Data Structures for Dynamic and Kinetic Multidimensional Point Sets

David Mount

University of Maryland, College Park

MURI Kickoff Meeting – 2008
Motivation

Latent-Space Embedding for Large Networks

Performed through iterative search. Large networks involve:

- solution spaces of very high dimension
- a large number of iterations to achieve convergence
- inner-loop computations of quadratic complexity

Recent results in spatial data structures

- Simple and practical dynamic structures
- Intrinsic data structures for kinetic updates
- Fast approximations through hierarchical sketching
- Space/query-time tradeoffs
**Quadtree**: Hierarchical structure based on subdivision of squares.

**Path Compression**: Chains of trivial splits are compressed.

- **Splitting**
- **Compression (shrinking)**
Basic Structures: Balanced Partition Trees

**Problem:** Compressed quadtrees may have depth $\Omega(n)$.

**Balanced Structures**

Achieving $O(\log n)$ depth:

**BBD-tree:** [AMN98] Combines splitting with centroid shrinking.
- Produces an inner cell and outer cell.

**BAR-tree:** [DGK01]
- Uses slanted splitting planes to split clusters. (Produces convex cells.)
Basic Structures: Balanced Partition Trees

**Problem:** Compressed quadtrees may have depth $\Omega(n)$.

**Balanced Structures**

Achieving $O(\log n)$ depth:

**BBD-tree:** [AMN98] Combines splitting with centroid shrinking. Produces an inner cell and outer cell.

**BAR-tree:** [DGK01] Uses slanted splitting planes to split clusters. (Produces convex cells.)
Approximate Spherical Range Searching

Example: Count (or report) the points lying within a given spherical range. Allow an error of $\varepsilon$.

Preprocessing: Build BBD tree.

Query Processing:

- Find maximal cells lying within the outer range and covering the inner range
- Access counts for each cell
- Return the total

Query time: $O(\log n + (1/\varepsilon)^{d-1})$.
Approximate Spherical Range Searching

**Example:** Count (or report) the points lying within a given spherical range. Allow an error of $\varepsilon$.

**Preprocessing:** Build BBD tree.

**Query Processing:**
- Find maximal cells lying within the outer range and covering the inner range
- Access counts for each cell
- Return the total

**Query time:** $O(\log n + (1/\varepsilon)^{d-1})$. 
How to support insertion and deletion?

**Analogy — Skiplist [Pug90]**

A randomized structure for 1-dimensional data. Build linked lists for successive random samples.

- $S_0 \leftarrow S$ (the original point set).
- $S_1 \leftarrow$ sample $S_0$ with probability $\frac{1}{2}$.
- $S_2 \leftarrow$ sample $S_1$ with probability $\frac{1}{2}$.
- The process ends after $O(\log n)$ stages in expectation.
Skip Quadtree

Skip Quadtree [EGS05]

Same idea, but applied to a sequence of compressed quadtrees.

- \(Q_0 \leftarrow \) quadtree for \(S_0\).
- \(Q_1 \leftarrow \) quadtree for \(S_1\).
- \(Q_2 \leftarrow \) quadtree for \(S_2\).
- ... Each node of \(Q_i\) is linked to its counterpart in \(Q_{i-1}\).

Although each quadtree may be unbalanced (like a linked list) it is possible to access each node in \(O(\log n)\) time through the links.
Skip Quadtree

Lemma 2 For any point $x$, the expected number of searching steps within any individual $Q_i$ is constant.

Figure 2: A randomized skip quadtree consists of adjacent compressed quadtrees linked by a double-headed arrow between the square centers.
Latent embedding algorithms involve adjusting the location many points with each iteration. The motion of each point may be small.

**Kinetic Updates**

Update a data structure after a *small motion* involving *many points* of the set.

- Data structures that are defined in terms of a fixed *coordinate frame* are sensitive to changes in *absolute position* of points, even if the *relative position* remains unchanged.
- Can we define spatial data structures that are *independent* of any coordinate frame?
**r-Net**

... is a subset $X \subseteq S$ such that

(i) every point of $S$ is within distance $r$ of some point of $X$

(ii) the pairwise distance between any two points of $X$ is $\geq r$

The balls of radius $r$ form a cover, which is similar to the partition provided by the square cells of one level of a quadtree.
Kinetic Structures

Net Tree

The leaves are $S_0 = S$.
Let $r$ be the minimum distance between any two points of $S$.
Let $S_1$ be an $r$-net of $S_0$.
Let $S_2$ be a $(2r)$-net of $S_1$.
Let $S_j$ be a $(2^j r)$-net of $S_{j-1}$.
... 
Until only one remains — the root.

Similar to the quadtree in spirit, but intrinsic to the point set.
Challenges

Depth vs. Breadth: Most innovations have focused on maintaining structures of low depth. However, search times are dominated by the search’s breadth, that is, the number of cells visited.

Tight vs. Slack: The best static structures achieve efficiency by enforcing tight constraints on subdivision properties (e.g. tree depth, cell size, aspect ratio). However, tight constraints result in frequent certificate failures, as used by kinetic structures.

The Right Mixture: What is the proper mixture of methods to achieve best overall performance, in terms of accuracy and execution times?
S. Arya, D. M. Mount, N. S. Netanyahu, R. Silverman, and A. Wu.
An optimal algorithm for approximate nearest neighbor searching.

S. Arya and D. M. Mount.
Approximate range searching.
Bibliography


J. Gao, L. J. Guibas and A. Nguyen.
Deformable Spanners and Applications.

S. Har-Peled and M. Mendel.
Fast construction of nets in low dimensional metrics, and their applications.

W. Pugh.
Skip lists: A probabilistic alternative to balanced trees.