Faculty of Information Technology
Monash University

Literature Review — Semester 2, 2014

Isolated Region Spatial Query

Jian Loong Liew 22545727

Supervisor: Associate Professor David Taniar

Contents

1 Introduction

The *Isolated Region* spatial query is a distinctive spatial query because of its requirements. It aims to achieve the very nature of being near a crowd yet not too far that it is isolated. It is visually easy to determine an isolated region given any number of objects bound within a region. However, it mathematically solving this in the terms of both the Euclidean and Spatial distance is challenging. The input of this query will be a set of static data points. $P = \{p_1, ..., p_n\}$. Each of these points will denote objects of interests (OOIs). The output of this query will be a region, R. This region will be the coined *Isolated Region*.

In order to propose a query processing solution to this query, the various nearest neighbour queries with their query processing solutions will be explored. This literature review attempts to categorize the various spatial query into different categories and identify the ones of interest with regards to solving the spatial query processing scenario at hand. The tradional reverse nearest neighbour query only returns objects, however the inverse version of these queries will return a region. These queries are also called the reverse queries.

Below is a graphical representation of the various spatial queries and the various algorithms proposed by authors to solve the scenario at hand. In order to begin attempting to propose a query processing solution for the problem at hand the understanding of the work done by previous authors in order to solve similar problems needs to be understood.

The important works here are

[4].

- Reverse Nearest Neighbour by Korn et al. (2000) [15]
- Group Nearnest Neighbour by Papadias et al. (2004) [19].
- Group Nearest Group by Deng et al. (2012) [7]
- Optimum Region by Xuan (2012) [30]

In order to process the various queries in the spatial road network are as follows

- Voronoi Diagrams
- VN^3 [14]
- Network Voronoi Diagram [32]

Formal defintion:

1.1 Overview of Spatial Queries

This diagram is based on deriving information from [27].

2 Nearest Neighbour

[13]

[22]

The nearest neighbour problem is considered to be one of the best known problems in computer science.

[23]

2.1 Range Euclidean Restriction

[21].

2.2 Aggregate Nearest Neighbour Query

Aggregate nearest neighbour queries returns the object that minimizes an aggregate distance function with respect to a set of query points [33]. A scenario for ANN is as follows: Assumming for example, n users at location $(q_1, ..., q_2)$. An ANN query outputs the facility that minimises the sum of distances that the users have to travel in order to meet there [20]. Thus, the input for an ANN query will be the set of static data points given to it.

2.3 Constrained Nearest Neighbour

[10].

2.4 Reverse k-Nearest Neighbour (RkNN)

The goal of a reverse k-nearest neighbour query is to identify the “influence” of query object on the whole data set [1].

A Reverse Nearest Neighbour (RNN) search is a method to retrieve all objects that consider the query point as the nearest neighbour. An example would be like when a marketing application in which the issue is to determine the business impact of opening an outlet of Company A at a given location. A simple task is to determine the segment of A's customer who would be likely to use this new facility [15]. Korn et al (2000) broken this down into two cases which are static and dynamic cases and also formalised a novel notion of influence based on reverse nearest neighbour queries and its variants. Influence sets based on reverse nearest neighbour (RNN) queries seem to caputre the notion of influence based on the example stated by them.

Cheema et al (2011), suggested a more generic concept called **influence zone** and showed that the influence zone can be used to efficiently compute the influence sets. This influence zone has various applications in location based services and decision support systems. This is because the influence zone may be used for market analysis as well as targeted marketing [5]. This concept is more generic compared to the notation of influence set proposed by Korn et al (2000).

This methodology to generate the influence zone of each location will be used in order to generate the influnce zone of each object of interest.

3 Constrained Range Search

[31].

[14]

[34]

[12]

4 Reverse Region Query

The input of this are objects and the output of this query is a region.

5 Group Nearest Neighbour

This was proposed by Papadias et al [19]. It is considered to be a novel form of Nearest Neighbour search. The GNN query is different from the traditional $k$nn query which only specifies a single query point, the GNN query has multiple query points. The input of this problem consist of a set of $P = \{p_1, ..., p_n\}$ of static data points in multi dimensional space and a group of query points. $Q = \{q_1, ..., q_n\}$. The ouput of this contains the $k > 1$ data points [19].

An example scenario for the GNN would would be to select a meeting place from all available meeting places. For example, for three executive directors who are located in three different places, to meet at the closest meeting place [27].

This is considered to be an expensive problem by definition because of the number of data points as well as query points it needs to process as it is considerably more complex than the traditional $k$nn queries. The reason for this complexity is mainly due to two reasons [16]. The first is because there are multiple query points being specified which requires more distance computation and the other is because the fact that the query point can be distributed within the data space in arbitary ways, creating a large search region.

5.1 Multiple Query Method (MQM)

This algorithm utilises the threshold algorithm where it performs incremental NN queries for each point in Q and combines their results [19]. The MQM retrieves the NN for every point in query set Q, it sometimes accesses the same tree nodes for different query points and this causes its cost to increase fast with the query set cardinality.

5.2 Single Point Method

5.3 Minimum Bounding Method (MBM)

5.4 Group Cloest Pair Method

5.5 Two Ellipse-based Pruning Method

Li et al (2005) suggested a distance pruning method using an ellipse. The pruning method is “If a point or an MBR is far away enough with respect to the two points we choose as the approximate ellipse, they cannot be in the final answer." This method is compared to the SPM method and the MBM method. The authors claim that this method works more efficiently than both SPM and MBM because the ellipse used in the methods are less distance computational during the search and can prune unqualified nodes more efficiently. IT is noted that the two ellipse-based pruning method can be used in both the depth-first and best-first travel paradigms.

6 Probablistic Group Nearest Query

[17].

7 Group Nearest Group

Group Nearest Group (GNG) query can be defined as a query which finds one data point p from a data point set D such that the total distance from p to the points in a query point set Q is minimal [7]. This is regarded as the generic version of the GNN query. When $k = 1$, a GNG query is reduced to a GNN query. A GNG query can also be called as a k-median clustering in operations search which is a partition based clustering problem with group data points into k which is a given number of clusters based on an optimization objective function.

The scenario of a GNG query is as follows. A security service provider plans to set up several new branches to serve several business districts which can be represented by a set of land marks, such as well-known buildings. The max number of branches is usually on constrainted by factors such as business cost. Under this constraint the provider wishes to select branch locations from many choices such that the reponse time to security alarms can be minimised that is the average distance between these landmarks to the nearest branch is minimal.

The GNG algorithm has its useful because it will find a meeting point from a set of groups, however it requires 2 set of inputs which are the

Query points with different weights are also useful in many situations [7].

7.1 Exhaustive Hierarchical Combination Algorithm (EHC)

7.2 Subset Hierarchical Replacement Algorithm (SHR)

7.3 Comparison of query processing solutions

7.4 Zone of Influence

8 Optimum Region

An optimum region query is as follows:

Another spatial query of interest is the *Optimum Region.* This problem appears in Euclidean space. Optimum region is defined as a region which includes all the points which can cover the maximum number of objects in the finite set as the center of a radius r [30]. The objective of the optimum region search is to find a region where the center of a circle with a fixed radius can be located to cover most objects of interest in the given set. An example of an optimum region is as follows:

- Build a hospital in a community and let most resident reach to the hospital within 20 minutes or 15 km.
- A company plans to establish a new Wi-Fi base station, the coverage range of which is 20 meters, to cover most Wi-Fi devices in the company.

The input for the optimum region are a set of object of interest and the output is the optimum region.

8.1 Circle Partition and Arcs Superposition (CPAS)

This algorithm relies on the polar coordinates of the system.

9 Skyline Query

10 Isolated Region

11 R-Trees

In order to efficiently solve the query processing in the multi dimensional splace, an efficient indexing system has been proposed by various authors. An example would be the *R-trees* [11]. This has then been further developed and enhanced by various authors. For example, a variation of the *R-trees* is the $R^+ - trees$ that avoids overlapping rectangles in intermediate nodes of the tree that is introduced. [3].

11.1 Rectangle Handling

Table from R+ tree goes here.

12 TPR Tree

[29].

13 Voronoi Diagram

[25]

Okabe et al in [18] presented a thorough discussion on regular and network Voronoi diagrams.

[6].

[9]

[2]

[24].

[26]

[28]

[8]

Kolahdouzan et al in [14] suggested an approach termed Voronoi based Network Nearest Neighbour (VN^3), which reduces the problem of distance computation in a very large network, in to the problem distance computation in a number of much smaller networks plus some additional table lookups. This is achieved by generating a first-order network Voronoi diagram over the points of interest.

13.1 On-line Progressive Expansion (OPE)

13.2 Off-line Precalculation (OPC)

14 Comparison

The table below list the various spatial queries based on their input, output as well as the various algorithms used.

Item		
Name	Input & Output	Algorithm
Aggregate Nearest Neighbour	per gram	13.65
Reverse Nearest Neighbour	each	0.01
Gnu	stuffed	92.50
Group Nearest Group (GNG)	stuffed	33.33
Optimum Region	frozen	Circle Partition and Arcs Superposition (CPAS)

References

[1] Elke Achtert, Christian Böhm, Peer Kröger, Peter Kunath, Alexey Pryakhin, and Matthias Renz. Efficient reverse k-nearest neighbor search in arbitrary metric spaces. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 515–526. ACM, 2006.

[2] Franz Aurenhammer. Voronoi diagramsa survey of a fundamental geometric data structure. *ACM Computing Surveys (CSUR)*, 23(3):345–405, 1991.

[3] Norbert Beckmann, Hans-Peter Kriegel, Ralf Schneider, and Bernhard Seeger. *The R*-tree: an efficient and robust access method for points and rectangles*, volume 19. ACM, 1990.

[4] Stefan Berchtold, Bernhard Ertl, Daniel A Keim, H-P Kriegel, and Thomas Seidl. Fast nearest neighbor search in high-dimensional space. In *Data Engineering, 1998. Proceedings., 14th International Conference on*, pages 209–218. IEEE, 1998.

[5] Muhammad Aamir Cheema, Xuemin Lin, Wenjie Zhang, and Ying Zhang. Influence zone: Efficiently processing reverse k nearest neighbors queries. In *Data Engineering (ICDE), 2011 IEEE 27th International Conference on*, pages 577–588. IEEE, 2011.

[6] Mark De Berg, Marc Van Kreveld, Mark Overmars, and Otfried Cheong Schwarzkopf. *Computational geometry*. Springer, 2000.

[7] Ke Deng, Shazia Sadiq, Xiaofang Zhou, Hu Xu, Gabriel Pui Cheong Fung, and Yansheng Lu. On group nearest group query processing. *Knowledge and Data Engineering, IEEE Transactions on*, 24(2):295–308, 2012.

[8] Ke Deng, Xiaofang Zhou, and Heng Tao Shen. Multi-source skyline query processing in road networks. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pages 796–805. IEEE, 2007.

[9] Qiang Du, Vance Faber, and Max Gunzburger. Centroidal voronoi tessellations: applications and algorithms. *SIAM review*, 41(4):637–676, 1999.

[10] Hakan Ferhatosmanoglu, Ioanna Stanoi, Divyakant Agrawal, and Amr El Abbadi. Constrained nearest neighbor queries. In *Advances in Spatial and Temporal Databases*, pages 257–276. Springer, 2001.

[11] Antonin Guttman. *R-trees: a dynamic index structure for spatial searching*, volume 14. ACM, 1984.

[12] Haibo Hu, Dik Lun Lee, and Jianliang Xu. Fast nearest neighbor search on road networks. In *Advances in Database Technology-EDBT 2006*, pages 186–203. Springer, 2006.

[13] Christian S Jensen, Jan Kolářvr, Torben Bach Pedersen, and Igor Timko. Nearest neighbor queries in road networks. In *Proceedings of the 11th ACM international symposium on Advances in geographic information systems*, pages 1–8. ACM, 2003.

[14] Mohammad Kolahdouzan and Cyrus Shahabi. Voronoi-based k nearest neighbor search for spatial network databases. In *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*, pages 840–851. VLDB Endowment, 2004.

[15] Flip Korn and S Muthukrishnan. Influence sets based on reverse nearest neighbor queries. In *ACM SIGMOD Record*, volume 29, pages 201–212. ACM, 2000.

[16] Hongga Li, Hua Lu, Bo Huang, and Zhiyong Huang. Two ellipse-based pruning methods for group nearest neighbor queries. In *Proceedings of the 13th annual ACM international workshop on Geographic information systems*, pages 192–199. ACM, 2005.

[17] Xiang Lian and Lei Chen. Probabilistic group nearest neighbor queries in uncertain databases. *Knowledge and Data Engineering, IEEE Transactions on*, 20(6):809–824, 2008.

[18] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, and Sung Nok Chiu. *Spatial tessellations: concepts and applications of Voronoi diagrams*, volume 501. John Wiley & Sons, 2009.

[19] Dimitris Papadias, Qiongmao Shen, Yufei Tao, and Kyriakos Mouratidis. Group nearest neighbor queries. In *Data Engineering, 2004. Proceedings. 20th International Conference on*, pages 301–312. IEEE, 2004.

[20] Dimitris Papadias, Yufei Tao, Kyriakos Mouratidis, and Chun Kit Hui. Aggregate nearest neighbor queries in spatial databases. *ACM Transactions on Database Systems (TODS)*, 30(2):529–576, 2005.

[21] Dimitris Papadias, Jun Zhang, Nikos Mamoulis, and Yufei Tao. Query processing in spatial network databases. In *Proceedings of the 29th international conference on Very large data bases-Volume 29*, pages 802–813. VLDB Endowment, 2003.

[22] Franco P Preparatat and Michael Ian Shamos. Computational geometry: an introduction. 1985.

[23] Nick Roussopoulos, Stephen Kelley, and Frédéric Vincent. Nearest neighbor queries. In *ACM sigmod record*, volume 24, pages 71–79. ACM, 1995.

[24] Maytham Safar, Dariush Ibrahimi, and David Taniar. Voronoi-based reverse nearest neighbor query processing on spatial networks. *Multimedia systems*, 15(5):295–308, 2009.

[25] Cyrus Shahabi, Mohammad R Kolahdouzan, and Mehdi Sharifzadeh. A road network embedding technique for k-nearest neighbor search in moving object databases. *GeoInformatica*, 7(3):255–273, 2003.

[26] Robin Sibson. A brief description of natural neighbour interpolation. *Interpreting multivariate data*, 21, 1981.

[27] David Taniar and Wenny Rahayu. A taxonomy for nearest neighbour queries in spatial databases. *Journal of Computer and System Sciences*, 79(7):1017–1039, 2013.

[28] David Taniar, Maytham Safar, Quoc Thai Tran, Wenny Rahayu, and Jong Hyuk Park. Spatial network rnn queries in gis. *The Computer Journal*, 54(4):617–627, 2011.

[29] Yufei Tao, Dimitris Papadias, and Jimeng Sun. The tpr*-tree: an optimized spatio-temporal access method for predictive queries. In *Proceedings of the 29th international conference on Very large data bases-Volume 29*, pages 790–801. VLDB Endowment, 2003.

[30] Kefeng Xuan. Range / Region-Related Queries Processing in Spatial Databases by Clayton School of Information Technology Monash University. (0190), 2012.

[31] Kefeng Xuan, Geng Zhao, David Taniar, Maytham Safar, and Bala Srinivasan. Constrained range search query processing on road networks. *Concurrency and Computation: Practice and Experience*, 23(5):491–504, 2011.

[32] Kefeng Xuan, Geng Zhao, David Taniar, Bala Srinivasan, Maytham Safar, and Marina Gavrilova. Network voronoi diagram based range search. In *Advanced Information Net-*

working and Applications, 2009. AINA'09. International Conference on, pages 741–748. IEEE, 2009.

[33] Man Lung Yiu, Nikos Mamoulis, and Dimitris Papadias. Aggregate nearest neighbor queries in road networks. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):820–833, 2005.

[34] Geng Zhao, Kefeng Xuan, David Taniar, and Bala Srinivasan. Incremental k-nearest-neighbor search on road networks. *Journal of Interconnection Networks*, 9(04):455–470, 2008.