Approx Nearest Neighbor

Ariya-Mount Journal Paper

Last Time Cell-Decomp
Prob: Not clear how to get efficient point location

Recall: Box = Red \( \frac{ls}{ss} \leq 3 \) or aligned

Cell = Box
2) Diff of 2 boxes (sticky)

Sticky: \( b_i \) sticky in box \( b \)

\( d = 0 \) or \( d_i \geq h_i \)
\textbf{Split} \((b, S)\)

1) even
2) \(b \cap S = \emptyset \) \& \( b_2 \cap S = \emptyset \)
else "not possible"

\textbf{Shrink} \((b, S)\)

Minimal \(\{S \subseteq b' \cap b' \text{in a sticky box}\}\)
\(b' \subseteq b\)

\textbf{Note} Shrink in \(O(1)\) time

Problem split: We need to split \(S\) into \(S_1\) \& \(S_2\)...

Idea: \(S = \{p'_1, \ldots, p'^k\}\) \(p^i = (p^i_x, p^i_y)\)

Maintain linked lists \(L_x, L_y\)

\(L_x = p'_1 \cdots p'^k\)
\(L_y = p'_1 \cdots p'^k\)

Cross linked

\(p^i_x \leftrightarrow p^i_x\)
\(p^i_y \leftrightarrow p^i_y\)

\(x\)-list

\(y\)-list
Suppose we split vertically:

Let's work on the x-list first.

We need only break the list i.e.

find set \( P_x \in b \), & \( P_x \in b \).

Solution: work from both ends of list.

\(
\text{Cost} = \min \{ |S_1|, |S_2| \}
\)

What about y-list?

Assume \( |S_1| \leq |S_2| \)

We splice out \( S_1 \) from the y-list.

\( \text{Cost} = 20|S_1| \)

In last week's alg we now need to sort y-list of \( S_1 \).
If y-list also included rank, numbers \( \in \{1, \ldots, k\} \)

Then sorty would be \( O(k) \) (bucket sort)

\( \text{Prob: } |S_1| \ll k \). \)
New alg: Idea: work on $S_a$ until $|S_a| \leq \frac{a}{3}k$.

**Alg centroid-shrink** ($b, S, k$) \( |S|=k \) (case $b \in \text{box}$)

While \( |S| > \frac{a}{3}k \) do

\( b = \text{Shrink} (b, S) \)

\( (b_1, S_1), (b_2, S_2) = \text{split} (b, S) \quad |S_1| \leq |S_2| \)

Centroid-shrink ($b_2, S_2$)

---

Case: $b$ in a cell $b_0$ & $b_1$

Shrink ($b_0, b_1, S$) = shrink but include $b_1$.

Split ($b_0, b_1, S$) = if ($b_1, S$) & ($b_2, S_2$)

If $b_1 \subseteq b_2$ & $|S_1| \leq |S_2|$ continue

Else ($b_1, b_1, S_1$) & centroid-shrink ($b_2, S_2, k$)
Cost = \( T(\ell) \)

Claim: \( T(\ell) = O(\ell) \)

Cost of shrink-splits = \( \sum C_{\ell_i} \) when \( \sum \ell_i = \ell \)

At the end we do 2 spliceouts at a cost of \( O(\ell) \)

Balanced-Box Decomposability (BBD)

\( T(\ell) = O(n^{\frac{1}{2}}n) \quad T(n) = 2T(\sqrt{n}) + c \cdot n \)

Each cell is assigned a bounded size list of points

\( C = (b_0, b_1) \) gets a point from \( b_1 \)
T = BST  query point q

r = root(T)

Q = priority queue (dist(box, q))

Search(q, T)

enqueue(r)