbor query algorithm of keywords is provided. Next, we will further study other query issues under wireless broadcast environment.

Acknowledgements

References

[1] T. Imielinski, S. Viswanathan and B. R. Badrinath, "Energy Efficient Indexing on Air", In: Proceedings of the 1994 ACM SIGMOD International Conference on Management of Data, **(1994)**, pp. 25-36.

[2] S. Hambrusch, C. M. Liu, W. G. Aref and S. Prabhakar, "Efficient Query Execution on Broadcasted Index Tree Structures", Data Knowledge Eng, vol. 60, no. 3, **(2007)**, pp. 511-529.

[3] S. Hambrusch, C. M. Liu and S. Prabhakar, "Broadcasting and Querying Multi-Dimensional Index Trees in a Multi-Channel Environment", Inf. Syst, vol. 31, no. 8, **(2006)**, pp. 870-886.

[4] S. Khanna and S. Zhou, "On Indexed Data Broadcast", J, Comput, Syst, Sci, vol. 60, no. 3, **(2000)**, pp. 575-591.

[5] J. X. Yu and K. L. Tan, "An Analysis of Selective Tuning Schemes for Non-Uniform Broadcast", Data Knowledge Eng, vol. 22, no. 3, **(1997)**, pp. 319-344.

[6] J. Zhang and L. Gruenwald, "Optimizing Data Placement Over Wireless Broadcast Channel for Multi-Dimensional Range Query Processing", In Proceedings of the 5th IEEE International Conference on Mobile Data Management（MDM 2004), **(2004)**, pp. 256-265.

[7] S. Acharya, R. Alonso, M. Franklin and S. Zdonik, "Broadcast Disks: Data Management for Asymmetric Communication Environments", In Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, **(1995)**, pp. 99-210.

[8] S. Hameed and N. H. Vaidya, "Efficient Algorithms for Scheduling Data Broadcast", Wireless Networks, vol. 5, no. 3, **(1999)**, pp. 183-193.

[9] H. P. Hung and M. S. Chen, "On Exploring Channel Allocation in the Diverse Data Broadcasting Environment", In Proceedings of the 25th IEEE International Conference on Distributed Computing Systems（ICDCS' 05）, **(2005)**, pp. 729-738.

[10] C. J. Su, L. Tassiulas and V. J. Tsotras, "Broadcast Scheduling for Information Distribution", CM/Baltzer J, Wireless Network 5, no. 2, **(1999)**, pp. 137-147.

[11] B. Zheng, X. Wu, X. Jin and D. L. Lee, "Tosa: A Near –Optimal Scheduling Algorithm for Multi-Channel Data Broadcast", In Proceedings of the 6th International Conference on Mobile Data Management, **(2005)**, pp. 29-37.

[12] J. L. Huang, M. S. Chen and W.C. Peng, "Broadcasting Dependent data for Ordered Queries without Replication in a Multi-Channel Mobile Environment", In Proceedings of the 19th International Conference on Data Engineering, **(2003)**, pp. 692-694.

[13] A. R. Hurson and Y. Jiao, "Data Broadcasting in a Mobile Environment", In Wireless Information Highway, IRM Press, **(2004)**, pp. 96-154.（Chapter 4）

[14] C. M. Liu and K. F. Lin, "Disseminating Dependent Data in Wireless Broadcast Environments", Distributed Parallel Databases, Available Online, **(2007)**.

[15] K. F. Lin and C. M. Liu, "Schedules with Minimized Access Latency for Disseminating Dependent Information on Multiple Channels", In Proceedings of the IEEE International Conference on Sensor Networks, Ubiquitous and Trustworthy Computing (SUTC 2006),no. 1, **(2006)**, vol. 344-351.

[16] C. M. Liu and K. F. Lin, "Efficient Scheduling Algorithms for Disseminating Dependent Data in Wireless Mobile Environments", In Proceedings of the IEEE 2005 International Conference on Wireless Networks. Communications and Mobile Computing（WirelessCom 2005）, **(2005)**, pp. 375-380.

[17] B. Zheng, J. Xu, W. C. Lee and L. Lee, "Grid-Partition Index: a Hybrid Method for Nearest - Neighbor Queries in Wireless Location-Based Services", VLDB J, vol. 15, 1, **(2006)**, pp. 21-39.

[18] J. Xu, B. Zheng, W. C. Lee and D. L. Lee, "The D-Tree: an Index Structure for Planar Point Queries in Location Based Wireless Services", IEEE Trans, Knowledge Data Eng, vol. 16, no. 12, **(2004)**, pp. 1526-1542.

[19] B. Gedik, A. Singh and L. Liu, "Energy Efficient Exact kNN Search in Wireless Broadcast Environments", In Proceedings of the 12th Annual ACM International Workshop on Geographic Information Systems, **(2004)**, pp. 37-146.

[20] N. Roussopoulos, S. Kelley and F. Vincent, "Nearest Neighbor Queries", In Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, **(1995)**, pp. 71-79.

[21] C. M. Liu and S. Y. Fu, "Effective Protocols for kNN Search on Broadcast Multi-Dimensional Index Trees", Information Systems, vol. 33, **(2008)**, pp. 18-35.

[22] B. Zheng, W. C. Lee, C. Ken, K. Lee, D. L. Lee and M. Shao, "A Distributed Spatial Index for Error-Prone Wireless Data Broadcast", The VLDB Journal, vol. 18, **(2009)**, pp. 959-986.

[23] K. Park, H. Choo and P. Valduriez, "A Scalable Energy -Efficient Continuous Nearest Neighbor Search in Wireless Broadcast Systems", Wireless Network, vol. 16, **(2010)**, pp. 1011-1031.

[24] L. F. Lin and C. C. Chen, "Study on RNN Query in Broadcasting Environment", Ubiquitous Computing and Communication Journal, **(2008)**.

[25] S. L. Hakimi, M. Labbe and E. Schmeichel, "The voronoi partition of a network and its implications in location theory", In ORSA Journal on Computing, vol. 4, no. 4, **(1992)**, pp. 412-417.

[26] M. R. Kolahdouzan and C. Shahabi, "Voronoi-based k nearest neighbor search for spatial network databases", In Proceedings of Intl,Conf, on Very Large Database, **(2004)**, pp. 840-851.

[27] T. Imielinski, S. Viswanathan and B. R. Badrinath, "Data on air: Organization and access[J]", Knowledge and Data Engineering, IEEE Transactions on, vol. 9, no. 3, **(1997)**, pp. 353-372.

Author

Chen Wen is an Associate Professor in the School of Mathematics and Computer Science, Tongling College,Tongling, P.R.China. He holds a master degree in Computer Science and Technology from the Anhui University, Anhui, P.R.China. His previous research areas include privacy preserving.