Marginal Object Weight Ranking for Nearest Neighbor Search in Spatial databases

K.Balasaravanan [1], Dr.K.Duraiswamy [2]

**[1] Research Scholar and Assistant Professor (Senior Grade),
Department Of Computer Science and Engineering,
Velalar College of Engineering and Technology,
Erode - 638012, Tamil Nadu, India.**

**[2] Dean(Academic) and Research Supervisor,
K.S.Rangasamy College of Technology,
Tiruchengode-637215, TamilNadu, India.**

Abstract

Uncertainty prevails in spatial databases whose objects are defined as multidimensional probable density function. Identifying an object is tedious task in uncertain spatial databases. Recent work on searching objects and its closeness to other objects was presented as superseding nearest neighbor search technique. In superseding neighbor search, one of all other objects in the spatial database, which was more close to the querying point is called as superseding nearest neighbor object. However there are cases where no object was unavailable to supersede to all other nearest neighbor object, ambiguity arises.

To overcomes the situation of ambiguity, in this paper Marginal Object Weight(MOW) is introduced to all nearest neighbor object for any respective query point, from there on most highly ranked object is identified as superseding nearest neighbor object. The performance of MOW ranking on nearest neighbor search is carried out for optimal amount of query point in the spatial databases of real and synthetic dataset. The superseding nearest neighbor object obtained with MOW ranking scheme identify more optimal to the query points compared to that of unranked superseding nearest neighbor object search scheme.

Keywords: *Spatial Databases, Nearest Neighbor Search, Superseding, Marginal Object Weight*

1. Introduction

With growing trend in spatial databases ([2], [3]) most of the data are uncertain about its where about. The data objects position are evaluated with the nearest neighbor search and evaluated in terms of probability density function. The identification of uncertain objects as instance reference was referred in most of recent works [4], [5], [6]. In spatial databases no object is considered to be nearest neighbor (NN) for the querying point in any instances, only comparative nearest position is only certain that to at any specific spatial event.

We consider nearest neighbor (NN) queries on uncertain objects. In general, there may not exist any object that is guaranteed to be the NN. We say that an object is an NN-candidate if it may be the NN. Assume that the query point is q. A is an NN-candidate. However, D is not an NN-candidate, as its distance to q is larger than that of C. Apparently, when the number of NN-candidates is large, returning all of them to the user is a poor choice. Hence, it is important to select the best few NN-candidates.

The uncertain nature of objects nearest neighbor query where about's were handled in the recent past with superseding nearest neighbor search [1] scheme. In superseding nearest neighbor, the object of NN was ambiguous at points where the query made comprises of more frequently varying densed object position. The multidimensional nature of the object creates further confusion in identifying even the comparative nearness to the query point. The ambiguous nearest neighbor search for object to any query point in the spatial databases need to be overcame.

The proposal in this work, briefs about a ranking scheme based on marginal weight being assigned to the objects at the optimal nearest ambiguous objects. Among the ambiguous nearest neighbor object high ranked marginal weighted object is identified as more optimal to its respective query point. The implementation of MOW is tested with highly dense and sparsely dense spatial databases, to shows its performance of obtaining optimal superseding nearest neighbor object at any specific point of query search.

2. Related Works

Uncertain similarity query processing has focused on various aspects. A lot of existing work dealing with uncertain data addresses probabilistic nearest neighbor (NN) queries for certain query objects [1], [8] and for uncertain queries [7]. To reduce computational effort, [9] add threshold constraints in order to retrieve only objects whose probability of being the nearest neighbor exceeds a user-specified threshold to control the desired confidence required in a query answer.

In [10] introduced two frameworks for the spatial network kNN query: the incremental euclidean restriction (IER) and incremental network expansion (INE). IER applies the property that the Euclidean distance between any two network nodes is a lower bound of their network distance to prune the search space. INE performs network expansion similar to the Dijkstra’s algorithm [9] from the query point and examines data objects in the order they are encountered. They showed that INE performs better than IER in general. As an optimization of IER, [8] proposed incremental lower bound constraint (LBC). The LBC method calculates distance lower bounds of objects for pruning purposes. Hence, the workload from network distance calculation is greatly reduced.

In [5] propose another NN algorithm that performs depth-first search on an R-tree. This algorithm requires less memory than BF, but may need to access more nodes. Solutions based on R-trees, however, have poor performance in high-dimensional spaces [9], because the structure of the R-tree deteriorates significantly as the dimensionality increases. This observation leads to several algorithms specifically designed for high-dimensional NN search (see [10] and the references therein). The above solutions assume that the distance between two objects can be calculated quickly, whereas [6] consider the case where distance evaluation is expensive.Finally, it is worth mentioning that NN retrieval has numerous variations such as reverse NN search [2], aggregate NN search [3], continuous NN search [5], etc. In [4] address a different version of uncertain NN search. Specifically, they assume existentially uncertain objects. Namely, an object may not belong to the database, but in case it does, its location is precise. In our context, an object definitely exists, but its location is uncertain. The solution of [8] is specific to its settings, and cannot be adapted to our problem.

An NN query can be regarded as an instance of top-1 search. If we define the score of an object o as its distance to the query point q, then the goal is to find the top-1 object with the lowest score. This creates the opportunity of applying top-k methods to NN queries. Several top-k algorithms [9], [7], [10] have been proposed for uncertain data. At k ¼ 1, they extract the object that has the smallest score with the largest probability. In other words, they advocate the same result as the PR-principle.

3. Superseding Nearest Neighbor Search Scheme

In Superseding Nearest Neighbor Search Scheme initial assumption made are briefed to explain the nature and function of the scheme. Initial with a set of uncertain objects, each object is associated with a specified set of points. The specified set of points is referring as instances to its associated object and its probability of occurrence is evaluated. The discrete probability density function considers every object contains multiple numbers of instances. Each instance is associated with multiple probability ratios. The PR-principle is a reasonable way to define the results of NN queries on uncertain data. A common criticism is that sometimes even the highest NN probability can be quite low, and multiple objects may have almost the same NN probabilities. In any case, the PR-principle is orthogonal to our SNN approach. As will be shown in the experiments for most queries, the SNN-core contains only a single object that is not the object with the greatest NN-probability.

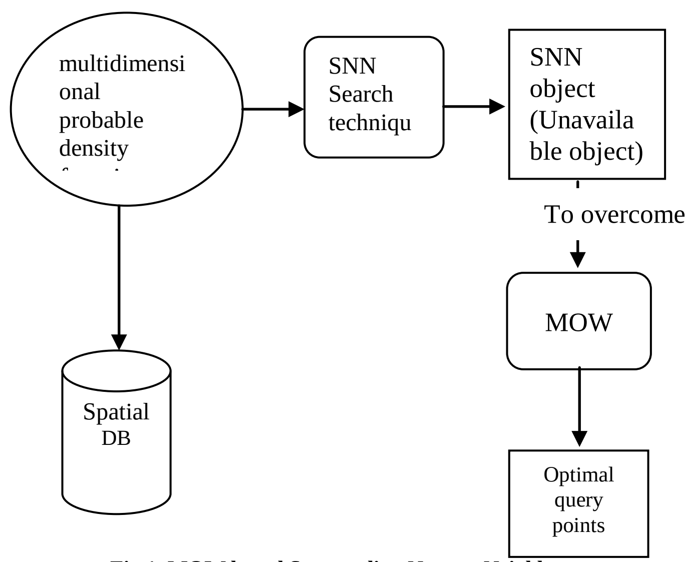


Fig 1. MOW based Superseding Nearest Neighbor

In practice, the PR-principle and our SNN method are nice complements to each other. First, they provide two interesting options to a user, each with its unique features. Second, they can even be combined to provide more reliable results. For example, a user may want to find objects that 1) objects are in the SNN-core, and 2) their NN-probabilities are among the top-t in the data set, where t is a user parameter.

SNN search can be applied in any application where it makes sense to issue NN queries on uncertain objects. In particular, we show that the SNN-core is always unique, thus eliminating the question of which SNN-core would be better if there were several. Our second contribution is a set of algorithms for finding SNN-cores. These algorithms utilize a conventional R-tree (commonly available in commercial DBMS) to drastically prune the search space to achieve I/O efficiency.

The probable value of minimum to maximum is used to identify objects with to its query points. The ranking instance list sorts the objects in chronological order of their distance to the query of the spatial data set position. These distance based nearest neighbor search are carried out in the superseding scheme. The superseding nearest neighbor search scheme, identify the comparative superseded object among the nearest neighbor objects of distance metrics based query point in the spatial data events.

The superseded objects are the ones which have closest possible neighbor among the nearest ones in the required instance of any specific reference point. The more specific reference point identifies the exact object certainty. However at some instance closest proximity of all available objects for the reference point may be ambiguous due to varying distance of objects relativity. These motivate us to introduce Marginal Object Weight ranking model to introduce to the superseding nearest neighbor scheme, explained briefly below.

4. Marginal Object Weight Ranking

.In the spatial databases of uncertain objects, the nearest object to the querying point needs to be identified at its optimality to present a discrete probability of objects. The superseding nearest neighbor object search creates few ambiguity on identification of instance points at which specific spatial object is closest than its other entire neighbor to the referring point. In order to provide a clear demarcation to the object at an instance, marginal weight is introduced to all the nearest neighbor objects of the reference points. The marginal weight assigned to each object is based on its proximity of reference points at the spatial events.

The marginal weight of object keeps on changing with the multiple dimensional of the spatial events. The marginal weight is calculated based on the instance event generated with objects nearest to the reference / querying point in terms of frequency of object being at the closest position and the objects distance to the querying point on the spatial database. All the closest objects are assigned with the calculated marginal weight and ranking is made based on the chronological order. With the rank assigned for each objects for nearest neighbor search, the superseding object is identified for any specific spatial instances to its optimality without any ambiguity.

5. Performance of MOW based Superseding nearest Neighbor

The experiments are conducted for the MOW based superseding nearest neighbor with 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 data set. The MOW calculated for the closest objects of the reference points are clearly indicates its clarity of the nearest neighbor of superseding nature among all the available objects attributes in the Climate Forecast spatial data sets. In addition the discrete probability density function associated to the reference points provides better nearest neighbor search rate of identifying the superseding object for the given spatial climatic events to forecasts its temperature and humidity conditions.

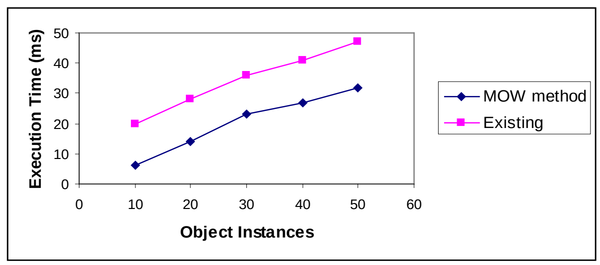


Fig 2. Object Instances Vs Execution Time

The above figure shows that the efficiency of our proposed model MOW method. It shows the Execution time of our proposed method is low compared to that of existing method.

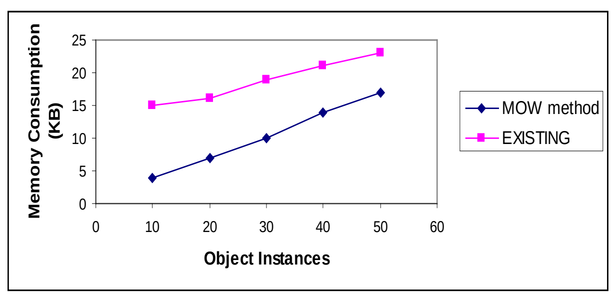


Fig 3. Object Instances Vs Memory Consumption

Figure 3 shows that the efficiency of our proposed model MOW method. It shows the effective Memory consumption.

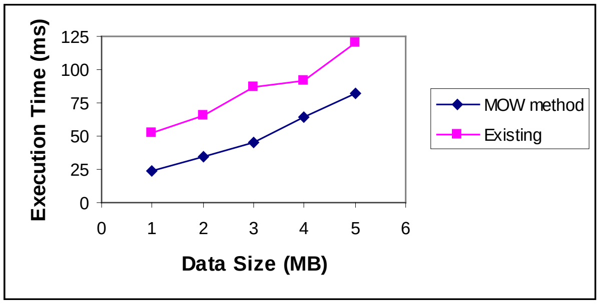


Fig 4. Data size Vs Execution Time

The above figure 4 shows that the efficiency of our proposed model MOW method. Compared to existing method, MOW method having the low execution time.

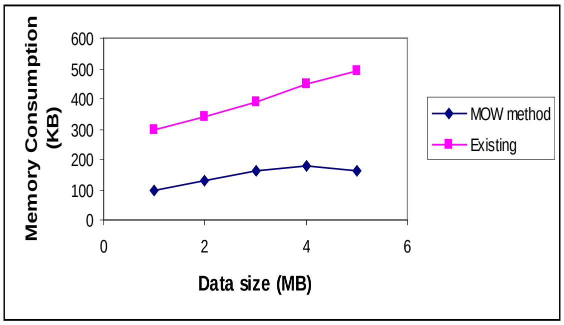


Fig 5. Data size Vs Memory Consumption

The above figure shows Data size Vs Memory Consumption. Our proposed model, MOW method having efficient memory consumption.

6. Conclusion

The marginal object weight ranking scheme presented in this paper improved superseding nearest neighbor search on uncertain objects in the spatial database obtain an optimal nearest neighbor object without any ambiguity among the closer objects of the reference points. The experiments shows that most of the query points are identified with better clarified optimal nearest neighbor object to make certainty on the uncertain object positions in the climate forecast data set instances extracted from UCI Repository. The rate of nearest neighbor search is improved with the marginal weights calculated and assigned to the objects of closer proximity to the reference / query points. The weight assigned to the object is ranked to obtain its closest positional points at multiple instances, by which the uncertainty of the objects are almost reduced to zero.

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 EngineeringVOL.22, NO.7,July 2010.

[2] R. Cheng, D.V. Kalashnikov, and S. Prabhakar, "Querying Imprecise Data in Moving Object nvironments," IEEE Transaction on Knowledge and Data Eng., vol. 16, no. 9, pp. 1112-1127, Sept. 2004.

[3] X. Dai, M.L. Yiu, N. Mamoulis, Y. Tao, and M. Vaitis, "Probabilistic Spatial Queries on Existentially Uncertain Data," Proc. Symp. Advances in Spatial and Temporal Databases (SSTD), pp. 400-417, 2005

[4] 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.

[5] 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

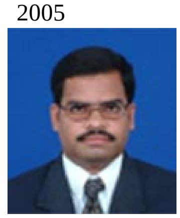
[6] J. Pei, B. Jiang, X. Lin, and Y. Yuan, "Probabilistic Skylines on Uncertain Data," Proc. Very Large Data Bases (VLDB), pp. 15-26, 2007.

[7] T.H. Cormen, C.E. Leiserson, R.L. Rivest, and C. Stein, Introduction to Algorithms. MIT Press, 2001.

[8] G.R. Hjaltason and H. Samet, "Distance Browsing in Spatial Databases," ACM Trans. Database Systems, vol. 24, no. 2, pp. 265- 318, 1999

[9] H.V. Jagadish, B.C. Ooi, K.-L. Tan, C. Yu, and R. Zhang, "idistance: An Adaptive bþ-Tree Based Indexing Method for Nearest Neighbor Search," ACM Trans. Database Systems, vol. 30, no. 2, pp. 364-397, 2005.

[10] D. Papadias, Y. Tao, K. Mouratidis, and C.K. Hui, "Aggregate Nearest Neighbor Queries in Spatial Databases," ACM Trans. Database Systems, vol. 30, no. 2, pp. 529-576, 2005



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,VOIP, Internet Computing and security, Digital Image Processing, Computer Graphics, Data warehousing and Data mining. Dr.K.Duraiswamy has published many papers in refereed international and national journals and conferences proceedings.