ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

A Survey on Scalable Distributed Processing of K nearest Neighbour Queries over Moving Objects

Prof.A.N Banubakode , Gajanan Payghan , Kokane Anirudh, Panchal Suraj, Kamble Rohit

Department of Information Technology, JSPM's RajarshiShahu College of Engineering, Pune, Savitribai Phule Pune University, India

ABSTRACT: Many applications involving moving objects is the task of processing k-nearest neighbor (k-NN) queries. Most of the existing approaches to this problem are designed for the centralized setting where query processing takes place on a single server; it is difficult, if not impossible, for them to scale to a distributed setting to handle the vast volume of data and concurrent queries that are increasingly common in those applications. To address this problem, we propose a suite of solutions that can support scalable distributed processing of k-NN queries. We first present a new index structure called Dynamic Strip Index (DSI), which can better adapt to different data distributions than exiting grid indexes. Moreover, it can be naturally distributed across the cluster, therefore lending itself well to distribute processing. We further propose a distributed k-NN search (DKNN) algorithm based on DSI.
The advances of GPS technology and wide-ranging usage of wireless communication devices have facilitated the collection of large amount of spatiotemporal data. Of special interest are data related to moving objects. By modeling and analyzing moving objects data, we learn about the moving objects behavior and even predict their future locations. Discovery of patterns and prediction of future movement can greatly influence different fields. Devising the correct inference is a scientific and computational challenge. Fundamentals of moving objects' representation and Space partitioning are presented. Markov Models are described in general along with their application to spatiotemporal prediction. Process of prediction and possible enhancements are shown.

KEYWORDS: k Nearest Neighbor, Multiple kNN queries, spatio-temporal database, information sharing strategy, queries processing.

I.INTRODUCTION

All moving things in the real world can comprise spatiotemporal data, containing time and space attributes simultaneously [2]. Data that are moving in time with changes of their location or shape describe a moving object. In many applications, knowing moving objects location in advance is very appreciable. Discovery of patterns and prediction of future movement can greatly influence different fields. Some of examples are analyzing wild animals' movement in order to predict their migrations, monitoring vehicles and analyzing their movement in order to predict traffic congestions, predicting the movement of a mobile user roaming around and changing access points to assure given level of quality of service in wireless networks or analyzing and predicting the movement of aircraft in combat in order to develop defending strategies. There are also situations in which determining the exact position of a moving object is not possible, for example when the moving object enters a shadow area of GPS, and estimating the location by referencing the previous ones which were provided in visible regions, becomes necessary. As technology advances, we encounter more available data on moving objects, thus increasing our ability to mine spatiotemporal data [3]. We can use these data to analyze moving objects behavior and to predict their future locations. k nearest neighbor (k-NN) queries over moving objects is a fundamental operation in many location-based applications.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

II.RELATED WORK

Although the movement is continuous, GPS and communication technologies allow us to sample an object's position, i.e. to obtain the position at discrete instances of time [1]. Interpolating these samples, we can extract object's movement. The simplest approach is to use linear interpolation. The sampled positions become the end points of line segments of polylines, and the movement of an object is represented by entire polyline in three-dimensional space. The movement of an object (i.e. the trace of an object) is called trajectory. Trajectories can have different characteristics depending on the characteristics of moving objects they represent, or depending on the application requirements. Two most important general characteristics of trajectories are direction and speed of the movement. Trajectories relate to pertaining spatial environment (such as ground cover, nature objects or objects in urban environment) or to other trajectories (moving objects of the same class or other moving objects). Further, trajectories can enter or can cross a spatial environment, or trajectories can (not) be within a spatial environment; they intersect, meet and can be near or far to other trajectories.

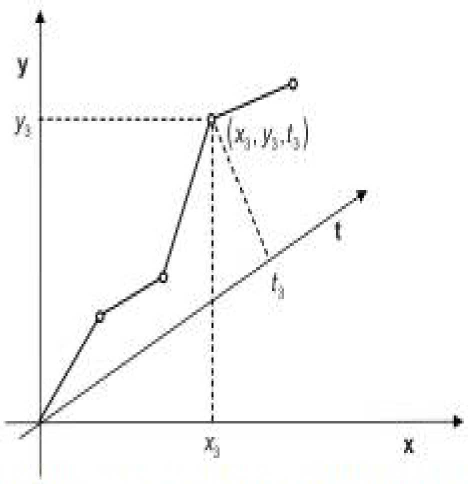


Figure 1 Trajectory - x and y represent coordinates and t represents time

Deploying DKNN We now discuss how to run the DKNN algorithm on S4 to process k-NN queries. The DKNN algorithm can be decomposed into several steps, and each step can be handled by a particular type of PEs. We call the resulting scheme DKNNS4.

The most notable advantage of DKNN is that even though the master node does not store the positions of objects, it can still determine the search space that contains the k-NNs in just two steps, by first directly determining the candidate strips using the DCS algorithm, and then identifying the final set of strips to search by computing the circle Cq. This is highly beneficial when the algorithm is running in a distributed system. In contrast, most existing search algorithms do not have this property. With those algorithms, the master cannot determine the final region for k-NN search without involving an uncertain number of rounds of communication between the master and slaves, incurring significant communication costs.

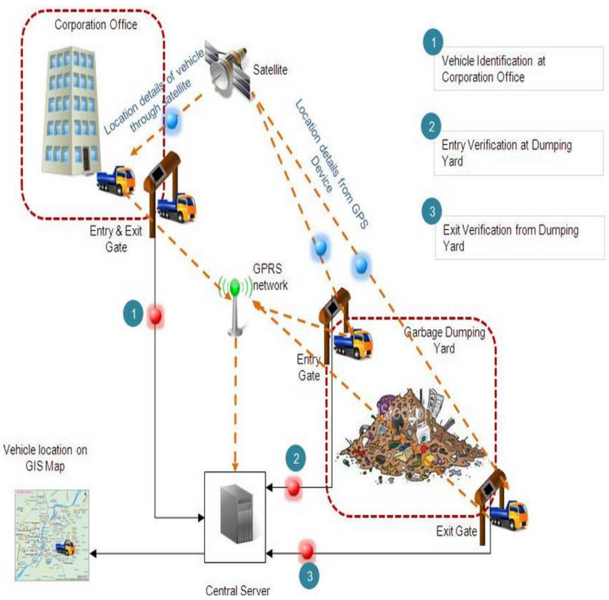
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

III.SYSTEM ARCHITECTURE



IV.PROPOSED SYSTEM MECHANISM

The large volume of data and heavy query workloads call for new scalable solutions. We propose DSI, a distributed strip index, and DKNN, a distributed k-NN search algorithm, to address this challenge. Both DSI and DKNN are designed with distributed processing in mind, and can be easily deployed to a distributed system. DSI is a data partitioning index and is able to adapt to different data distributions. Based on DSI, we present the DKNN algorithm that can directly determine a region that is guaranteed to contain the k-NNs for a given query with only two iterations. This has a clear cost benefit when compared with existing approaches, such as grid-based methods, which require an uncertain number of iterations. We show how the proposed index and algorithm can be implemented with S4. Extensive experiments confirm the superiority of the proposed method. We would like to explore how to evaluate continuous k-NN queries over moving objects using the strip index. For a given k-NN query q, it is very possible that its result (a list of objects) remains relatively stable when objects move with reasonable velocities. Therefore, it is promising to investigate how the k-NN results can be incrementally updated as objects move.

V. CONCLUSION AND FUTURE WORK

In this paper, we introduced a technique that can quickly identify the related queries of a given query q. After finding the related queries of q, our algorithm uses the information to infer a search region that covers the kNN results of q. We employ an R-Grid to reduce the distance computations when traversing an R-tree. The performance results show that the techniques proposed in the paper can speed up the performance and reduce the disk I/O cost of kNN searching. Currently, our algorithm can only deal with kNN queries. In the future, we plan to extend our work to permit various spatial queries (e.g., reverse kNN, aggregate NN and range queries). We plan to explore this subject further in the future.

REFERENCES

[1] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, Big Data: The Next Frontier for Innovation, Competition, and Productivity. New York, NY, USA: McKinsey Global Institute, 2011.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

[2] Wechat [Online]. Available: http://en.wikipedia.org/wiki/ WeChat, 2014. [3] X. Yu, K. Pu, and N. Koudas, "Monitoring k-nearest neighbor queries over moving objects," in Proc. 21st Int. Conf. Data Eng., 2005, pp. 631–642.
[4] B. Zheng, J. Xu, W.-C. Lee, and L. Lee, "Grid-partition index: A hybrid method for nearest-neighbor queries in wireless locationbased services," VLDB J., vol. 15, no. 1, pp. 21–39, 2006.
[5] K. Mouratidis, S. Bakiras, and D. Papadias, "Continuous monitoring of spatial queries in wireless broadcast environments," IEEE Trans. Mobile Comput., vol. 8, no. 10, pp. 1297–1311, Oct. 2009.
[6] J. Dean and S. Ghemawat, "MapReduce: Simplified data processing on large clusters," Commun. ACM, vol. 51, no. 1, pp. 107–113, 2008.