

An Efficient Bayesian Nearest Neighbor Search Using Marginal Object Weight Ranking Scheme in Spatial Databases

¹K. Balasaravanan and ²K. Duraiswamy

¹Department of CSE,

Velalar College of Engineering and Technology, Erode, Tamil Nadu, India

²Department of CSE, K.S. Rangasamy College of Technology, Namakkal, Tamil Nadu, India

Abstract: Problem statement: A database that is optimized to store and query data that is related to objects in space, including points, lines and polygons is called spatial database. Identifying nearest neighbor object search is a vital part of spatial database. Many nearest neighbor search techniques such as Authenticated Multi-step NN (AMNN), Superseding Nearest Neighbor (SNN) search, Bayesian Nearest Neighbor (BNN) and so on are available. But they had some difficulties while performing NN in uncertain spatial database. AMNN does not process the queries from distributed server and it accesses the queries only from single server. In SNN, the high dimensional data structure could not be used in NN search and it accesses only low dimensional data for NN search. **Approach:** The previous works described the process of NN using SNN with marginal object weight ranking. The downside over the previous work is that the performance is poor when compared to another work which performed NN using BNN. To improve the NN search in spatial databases using BNN, we are going to present a new technique as BNN search using marginal object weight ranking. Based on events occurring in the nearest object, BNN starts its search using MOW. The MOW is done by computing the weight of each NN objects and rank each object based on its frequency and distance of NN object for an efficient NN search in spatial databases. **Results:** Marginal Object Weight (MOW) is introduced to all nearest neighbor object identified using BNN for any relevant query point. It processes the queries from distributed server using MOW. **Conclusion:** The proposed BNN using MOW framework is experimented with real data sets to show the performance improvement with the previous MOW using SNN in terms of execution time, memory consumption and query result accuracy.

Key words: Marginal Object Weight (MOW), Superseding Nearest Neighbor (SNN), Authenticated Multi-step NN (AMNN), Bayesian Nearest Neighbor (BNN)

INTRODUCTION

An emergent trend in spatial databases is, most of the data are uncertain. In spatial databases, an object is computed by probability density function (pdf). Applications which sustain high dimensional data structure will achieve the Nearest Neighbor (NN) and similarity search. To perform nearest neighbor (NN) queries search on uncertain objects, it is necessary to choose the best few NN candidate set. An object is said to be an NN-candidate if it may be the NN. For an instance, assume three objects with its corresponding probabilities and identify the instances of those objects. For NN search in uncertain spatial databases, many NN search techniques have been presented. In fact, when the amount of NN-candidates is large, recurring all those NN candidates to the user is a poor choice.

Therefore, it is vital to choose the preeminent few NN-candidates in an uncertain spatial database.

The multi-step NN structure is provoked by applications that involve exclusive distance computations. Especially, let $DST(B,A)$ be the definite distance between a query B and a data point A that is available in DB. The applicability of the multi-step structure takes a break on the subsistence of a filter distance metric DST, which is inexpensive to assess and suits the lower bounding property. In SNN, an NN object is identified, if it has a nonzero probability. Only the smallest amount of NN-candidates supersedes all the NN-candidates outside the SNN-core. So, the main concern over SNN and AMNN is that it performs the NN and similarity search for the low dimensionality data structures.

To make a query processing more secure in spatial databases, a server sustains a data set DB marked by a

Corresponding Author: Balasaravanan, K., Department of Computer Science and Engineering, Velalar College of Engineering and Technology, Erode, Tamil Nadu, India

trusted authority (e.g., the data owner). The signature SIG is typically supported on a public-key cryptosystem. The server obtains and practices queries from clients. Each query responded as a result set RS_DB that suits definite predicates. In addition, the client must be capable to create that RS is correct, i.e., that it includes all records of DB that assure the query conditions, and that these records have not been adapted by the server or another entity. Because SIG detains the whole DB, the server precedes A Verification Object (VO). After giving VO, only the authorized users can access the records in spatial database.

The contribution of the proposed BNN with MOW described about a ranking method based on marginal weight which is dispersed to the objects at the best nearest ambiguous objects done by BNN in a secure manner. Among the uncertain nearest neighbor object high ranked marginal weighted object is recognized as the best possible to its relevant query point. The implementation of MOW in BNN is tested with greatly intense and lightly intense spatial databases, to shows its performance of obtaining nearest neighbor object at any definite point of query search.

Literature review: In a spatial database, uncertain similarity query processing is an important aspect in data mining. A lot of work has been performed to deal with uncertain (Parker *et al.*, 2010) data addresses probabilistic closest neighbor (NN) queries for definite query objects (Yuen *et al.*, 2010), and for tentative queries (Cheng *et al.*, 2008). Kriegel *et al.* (2007) presented another NN algorithm that uses R- tree for depth-first search. This algorithm consumes less memory than BF, but it does not process many nodes. Normally, resolutions for R-trees, had poor performance in high-dimensional spaces, because the formation of the R-tree weakens considerably as the dimensionality increases. The above solutions imagine that the distance between two objects can be considered rapidly, while (Mouratidis and Papadias, 2007) distance estimation is expensive with NN search using continuous queries. For a spatial queries, (Hu *et al.*, 2011) presented a NN search using Voronoi Neighbors by identifying the best NN object integrity.

Query processing in spatial database for NN search (Papadopoulos *et al.*, 2010) has been described in different aspects. For accessing the uncertain spatial database, querying objects (He *et al.*, 2007) proposed adaptive algorithms to make a decision whether to utilize an key probe or a data examine for each step during the processing of a queries are used. The Nearest Neighbor (NN) can be evaluated using the probability density function in which the NN would ahead of the user-specified threshold to manage the preferred location in a query outcome to diminish the

computational effort (Mouratidis and Yiu, 2010). Papadopoulos *et al.* (2011), the NN has been initiated out using the multi-step processing, it diminishes the network overhead due to the broadcasting of false hits.

Safavi-Naini and Wild (2008), Authentication codes offer message reliability within a symmetric key setting. It believe a strong attack circumstances where the opponent is adaptive and has right to use to authentication and verification of the data objects can be recognized using the Location Based Service (LBS). Here two techniques have been proposed; (1) present unspecified LBS access to the users and (2) permit well-organized query processing at the LBS side. This realizes low computational task and low communication overhead. For an efficient search in uncertain database, (Beskales *et al.*, 2008), presented a NN search using Top-k queries in a given uncertain database.

To improve the NN search object in uncertain spatial databases, a new technique BNN search using MOW ranking scheme is presented.

MATERIALS AND METHODS

The proposed BNN search using MOW ranking in spatial database is framed efficiently to perform NN search in a secure manner by ranking each object based on its proximity of reference points at the spatial events. The proposed BNN search using MOW ranking framework comprises of three operations. The first operation is to identify the objects which are considered as NN to the specified object. The second operation describes the NN search using BNN based on for similarity retrieval. The third operation is to rank each object using MOW obtained through BNN in the spatial database. The architecture diagram of the proposed BNN search using MOW ranking in spatial database is shown in Fig. 1.

The first process is to identify the objects in the spatial database to perform NN and similarity search based on its respective query points. Then perform NN search using BNN. BNN using MOW perform the NN search efficiently for both high and low dimensional data structure by retrieving the information from the distributed servers and eliminate the server which is not involved in the result producing contribution. The process of BNN search for identifying the NN object is explained briefly in Fig. 2. In BNN, if the input data is high dimensional data structure in a query, then the time series is used for identifying the similarity retrieval for NN search. If the input data is low dimensional structure, NN search is performed directly with similarity search retrieval of data in query processing. A sequential scan is used for moderate dimensionality of data. The similarity and NN search is performed to map the record in the database from distributed servers. The query from several servers i.e., distributed server

has been processed with multi-step method. To reveal the lower bounding property, it combine multi- step search with a dimensionality reduction. After identifying the NN object using BNN, ranking scheme MOW is used to rank each NN object based on its proximity of reference points at the spatial events. The object which has the highest rank from the list of objects is identified as the best NN to access the query in spatial database.

Pseudo code used for BNN search: BNN search is done from distributed server. Access the spatial database present in the distributed server. For high dimensional data source, follow the steps described below and identify the best NN object using BNN.

MOW ranking scheme: The marginal weight of object derived from BNN is calculated based on the instance event produced with objects nearest to the reference/querying point in provisions of frequency of object being at the contiguous position and the objects distance to the querying point on the spatial database.

The above pseudo code explains the process of BNN search from distributed server (Fig. 3). Identify and mark the distributed server and access the queries from the servers based on its type of data like high dimensional data, low dimensional data and moderate dimensionality. After identifying the input data, plot the records in the queries to the low dimensional data and processed it based on the mapping. Then identify the nearest neighbor object related to the given dimensional data.

The above pseudo code described the process of MOW ranking scheme for each derived NN object from BNN (Fig. 4). The weight of each derived NN object is identified. Normally, in MOW ranking scheme, the ranking of each object is done based on the frequency and distance of the NN object.

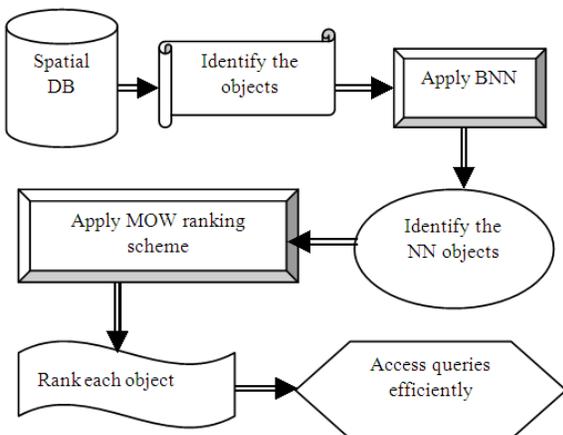


Fig. 1: Architecture Diagram of the proposed

Frequency is computed by how often the object is assigned as NN and the distance is computed by the distance between the NN object to the querying point in spatial database. After computing distance and frequency, the rank of each NN object is compared and identifies the best NN object in spatial database through MOW ranking scheme using BNN.

All the contiguous objects are dispersed with the intended marginal weight and ranking is completed based on the sequential order. With the rank allotted for each objects for NN search, the NN object is recognized using BNN without any ambiguity. The process of ranking scheme is shown in Fig. 5.

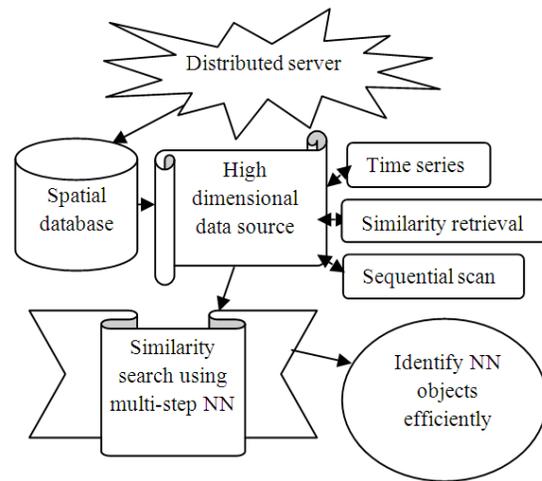


Fig. 2: Process of BNN search

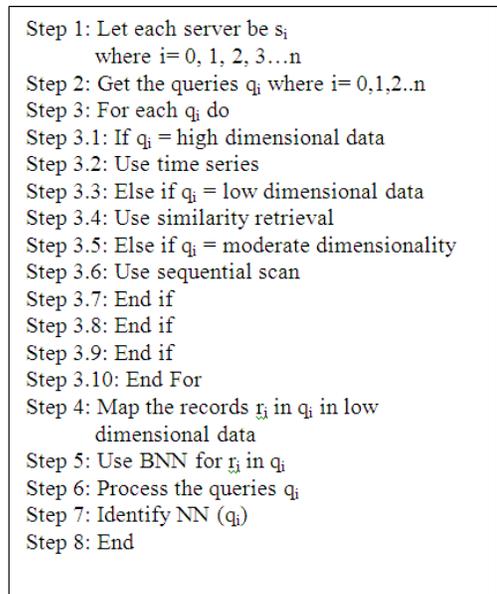


Fig. 3: Pseudo code for the BNN search

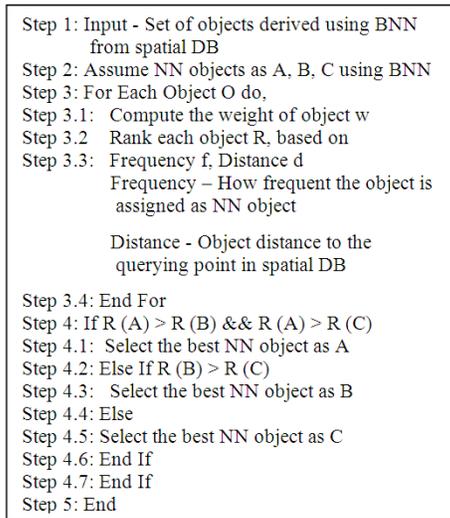


Fig. 4: Pseudo code for MOW ranking scheme

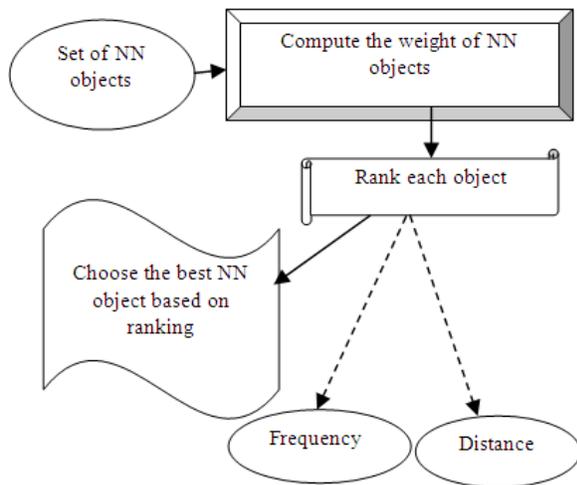


Fig. 5: Process of MOW ranking scheme

RESULTS

The experiments for BNN search using MOW ranking scheme 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. Local disk maintains the data sets. We ran our experiments with various data sets obtained from UCI repository. The proposed BNN search using MOW ranking scheme efficiently designed for NN search using BNN and assigned rank to each and every object, based on MOW ranking scheme, the best NN object is chosen. Based on frequency and distance of object, MOW is assigned.

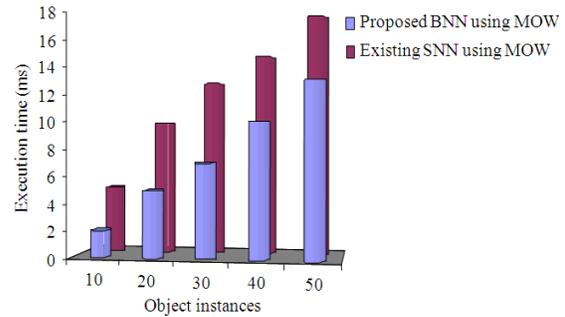


Fig. 6: Object instances Vs execution time

The MOW evaluated for the nearest objects of the reference points are points out its simplicity of the nearest neighbor using BNN amongst all the accessible objects attributes in the Climate Forecast spatial data sets. The performance of the proposed BNN search using MOW ranking scheme for identifying the best NN object based on is measured in terms of Execution time, Memory consumption and Query result accuracy.

Execution time: The time taken to search the NN object using BNN from a list of objects in a spatial database from distributed server is termed as execution time. Figure 6 described the execution time consumed to search NN object from spatial database. In the proposed BNN search using MOW ranking scheme, the process of identifying the NN object from spatial database consumed less time compared to an existing SNN using MOW.

Comparison result of BNN using MOW with an existing SNN using MOW based on execution time consumption variance, measured in terms of milliseconds (m/s). The performance graph of the proposed BNN search using MOW ranking scheme for identifying the NN efficiency is shown in the Fig. 6. The variance in the execution time consumption would be 12-20% low in the proposed BNN search using MOW ranking scheme.

Memory consumption: The memory consumption rate is the rate at which memory is consumed for identifying the best NN object. Figure 7, described the memory consumption of object instances from spatial database. In the proposed BNN search using MOW ranking scheme, the procedure of identifying the NN object and processing the best NN object query from spatial database consumed less memory compared to an existing SNN using MOW. Comparison result of BNN using MOW with an existing SNN using MOW based on memory consumption variance, measured in terms of Kilo Byte (KB).

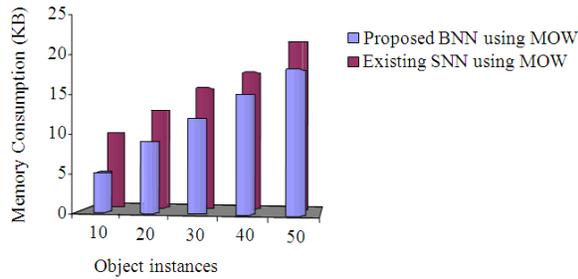


Fig. 7: Object instances Vs Memory

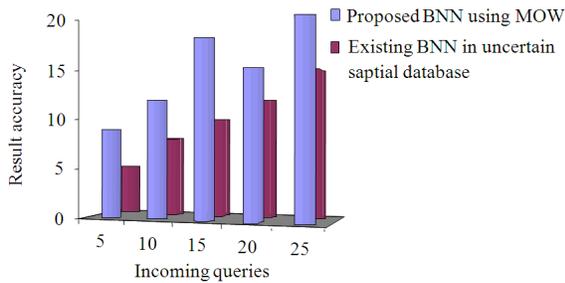


Fig. 8: Incoming queries Vs result accuracy

The performance graph of the proposed BNN search using MOW ranking scheme for memory consumption rate is shown in the Fig. 7. The variance in the execution time consumption would be 15-20% low in the proposed BNN search using MOW ranking scheme.

Query result accuracy: The query result accuracy is the outcome of the processed query after identifying the NN object assigned as the highest rank from a list of objects and process the query requested by that NN object.

Figure 8, described accuracy of the outcome of the processed query from spatial database. In the proposed BNN search using MOW ranking scheme, the result accuracy for processed queries is high contrast to an existing SNN using MOW. The result accuracy of the proposed BNN search using MOW ranking scheme for processing the queries is shown in the Fig. 7. The variance in the accuracy would be 20-25% high in the proposed BNN search using MOW ranking scheme.

DISCUSSION

In this study, we have seen how the nearest neighbor object is efficiently identified based on ranking scheme using the proposed BNN search using MOW ranking scheme with an existing SNN search using MOW and BNN search in spatial database written in mainstream languages such as Java. We used different sets of semantic data for comparing the results

of the proposed BNN search using MOW ranking scheme with an existing SNN search using MOW and BNN search in spatial database.

Finally, it is observed that the proposed BNN search using MOW ranking scheme efficiently achieved the NN object search efficiently in a less memory time, even in a less memory consumption. Based on MOW ranking scheme, each object is ranked depended on its proximity of reference points at the spatial events.

CONCLUSION

In this study, we efficiently performed the NN search in spatial database from distributed server by professionally introducing the proposed BNN search using MOW ranking scheme. The proposed BNN search rank the objects in spatial database accurately under a limited memory and it consumed less time to achieve the NN search using BNN. The communication overhead for BNN using MOW is low since it sustained diverse methods for diverse dimensionality of data. It reached a consistent data structure and practiced the query efficiently from the diverse servers. The proposed BNN using MOW eradicated the servers' data set which does not supply the result accuracy. The query verification and accuracy has been done efficiently in BNN. The experimental results showed that the proposed BNN search using MOW ranking scheme worked efficiently by improving the reliability of the NN search with less overhead and less execution time.

REFERENCES

Beskales, G., M.A. Soliman and I.F. Ilyas, 2008. Efficient search for the top-k probable nearest neighbors in uncertain databases Proc. VLDB Endowment, 1: 326-339.

Cheng, R., J. Chen, M.F. Mokbel and C.Y. Chow, 2008. Probabilistic verifiers: Evaluating constrained nearest-neighbor queries over uncertain data. Proceedings of the 24th International Conference on Data Engineering, Apr. 7-12, IEEE Xplore Press, Cancun, pp: 973-982. DOI: 10.1109/ICDE.2008.4497506

He, B., Q. Luo and B. Choi, 2007. Bayesian networks for knowledge-based authentication. IEEE Trans. Knowl. Data Eng., 19: 695-710. DOI: 10.1109/TKDE.2007.1024

Hu, L., W. Ku, S. Bakiras and C. Shahabi, 2011. Spatial query integrity with voronoi neighbors. IEEE Trans. Knowl. Data Eng. DOI: 10.1109/TKDE.2011.267

Kriegel, H.P., P. Kunath and M. Renz, 2007. Probabilistic nearest-neighbor query on uncertain objects. Adv. Database: Concepts Syst. Appli., 4443: 337-348. DOI: 10.1007/978-3-540-71703-4_30

- Mouratidis, K. and D. Papadias, 2007. Continuous nearest neighbor queries over sliding windows. *IEEE Trans. Knowl. Data Eng.*, 19: 1-15.
- Mouratidis, K. and M.L. Yiu, 2010. Anonymous query processing in road networks. *IEEE Trans. Knowl. Data Eng.*, 22: 2-15. DOI: 10.1109/TKDE.2009.48
- Papadopoulos, S., L. Wang, Y. Yang, D. Papadias and P. Karras, 2011. Authenticated multistep nearest neighbor search. *IEEE Trans. Knowl. Data Eng.*, 23: 641-654. DOI: 10.1109/TKDE.2010.157
- Papadopoulos, S., L. Wang, Y. Yang, D. Papadias and P. Karras, 2010. Authenticated multistep nearest neighbor search. *IEEE Trans. Knowl. Data Eng.*, 23: 641-654. DOI: 10.1109/TKDE.2010.157
- Parker, D.R., S.C. Gustafson, M.E. Oxley and T.D. Ross, 2010. Development of a bayesian framework for determining uncertainty in receiver operating characteristic curve estimates. *IEEE Trans. Knowl. Data Eng.*, 22: 31-45. DOI: 10.1109/TKDE.2009.50
- Safavi-Naini, R. and P.R. Wild, 2008. Information theoretic bounds on authentication systems in query model. *IEEE Trans. Inform. Theory*, 54: 2426-2436. DOI: 10.1109/TIT.2008.921683
- Yuen, S.M., Y. Tao, X. Xiao, J. Pei and D. Zhang, 2010. Superseding nearest neighbor search on uncertain spatial databases. *IEEE Trans. Knowl. Data Eng.*, 22: 1041-1055. DOI: 10.1109/TKDE.2009.137