International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868
Foundation of Computer Science FCS, New York, USA
Volume 1– No.7, March 2012 – www.ijais.org

Bayesian nearest Neighbor Search in a Spatial Database

K. Balasaravanan
Research Scholar and Asst.Professor,
Department of Computer Science and Engineering,
Velalar College of Engineering and Technology,
Erode – 638 012, Tamilnadu, India

K. Duraiswamy
Dean (Academic),
Research Supervisor,
K.S.Rangasamy College of Technology,
Tiruchengode-637215, Tamilnadu, India

ABSTRACT
In an uncertain spatial database, identifying nearest neighbor is the important task to perform. To perform the nearest neighbor search (NN), existing work have presented Authenticated Multi-step NN (AMNN) and Superseding Nearest Neighbor (SNN) search. The AMNN efficiently performed the NN search using query authentication and trusted authority centre in which NN search has done only in single server not for distributed server and communication overhead also increased. The main drawback of SNN is that it cannot be applied to high dimensional data structure.

To overcome all these issues, in this paper we implements a new technique named BNN (Bayesian Nearest Neighbor) for NN search and similarity search in a spatial database. BNN performs NN search efficiently and retrieve the distance information not only from single server but also from distributed servers. It can be applied to high dimensional data structure and it automatically reduces the communication overhead. The query result returned by BNN will be a reliable one. The experimental evaluation shows that BNN performs Nearest neighbor search and similarity search well than existing AMNN and SNN.

Keywords
Keywords: Query Authentication, BNN (Bayesian Nearest Neighbor), Distributed server

1. INTRODUCTION
In spatial databases, an object is represented by probability density function (pdf). Applications which have high dimensional data structure will perform the Nearest Neighbor (NN) and similarity search using multi-step processing. For an illustration, take four objects W, X, Y, Z. Specially, if the probabilities of an object W are 0.4 and 0.6 , then it is termed as W[1] and W[2] respectively. So it is referred as W[1] and W[2] as the instances of W. Similarly, if object X also has two instances X[1] and X[2]. Whereas Object Y as Y[1] and Z as Z[1], because it has only one instance and its location has no uncertainty. For these types of uncertain spatial databases, Superseding NN (SNN) has been presented.

NN retrieval has been broadly considered in databases, computational geometry, machine learning, etc. In the sequel, the most vital outcomes in the database literature are that paying attention to the best first (BF) algorithm. For the applications that are in need of result authentication, multi-step NN search has been used. The multi-step NN framework can be motivated by applications that entail expensive distance computations. If the object in the uncertain spatial database does not have location, then the NN has been performed by maintaining location servers. But there is a great extent for loss of data since anyone can retrieve the information from the location server anonymously.

After performing the NN search, the AMNN perform the query verification. The query verification is done with by maintaining trusted third party centre. The Trusted third party will verify the queries and produce the outcome to the source point by attaching the digital set signatures. In SNN, an object is said to be a nearest neighbor (NN), if it has a nonzero probability. If no object is able to supersede every other NN-candidate, the SNN-core is presented, the minimum set of NN-candidates each of which supersedes all the NN-candidates outside the SNN-core. But all the existing SNN and AMNN perform the NN and similarity search only be efficient for the low dimensionality data structures. And both these retrieve the NN information only form the single server. It consumes more time for NN search and communication overhead also increases due to more number of transmissions of False Hits.

In this paper, we present a technique, BNN (Bayesian Nearest Neighbor) for NN search and similarity search in an uncertain spatial database which achieves NN search powerfully and recover the distance information not only from single server but also from distributed servers. BNN performs NN and similarity search efficiently for the high dimensional data structure and it automatically reduces the communication overhead.

2. LITERATURE REVIEW
Query processing for spatial database for the nearest neighbor search has been explained in various aspects. A lot of existing work has been done with uncertain spatial database for querying objects [2]. This Nearest Neighbor (NN) can be identified by using the probability density function from which the nearest neighbor has beyond the user-specified threshold to control the desired confidence required in a query answer to reduce the computational effort [4].

In [3], the NN has been found out by using the multistep processing in which it reduces the network overhead arises due to the transmission of false hits. In [5], two frameworks have been used for the spatial kNN query: the incremental euclidean restriction (IER) and incremental network expansion (INE). To identify the search space, IER uses eucliean distance and INE uses Dijkstra Algorithm fro identifying the nearest data objects [6]. From this, we conclude that INE performs better than IER. An extension of

IER proposed in incremental lower bound constraint (LBC) which rapidly reduces the network overhead.

In [7], the data objects can be identified by using the LBS (Location Based Service). In which it reveals the exact user locations and substitutes the anonymity techniques to identify the data objects location for query processing. Here two techniques has been proposed; 1) provide anonymous LBS access to the users and 2) allow efficient query processing at the LBS side. This achieves low computational task and low communication overhead.

Another algorithm to identify NN is based on depth first search on an R-Tree. This algorithm needs fewer memories and this will not be an appropriate one for the high dimensional spaces. This issue will lead to the development of several algorithms for high dimensional spaces in terms of NN. From this the distance between the two objects can be calculated easily and for consumes more amounts. Another approach [8], will identify the NN based on the supersede data object.

Several approaches have been used to identify the NN search, reverse NN search [9], aggregate NN search [10], continuous NN search [11], etc. To identify the top-1 search for the NN query, top- k algorithms have also been proposed to improve the NN search without any ambiguity [12].

In [1], superseding nearest neighbor search, every object is illustrated by a multidimensional probability density function in which it return the SNNcore. Therefore, the minimum set of NN-candidates every one of which supersedes every NN-candidate outside the core. Various adaptation of uncertain NN search is addressed in [13]. In particular, consider existentially uncertain objects. Specifically, an object will not depend on the database, if it depends, its location is specific. In our framework, an object absolutely exists, but its location is uncertain.

3. BAYESIAN NEAREST NEIGHBOR SEARCH FOR QUERY PROCESSING

To perform the NN and similarity search efficiently, many existing techniques (SNN, AMNN) has been presented. An existing authenticated multi step processing (AMNN) for Nearest Neighbor search for high dimensional data. Even though it has been presented with high dimensional data, it produced a unreliable query outcome. It presented an ID-AMN (Incremental Distributed AMN) for retrieving the distant values form the servers to compute DST. But it did not eliminate the server which does not contribute the results in distributed server. It uses the false hit (FH) reduction algorithm to minimize the weight of FH rate.

To perform the NN search in an uncertain spatial database, SNN presented a supersede NN search where each object is termed as a multidimensional probability density function. For a query point q, an object is termed as NN, if it has a nonzero probability. For an object which has no location, SNN core has been used. SNN core is the minimum set of NN-candidates each of which supersedes all the NN-candidates. The main drawback of SNN is it can be applied only for low dimensional data. To overcome all these issues, BNN search is presented here to perform the NN search efficiently for both high dimensional data and retrieve the information from the distributed servers and eliminate the server which is not involved in the result producing contribution.

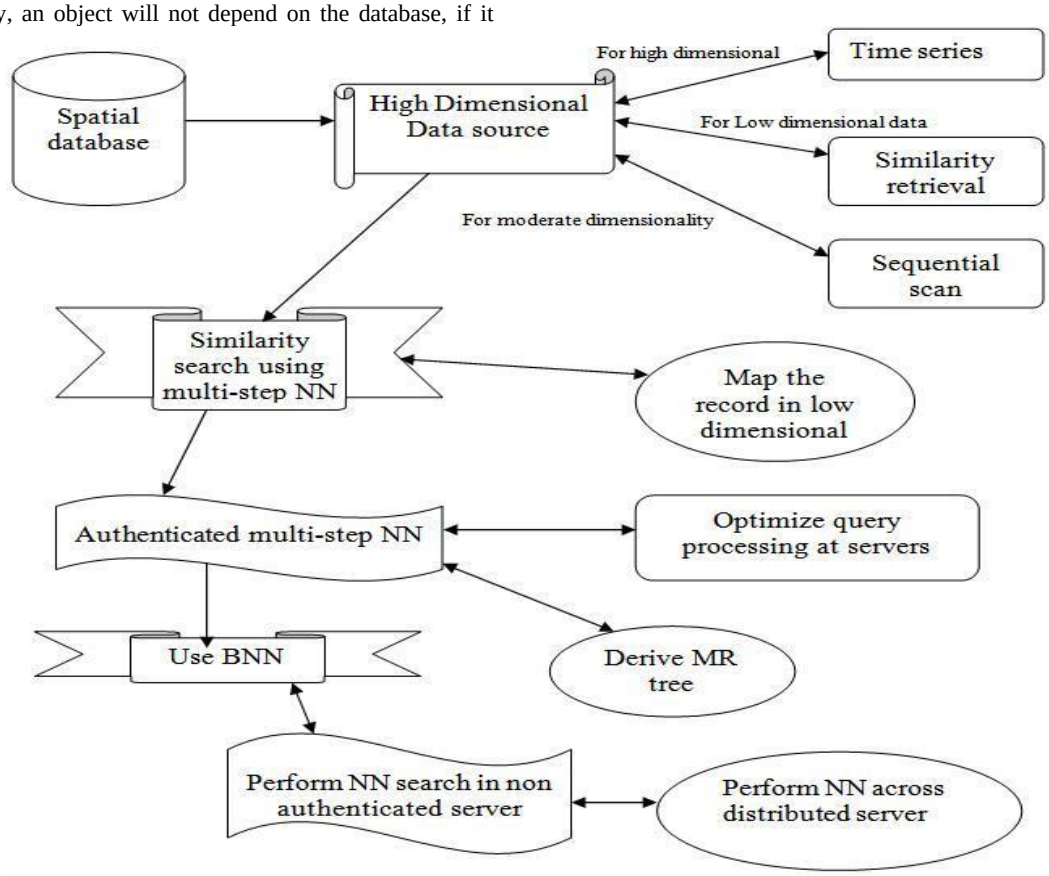

Fig 3.1 Authenticated NN search using BNN

The high dimensional data source used the specific technique for different types of dimensionality data structure. For high dimensional data NN search it use time series. After applying the data sources, the similarity search is performed to map the record in the database from distributed servers. The query from several servers has been processed with multi-step method. To exhibit the lower bounding property, it combine multi- step search with a dimensionality reduction. The pseudo code for the BNN search is:

1: Let each server be si where i= 0, 1, 2, 3…n

2: Get the queries qi where i= 0,1,2..n

3: For each qi do

4: If qi = high dimensional data

5: Use time series

6: Else if qi = low dimensional data

7: Use similarity retrieval

8: Else if qi = moderate dimensionality

9: Use sequential scan

10: End if

11: End if

12: End if

13: End For

14: Map the records ri in qi in low dimensional data

15: Use BNN for ri in qi

16: Process the queries qi

17: Identify NN(qi)

18: End

Authenticated multi-step NN is used for query processing and verification across distributed servers. It uses MR tree for indexing purpose for both low and high dimensionality representations. BNN search is used to prove efficiency in non authenticated similarity retrieval. The BNN core which has been obtained and it is integrated with other NN search and describes the clear NN search. The BNN search will produce the reliable query and also it consumes less time to perform the NN search across distributed server.

4. EXPERIMENTAL EVALUATION AND RESULT DISCUSSIONS

The experiments for BNN search is evaluated with the spatial data sets time series, climate condition forecasts obtained from UCI repository. The experiment is implemented in Java 1.6 SDK and core java concept with over 1200 instances of climate conditional forecast dataset. The time taken to perform the NN search using BNN is less when compared to an existing AMNN and SNN. Fig 4.1 describes the execution time taken by queries processing using different search techniques. When an incoming queries increases, the execution time taken by BNN is low when contrast to an existing SNN, AMNN.

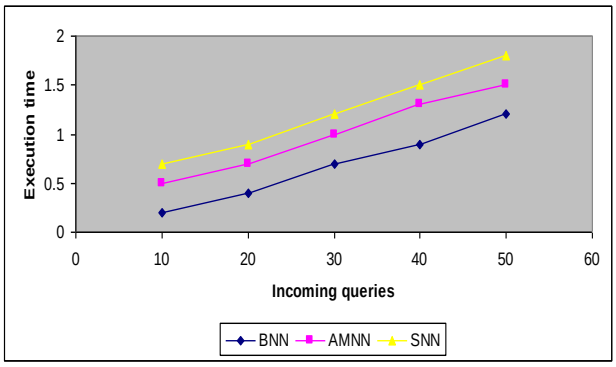

Fig 4.1 Incoming queries Vs Execution Time

Fig 4.2 describes the query result accuracy of query processing using different search techniques. When a query processing rate increases, the query result accuracy time taken by BNN is high when contrast to an existing SNN, AMNN.

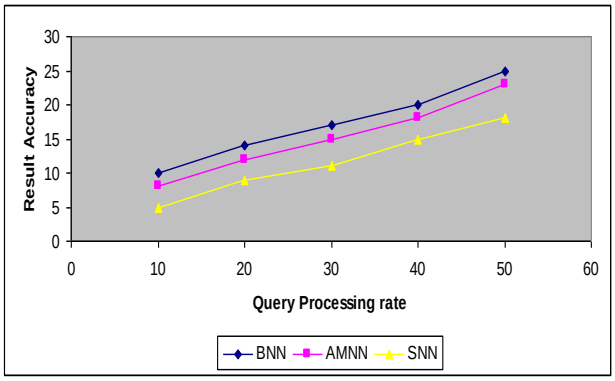

Fig 4.2 Incoming queries rate Vs Result Accuracy

Fig 4.3 describes the performance rate of incoming query processing using different search techniques. When a query processing rate increases, the performance rate of query evaluation by BNN is high when contrast to an existing SNN, AMNN.

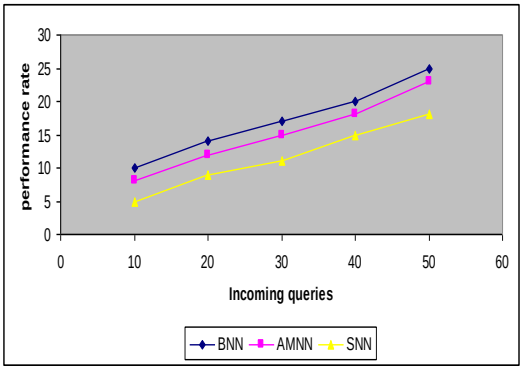

Fig 4.3 Incoming queries rate Vs Performance rate

The performance results showed that the result accuracy and performance rate of NN and similarity search using BNN is high contrast to an existing SNN and AMNN. The outcome of BNN would be reliable data and processed efficiently for high dimensional data structure.

5. CONCLUSION

The Bayesian Nearest Neighbor search in spatial database presented in this work provided an efficient method to identify the nearest neighbor on spatial database without any ambiguity occurred. The communication overhead for BNN is low since it maintained different methods for different dimensionality of data. It achieved a reliable data structure and processed the query efficiently from the different servers. The BNN eliminated the servers' data set which does not contribute the result accuracy. The query verification has been done efficiently in BNN. Finally, to the best of our knowledge, the NN and similarity search using BNN is better than an existing AMNN and SNN.

6. REFERENCES

[1] Sze Man Yuen, Yufei Tao, Xiaokui Xiao, Jian Pei, and Donghui Zhang, " Superseding Nearest Neighbor Search on Uncertain Spatial Databases" IEEE Transaction on Knowledge and Data Engineering VOL. 22, NO. 7, July 2010.

[2] Stavros Papadopoulos, Lixing Wang, Yin Yang, Dimitris Papadias, and Panagiotis Karras, "Authenticated Multistep Nearest Neighbor Search", IEEE Transactions on Knowledge and Data Engineering, vol. 23, no. 5, May 2011.

[3] H.-P. Kriegel, P. Kunath, and M. Renz, "Probabilistic Nearest- Neighbor Query on Uncertain Objects," Proc. Database Systems for Advanced Applications (DASFAA), pp. 337-348, 2007.

[4] Kyriakos Mouratidis and Man Lung Yiu " Anonymous Query Processing "ieee transactions on knowledge and data engineering, vol. 22, no. 1, january 2010.

[5] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh, "Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures," Proc. Int'l Conf. Very Large Data Base Endowment (VLDB '08), vol. 1, pp. 1542-1552, 2008.

[6] A. Kundu and E. Bertino, "Structural Signatures for Tree Data Structures," Proc. Int'l Conf. Very Large Data Base Endowment (VLDB '08), 2008.

[7] H. Pang and K. Mouratidis, "Authenticating the Query Results of Text Search Engines," Proc. Int'l Conf. Very Large Data Base Endowment (VLDB '08), 2008.

[8] Y. Tao, K. Yi, C. Sheng, and P. Kalnis, "Quality and Efficiency in High Dimensional Nearest Neighbor Search," Proc. ACM SIGMOD, 2009.

[9] Y. Yang, S. Papadopoulos, D. Papadias, and G. Kollios, "Spatial Outsourcing for Location-Based Services," Proc. Int'l Conf. Data Eng. (ICDE '08), 2008.

[10] K. Yi, F. Li, M. Hadjieleftheriou, G. Kollios, and D. Srivastava, "Randomized Synopses for Query Assurance on Data Streams," Proc. Int'l Conf. Data Eng. (ICDE '08), 2008.

[11] G. Beskales, M.A. Soliman, and I.F. Ilyas, "Efficient Search for the Top-k Probable Nearest Neighbors in Uncertain Databases," Proc. Very Large Data Bases (VLDB), vol. 1, no. 1, pp. 326-339, 2008.

[12] R. Cheng, J. Chen, M.F. Mokbel, and C.-Y. Chow, "Probabilistic Verifiers: Evaluating Constrained Nearest-Neighbor Queries over Uncertain Data," Proc. Int'l Conf. Data Eng. (ICDE), pp. 973-982, 2008.

[13] M. Hua, J. Pei, W. Zhang, and X. Lin, "Ranking Queries on Uncertain Data: A Probabilistic Threshold Approach," Proc. ACM SIGMOD, pp. 673-686, 2008.

[14] K. Yi, F. Li, G. Kollios, and D. Srivastava, "Efficient Processing of Top-k Queries in Uncertain Databases," Proc. Int'l Conf. Data Eng. (ICDE), pp. 1406-1408, 2008.

ABOUT AUTHORS K.Balasaravanan received the M.E. degree in Computer Science and Engineering from Nandha Engineering college, Perundurai, Erode, Tamilnadu, India. He is pursuing the PhD degree in Anna University of Technology, Coimbatore, Tamilnadu, India. His research interests include Data warehousing, Data mining and Image mining.

Professor, Dr.K.Duraiswamy**,** He is working as a Dean (Academic) at K.S.Rangasamy College of Technology, tiruchengode, Tamilnadu, India. His area of Specification is Computer Architecture and his research interests include in Computer Networks, Internet Computing and security, Digital Image Processing, Data warehousing and Data mining. he published many papers in refereed international and national journals and conferences proceedings.