La eficiencia de la operación de búsqueda en este ADT es sensible a la distribución de los datos. Es decir, a igual tamaño de espacio y número de puntos, la búsqueda no toma el mismo tiempo si todos los datos están agrupados en una sub-región del espacio, que sí están distribuidos uniformemente.

Plantee al menos una hipótesis sobre la eficiencia de la búsqueda y valídela experimentalmente.

Pista:

puede utilizar distribuciones de puntos conocidas como uniforme, Gauss, Zipf, etc.

----------------------------------------------------------------------------------------------------------------

Como solo las hojas contienen información, en el peor de los casos buscar un punto que no esté en el árbol se recorre hasta el fondo del árbol

Si c es la menor distancia entre cualquier par de puntos y s es el largo del lado del cuadrado original que contiene todos los puntos. Entonces

la profundidad(altura) del quadtree es ≤ log(s/c) + 3/2

Un quadtree con profundidad d almacenando n puntos tiene O((d+1)n) nodos.

Hipotesis

Mientras más juntos estén los puntos, más subdivisiones tendrán los cuadrantes. Por lo que el árbol tendrá mayor altura que si estuvieran distribuidos uniformemente.

Entonces en teoría es más eficiente buscar cuando están distribuidos uniformemente, porque al haber menos subdivisiones la altura del árbol será menor.

Validación experimental

Se plantean 4 casos de estudio.

1. Cuando están todos distribuidos uniformemente
2. Cuando los puntos están distribuidos aleatoriamente.
3. Cuando los puntos están agrupados y el punto a buscar pertenece al cuadrante donde están los puntos agrupados.
4. Cuando los puntos están agrupados y el punto a buscar pertenece al cuadrante donde NO hay punto.

Estos casos fueron probados con aproximadamente 500000 puntos en un plano de 4096\*4096

Caso 1 (equilibrado)

4.3e-05 sec -> 0.000043 sec

Caso 2 (aleatorio)

4.3e-05 sec -> 0.000043 sec

Caso 3 (acumulados NW)

1e-04 sec -> 0.000100 sec

Caso 4 (acumulados NW\_no\_esta)

1.5e-05 sec -> 0.000015 sec

Experimentalmente se confirma que cuando están acumulados requiere más tiempo que si estuvieran bien distribuidos.

