ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 3, September 2012

Clustering and Indexing For Uncertain Objects Using Pruning Techniques of Voronoi Diagram and R-Trees

[1]Ravindra Changala, [2]Kareemunnisa [3]Thaduri Venkata Ramana, [4]Y Sowjanya, [5]K Chandrika
[1, 2] IT Dept, Guru Nanak Institutions Technical Campus, School of Engineering & Technology
[3]Professor&HOD, CSE Department, SLC's Institute of Engineering and Technology
[4,5]**M.Tech** Scholars, SLC's Institute of Engineering and Technology, Hyderabad
changalaravindra@gmail.com

Abstract: - Clustering Uncertainty data is the major issue of the research now a day. The k-means algorithm handle uncertain objects, is very inefficient by probability distribution functions with more expected distances. In paper gives few pruning techniques based on Voronoi diagrams to reduce expected distance calculation and the R-trees index organize the uncertain objects to reduce pruning complexities. This approach gives solution for clustering of the uncertain objects.

Key Words: **K-Means, Uncertain Objects, PDF, Pruning, Voronoi Diagrams, R-Trees.**

I. INTRODUCTION

Clustering is a technique that has been widely studied and applied to many real-life applications. Many efficient algorithms, including the well known and widely applied k-means algorithm, have been devised to solve the clustering problem efficiently. Traditionally, clustering algorithms deal with a set of objects whose positions are accurately known. The goal is to find a way to divide objects into clusters so that the total distance of the objects to their assigned cluster centers is minimized. The data gathered from the different sources like satellite, sensor networks and many more application were get the uncertainty[1] in the data. The uncertainty may because of the precision of the device with which the readings are taken. Clustering or classification of the normal data with any algorithm like k-means is simple, but if the data is uncertain then it is quite difficult to cluster or classify the data. The problem model does not address situations where object locations are uncertain. Data uncertainty, however, arises naturally and often inherently in many applications.

[1]*Lacking or indicating lack of confidence or assurance, "uncertain of his convictions"; "moving with uncertain steps"; "an uncertain smile"; "touched the ornaments with uncertain fingers".*

For examples, consider the application of clustering a set of mobile devices [1]. By grouping mobile devices into clusters, a leader can be elected for each cluster, which can then coordinate the work within its cluster. For example, a cluster leader may collect data from its cluster's members, process the data, and send the data to a central server via an access point in batch [2], [3]. In this way, local communication within a cluster only requires short-ranged signals, for which a higher bandwidth is available. Long-ranged communication between the leaders and the mobile network only takes place in the form of batch communication. This results in better bandwidth utilization and energy conservation. Traditional clustering methods were designed to handle point-valued data and thus cannot cope with data uncertainty. One possible way to handle data uncertainty is to first transform uncertain data into point-valued data by selecting a representative point for each object before applying a traditional clustering algorithm. For example, the centroid of an object's pdf can be used as such a representative point. However, in [5], it is shown that considering object pdf's gives better clustering results than the centroid method. In this paper we concentrate on the problem of clustering objects with location uncertainty. So here we study the voronoi diagram based approach which will solve the problem uncertain data with minimum number of expected distances while adding the object to the specific cluster.

Table 1. Example of Uncertain Data

Certain Data	**Uncertain Data**
Age of Ram is 23 years	Age of Ram is in between [20-30]
Temperature of Pune at morning was 120F	Temperature of Pune at morning was 120±10F
Salary of Ram is 6lacs/pa	Salary of Ram is 5-7lacas/pa
Can't play cricket today, because the temperature is 100 degree Celsius.	Can/cant play cricket because the temp. is 50-90 degree Celsius.
Annual income is 110	Annual income is [90-120]
Season is rainy	Season is[rainy, summer, winter]

II. RELATED WORK

The uncertainty of the data firstly represented with the help of the algorithm in [2] the major challenging was to handle the uncertainty of the data. To handle uncertainty of the data the PDF has important role everywhere. The most traditional clustering methods aims to find a unique factor which will form a cluster for the same object with existing or with totally new cluster by minimizing the sum of squared error(SSE)[2] in case of the k-mean algorithm. The uncertainty in the data mining field, the tuples value is not only the important factor but also the presence of the uncertainty any where is the major

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 3, September 2012

concern of the field. The data having less uncertainty is more important than having the more uncertainty. There are two different types of the uncertainties, existential uncertainty and value uncertainty. The existential uncertainty occurs when it is uncertain whether an object or a data tuple exists there with uncertainty. The tuples in the relational database could be associated with the probabilistic data; the probability represents the existence of the object there in database [1]. The value uncertainty is the one, where the data is in existence but its value is not known precisely [1]. A data item with the value uncertainty is usually represented by the PDF functions over the bounded finite field of the possible values. Getting the valuable results from the uncertain data, till there are different attempts made for the same in terms of the UK-Mean algorithms [3]. There has been growing interest in uncertain data mining. In [5], the well-known k-means clustering algorithm is extended to the UK-means algorithm for clustering uncertain data. In that study, it is empirically shown that clustering results are improved if data uncertainty is taken into account during the clustering process. As we have explained, data uncertainty is usually captured by pdf's, which are generally represented by sets of sample values. Mining uncertain data is therefore computationally costly due to information explosion (sets of samples vs. singular values). To improve the performance of UK-means, CK-means [18] introduced a novel method for computing the EDs efficiently. However, that method only works for a specific form of distance function. For general distance functions, [6] takes the approach of pruning, and proposed pruning techniques such as min-max-dist pruning. In this paper, we follow the pruning approach and propose new pruning techniques that are significantly more powerful than those proposed in [6]. In [19], guaranteed approximation algorithms have been proposed for clustering uncertain data using k-means, k-median as well as k-centre. Clustering of uncertain data is also related to fuzzy clustering, which has long been studied in fuzzy logic. In fuzzy clustering, a cluster is represented by a fuzzy subset of objects. Each object has a "degree of belongingness" with respect to each cluster. The fuzzy c-means algorithm is one of the most widely used fuzzy clustering methods. Different fuzzy clustering methods have been applied on normal or fuzzy data to produce fuzzy clusters. A major difference between the clustering problem studied in this paper and fuzzy clustering is that we focus on hard clustering, for which each object belongs to exactly one cluster. Our formulation targets for applications such as mobile device clustering, in which each device should report its location to exactly one cluster leader. Voronoi diagram is a well-known geometric structure in computational geometry. It has also been applied to clustering. For example, Voronoi trees [7] have been proposed to answer Reverse Nearest Neighbour (RNN) queries. Given a set of data points and a query point q, the RNN problem [29] is to find all the data points whose nearest neighbour is q. The TPL algorithms proposed in [30] uses more advanced pruning techniques to solve this problem efficiently. An R-tree is a self-balancing tree structure that resembles a B+-tree, except that it is devised for indexing multi-dimensional data points to facilitate proximity-based searching, such as k-nearest neighbour (kNN) queries. Rtrees are well studied and widely used in real applications. They are also available in many RDBMS products such as SQLite, MySQL and Oracle. An R-tree conceptually groups the underlying points hierarchically and records the minimum bounding rectangle (MBR) of each group to facilitate answering of spatial queries. While most existing works on Rtree concentrate on optimizing the tree for answering spatial queries, we use R-trees in this paper in a quite innovative way: We exploit the hierarchical grouping of the objects organized by an R-tree to help us check pruning criteria in batch, thereby avoiding redundant checking.

III. UNCERTAINTY MODEL

In this section we will concentrate on the uncertain model for numerical data only. Which is the most common uncertain attribute type occurring in the data mining field. When the value of the numerical attribute is uncertain then attribute is called as uncertain numerical attribute (UNA)[4]. These uncertain numerical values represent the range of the values specified in the above table, table1. By applying the Probability Distribution Function (PDF) over the range specified in the tuples. Continuous probability Distribution function is applied on the range of the data items. It generates the probability of the single attribute specified in the dataset. To handle the uncertainty of the range valued numerical data, we apply the continuous PDF on it so that it gives the probability of occurring the tuple in the data set. The goal of clustering is to group the same data items into the single cluster. So that the expected distance with cluster representative will be minimized. The expected distance for an object Oi with the cluster representative is represented as

$$E\left(\sum_{j=1}^{k}\sum_{i\in C_j}\left\|\mathbf{c}_j-\mathbf{x}_i\right\|^2\right)=\sum_{j=1}^{k}\sum_{i\in C_j}\int\left\|\mathbf{c}_j-\mathbf{x}_i\right\|^2 f(\mathbf{x}_i)d\mathbf{x}_i$$

This is computing the expected distance of the object to be clustered with respect to the pdf value of the uncertain numerical attribute. Once the expected distance calculation are over the next task is to have the clustering of the object. The next section discusses the clustering process with voronoi diagram for uncertain numerical attributes and pruning techniques with voronoi and pdfs are associated with the nodes of the R- tree.

IV. CLUSTERING MODEL

The Figure shows the clustering process to be carried out. The block diagram explains about the clustering process how that is carried out. Uncertain numerical data is supplied to the PDF generator algorithm, which

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 3, September 2012

generates the PDFS for all the records those who falls in valued uncertainty. As shown in the above figure the generated PDFs are bulk loaded into the R-Tree, which goes through the different iterations and fix the final root node for the tree. all the PDFs are stored into the leaf node of the tree and all other intermediate nodes including the root node represents some Bounding box called as Minimum Bounding Rectangle (MBR), which holds more than one records from the given dataset. As the process of MBR generation finishes our next starts called as pruning process with the help of the VORONOI diagram. This process fixes the centroid from the given dataset only and passes over reading the MBRs. If that MBR completely lies inside one the cell then there is no need to find out EDs for all the data items present into the same MBR that is the biggest advantage of using the VORONOI diagrams for clustering process. So that when we compare the UK- mean algorithm do the clustering of uncertain data.

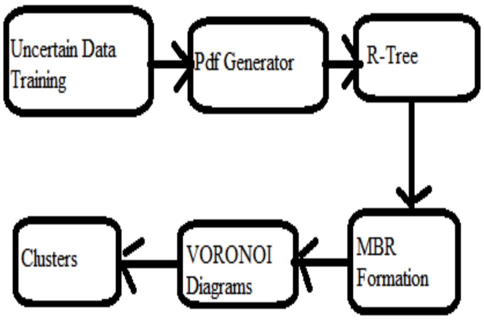


Fig 1. Uncertain Data Clustering Process

V. ALGORITHMS

We first give a short description of the UK-means algorithm [5] and existing pruning techniques [6] that improve UK-means. Then, we present our new pruning techniques that are based on Voronoi diagrams and finally, we introduce our performance booster based on R-trees.

A. MinMax Pruning

Algorithm 1 UK-means

Choose k arbitrary points as c_j $(j = 1, \ldots, k)$
repeat
  for all $o_i \in O$ **do** */*assign objects to clusters*/*
    for all $c_j \in C$ **do**
      Compute $\mathrm{ED}(o_i, c_j)$
    $h(i) \leftarrow \arg\min_{j:c_j \in C}\{\mathrm{ED}(o_i, c_j)\}$
  for all $j = 1, \ldots, k$ **do** */*readjust cluster representatives*/*
    $c_j \leftarrow$ centroid of $\{o_i \in O \mid h(i) = j\}$
until C and h become stable

UK-means (see Algorithm 1) [5] is an adaptation of the well known k-means algorithm to handle data objects with uncertain locations. Initially, k arbitrary points c1. . . ck are chosen as the cluster representatives. Then, UK-means repeats the following steps until the result converges. First, for each object oi, ED(oi, cj) is computed for all cj ∈ C. Object oi is then assigned to cluster cj_ that minimizes ED, i.e., h(i) ← j∗. Next, each cluster representative cj is recomputed as the centroid of all oi's that are assigned to cluster j. The two steps are repeated until the solution C = {c1, . . . , ck} and h(•) converge. The UK-means algorithm is inefficient. This is because UKmeans computes ED for *every* object-cluster pair in *every* iteration. So, given n objects and k clusters, UK-means computes nk EDs in *each iteration*. The computation of an ED involves numerically integrating a function that involves an object's pdf. In practice, a pdf is represented by a probability distribution matrix, with each element of the matrix representing a sample point in an MBR. To accurately represent a pdf, a large number of sample points are needed. The computation cost of integration is thus high.

B. MinMax Pruning

Several pruning techniques that are based on bounds on ED have been proposed in [6]. In the MinMax approach, for an object oi and a cluster representative cj, certain points in MBRi are geometrically determined. The distances from those points to cj are computed to establish bounds on ED. Formally, we define

$$\begin{aligned} \mathrm{MinD}(o_i, c_j) &= \min_{x \in \mathbf{MBR}_i} d(x, c_j) \\ \mathrm{MaxD}(o_i, c_j) &= \max_{x \in \mathbf{MBR}_i} d(x, c_j) \\ \mathrm{MinMaxD}(o_i) &= \min_{c_j \in C}\{\mathrm{MaxD}(o_i, c_j)\} \end{aligned}$$

It should be obvious that MinD(oi, cj) ≤ ED(oi, cj) ≤MaxD(oi, cj). Then, if MinD(oi, cp) > MaxD(oi, cq) for some cluster representatives cp and cq, we can deduce that ED(oi, cp) > ED(oi, cq) without computing the exact values of the EDs. So, object oi will not be assigned to cluster p (since there is another cluster q that gives a smaller expected distance from object oi). We can thus prune away cluster p *without* bothering to compute ED (oi, cp). As an optimization, we can prune away cluster p if MinD(oi, cp) > MinMaxD(oi). This gives rise to the MinMax-BB (bounding box) pruning algorithm (Algorithm 2).

Algorithm 2 MinMax-BB Pruning

for all $c_j \in C$ **do** */*for a fixed object o_i*/*
  Compute $\mathrm{MinD}(o_i, c_j)$ and $\mathrm{MaxD}(o_i, c_j)$.
Compute $\mathrm{MinMaxD}(o_i)$.
for all $c_j \in C$ **do**
  if $\mathrm{MinD}(o_i, c_j) > \mathrm{MinMaxD}(o_i)$ **then**
    Remove c_j from Q_i

C. Pruning with Voronoi Diagram

MinMax-based pruning techniques improve the performance of UK-means significantly by making use of

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 3, September 2012

efficiently evaluable bounds on ED to avoid many ED computations. However, these techniques do not consider the geometric structure of Rm or the spatial relationships among the cluster representatives. One important innovation of this paper is the introduction of Voronoi diagrams [7] as a method to exploit the spatial relationships among the cluster representatives to achieve a very effective pruning. We will show in this section that the Voronoi-diagram-based pruning techniques are theoretically strictly stronger than MinMax-BB. We start with a definition of Voronoi diagram and a brief discussion of its properties. Given a set of points C = {c1, . . . , ck}, the Voronoi diagram divides the space Rm into k cells V (cj) with the following property:

$$d(x, c_p) \;<\; d(x, c_q) \quad \forall x \in V(c_p), c_q \neq c_p$$

The boundary of a cell V (cp) and its adjacent cell V (cq) consists of points on the *perpendicular bisector*, denoted cp|cq between the points cp and cq. The bisector is the hyper plane that is perpendicular to the line segment joining cp and cq that passes through the mid-point of the line segment. This hyperplane divides the space Rm into two halves.

Algorithm 3 Voronoi-cell Pruning (VD)

Compute the Voronoi diagram for $C = \{c_1, \ldots, c_k\}$.
for all $c_j \in C$ **do**
  if $\mathrm{MBR}_i \subseteq V(c_j)$ **then**
    $Q_k \leftarrow \{c_j\}$ */*The one and only one candidate*/*

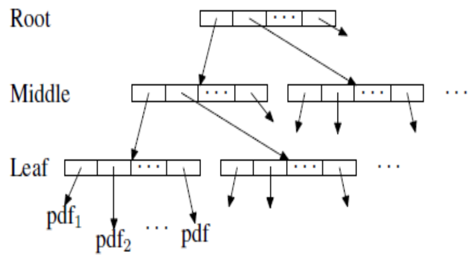
Algorithm 4 Bisector Pruning (Bi)

Extract all $H_{p/q}$ from Voronoi diagram for C
for all distinct $c_p, c_q \in C$ **do**
  if $\mathrm{MBR}_i \subseteq H_{p/q}$ **then**
    remove c_q from Q_i

D&E. Indexing the Uncertain Objects

The above pruning techniques all aim at reducing the number of ED calculations, which dominates the execution time of UK-means. As we will see later (Section V), our pruning techniques are so effective that in many cases more than 95% of the ED calculations are pruned. The cost of other computations, such as the pruning overhead, now becomes relatively significant. In order to further reduce the execution time, we have devised further techniques to reduce the pruning costs. Observing that Voronoi-diagram-based pruning techniques (namely VD and Bi) takes advantage of the spatial distribution of the cluster representatives, it is natural to ask whether we can also make use of the spatial distribution of the uncertain objects. Can we organize the objects so that nearby objects are grouped together and processed in batch to avoid repeating similar computations, such as similar pruning condition testing? If we first divide the uncertain objects into groups, we can obtain an MBR for each group (the minimum rectangle enclosing all objects in the group). With these MBRs, we can apply MinMax-BB, VD and Bi pruning onto the groups. This allows cluster candidate pruning at the group level. In the ideal case where a single cluster is assigned to a whole group, all the member objects of that group get assigned the cluster at once, saving the computations needed to assign clusters to each member individually. This saving is potentially significant. Furthermore, we can group the groups into super groups, forming a hierarchy. Our pruning techniques can then be applied to different levels in the hierarchy in a top down manner. To get good pruning effectiveness, the grouping should be done in a way that minimizes the volumes of the MBRs. A natural choice is to group objects using an R-tree structure.

1) R-Trees: The R-tree is a tree structure that can be considered a variant of B+-Tree. As such, it is a self-balancing tree with all leaf nodes at the same depth from the root node. The main difference is that R-tree is designed for indexing multi-dimensional spatial data to support faster proximity based queries. The tree has the property that each internal node stores also the



MBR for all the objects stored under that sub tree. Data items are not stored in the internal nodes. R-tree facilitates spatial query processing. As an example, to locate all objects that are within a distance d from a certain query point x, one starts from the root node, and only needs to descend (recursively) into the child nodes whose MBRs intersect the sphere with radius d centered at x. This brings the search to only those few leaf nodes containing the objects being searched for. In our implementation, we use an R-tree like the one depicted in Figure 3. Each tree node, containing multiple entries, is stored in a disk block. Based on the size of a disk block, the number of entries in a node is computed. The height of the tree is a function of the total number of objects being stored, as well as the fan out factors of the internal and leaf nodes. Each leaf node corresponds to a group of uncertain objects. Each entry in the node maps to an uncertain object. The following information are stored in each entry:

- The MBR of the uncertain object.
- The centroid of the uncertain object.
- A pointer to the pdf data of the object.

Note that the pdf data are stored outside the tree to facilitate memory utilization. Each internal node of the tree corresponds to a super-group, which is a group of groups. Each entry in an internal node points to a child group. Each entry contains the following information:

• The MBR of the child group.
• The number of objects under the sub tree at this child.
• The centroid of the objects under the sub tree at this child.
• A pointer to the node corresponding to the child.

Note that storing the number of objects under the sub tree at a child node and the corresponding centroid location allows efficient readjustment of cluster representatives at the end of each iteration of UK-means (steps 7–8, Algorithm 1). To build an R-tree from a database of uncertain objects, we use a bulk-load algorithm based on the Sort-Tile-Recursive algorithm [14]. It builds an R-tree from bottom up (as opposed to repeated insertion from the top) and has the advantages of building a more fully filled tree, with smaller MBRs for the internal nodes, and a shorter construction time. We illustrate this algorithm with a 2D example shown in Figure 4. The figure shows the MBRs of 36 uncertain objects. Suppose the fan out factor of leaf nodes is 4. Then, each leaf node will contain 4 uncertain objects and 9 leaf nodes are needed.

2) Group-Based Pruning: With an R-tree in place, we already have a multi-level grouping of the uncertain objects. In addition, the information kept at each node helps us do pruning in batch, thereby further boosting the performance of the pruning algorithms. Instead of repeating the cluster assignment to each uncertain object one after another as in UK-means, we traverse recursively down the tree, starting from the root node. We examine each entry of the root node. Each entry e represents a group (or super-group) of uncertain objects. The MBR of e is readily available from the R-tree. Using this MBR, we can apply our pruning techniques MinMax-BB, VD or Bi to prune away candidate clusters in the same way as explained before. These techniques work here because they are guaranteed to prune away a cluster representatives cp if there is for sure another cluster representatives cq 6= cp such that all points in the MBR are closer to cq than to cp. Since this property holds for all points within the group's MBR, it also holds for all sub-groups and uncertain objects under the sub tree.

VI. CONCLUSION

In this paper we have studied the problem of clustering uncertain objects whose locations are represented by probability density functions. We have discussed the UK-means algorithm [5], which was the first algorithm to solve the problem. We have explained that the computation of expected distances dominates the clustering process, especially when the number of samples used in representing objects' pdf's is large. We have mentioned the existing pruning techniques MinMax-BB and cluster-shift (CS) [6]. Although these techniques can improve the efficiency of UK-means, they do not consider the spatial relationship among cluster representatives, nor make use of the proximity between groups of uncertain objects to perform pruning in batch. To further improve the performance of UK-means, we have first devised new pruning techniques that are based on Voronoi diagrams. The VDBi algorithm achieves effective pruning by two pruning methods: Voronoi-cell pruning and bisector pruning. We have proved theoretically that bisector pruning is strictly stronger than MinMax-BB. Furthermore, we have proposed the idea of runing by partial ED calculations and have incorporated the method in VDBiP. Having pruned away more than 95% of the ED calculations, the execution time has been significantly reduced. It has been reduced to such an extent that the originally relatively cheap pruning overhead has become a dominating term in the total execution time. To further improve efficiency, we exploit the spatial grouping derived from an R-tree index built to organize the uncertain objects. This R-tree boosting technique turns out to cut down the pruning costs significantly. We have also noticed that some of the pruning techniques and R-tree boosting can be effectively combined. Employing different pruning criteria, the combination of these different techniques yield very impressive pruning effectiveness. We have conducted extensive experiments to evaluate the relative performance of the various pruning algorithms and combinations. The results show that our new pruning techniques outperform MinMax-BB consistently over a wide range of experimental parameters. The overhead of computing Voronoi diagrams for our Voronoi-diagram-based technique is paid off by the large number of ED calculations saved. The overhead of building an R-tree index also gets compensated by the large reduction of pruning costs. The experiments also consistently demonstrated that the hybrid algorithms can prune more effectively than the other algorithms. Therefore, we conclude that our innovative techniques based on Voronoi diagrams and R-tree index are effective and practical.

REFERENCES

[1] Ben Kao, Sau Dan Lee,Foris K. F Lee, "Clustering uncertain data using Voronoi diagrams and R-tree indexing" IEEE transaction on knowledge and Data Engineering Vol.22 No.9 September 2010.

[2] Cheng Zhang, Ming Gao, Aoying Zhou "Tracking high quality clusters over uncertain data streams" IEEE international conference on data engineering 2009.

[3] Wang kay Ngai, Ben Kao, Chun Kit Chui, Michael Chau, Reynold Cheng, Kevin Y. Yip "Efficient clustering of uncertain data" Sixth international conference on data mining. (ICDM 2006).

[4] Biao Qin, Yuni Xia, Sunil Prabhakar, Yicheng Tu "A rule based classification algorithms for uncertain data" IEEE international conference on data engineering 2009.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 2, Issue 3, September 2012

[5] P.Misra and P. Enge, Global Positioning System: Signals, Measurements, and Performance, 2nd ed. Ganga-Jamuna Press, 2006, iSBN 0-9709544-1-7.[2] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy efficient communication protocol for wireless micro sensor networks," in 33rd Annual Hawaii International Conference on System Sciences (HICSS), IEEE, Maui, Hawaii, U.S.A., 4th-7th Jan. 2000.

[6] S. Bandyopadhyay and E. J. Coyle, "An energy efficient hierarchical clustering algorithm for wireless sensor networks," in The 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM), San Francisco, U.S.A., 30th Mar.–3rd Apr. 2003.

[7] O. Wolfson and H. Yin, "Accuracy and resource consumption in tracking and location prediction," in SSTD, ser. Lecture Notes in Computer Science, vol. 2750. Santorini Island, Greece: Springer, 24-27 Jul. 2003, pp. 325–343.

[8] M. Chau, R. Cheng, B. Kao, and J. Ng, "Uncertain data mining: An example in clustering location data," in PAKDD, ser. Lecture Notes in Computer Science, vol. 3918. Singapore: Springer, 9–12 Apr. 2006, pp. 199–204.

[9] W. K. Ngai, B. Kao, C. K. Chui, R. Cheng, M. Chau, and K. Y. Yip, "Efficient clustering of uncertain data," in ICDM. Hong Kong, China: IEEE Computer Society, 18–22 Dec. 2006, pp. 436–445.

[10] F. K. H. A. Dehne and H. Noltemeier, "Voronoi trees and clustering problems." Inf. Syst., vol. 12, no. 2, pp. 171–175, 1987.

[11] B. Kao, S. D. Lee, D. W. Cheung, W.-S. Ho, and K. F. Chan, "Clustering uncertain data using Voronoi diagrams," in ICDM. Pisa, Italy: IEEE Computer Society, 15th–19th Dec. 2008, pp. 333–342.

[12] N. N. Dalvi and D. Suciu, "Efficient query evaluation on probabilistic databases," The VLDB Journal, vol. 16, no. 4, pp. 523–544, 2007.

[13] D. Barbar´a, H. Garcia-Molina, and D. Porter, "The management of probabilistic data," IEEE Trans. Knowl. Data Eng., vol. 4, no. 5, pp. 487–502, 1992.

[14] R. Cheng, D. V. Kalashnikov, and S. Prabhakar, "Querying imprecise data in moving object environments," IEEE Trans. Knowl. Data Eng., vol. 16, no. 9, pp. 1112–1127, 2004.

[15] R. Cheng, Y. Xia, S. Prabhakar, R. Shah, and J. S. Vitter, "Efficient indexing methods for probabilistic threshold queries over uncertain data," in VLDB. Toronto, Canada: Morgan Kaufmann, 31 Aug.–3 Sept. 2004, pp. 876–887.

[16] J. Chen and R. Cheng, "Efficient evaluation of imprecise location dependent queries," in ICDE. Istanbul, Turkey: IEEE, 15-20 Apr. 2007, pp. 586–595.

[17] H.-P. Kriegel and M. Pfeifle, "Density-based clustering of uncertain data," in KDD. Chicago, Illinois, USA: ACM, 21–24 Aug. 2005, pp.672–677.