# 文件介绍

## hierarchical\_clustering.py

定义POI类，Cluster类，HierarchicalClustering类

## min\_hierarchical\_clustering.py

定义MinHierarchicalClustering类

## write\_file.py

定义writeFile函数和dfs函数

## main.py

主函数

# 类介绍

## POI

作用：

描述POI信息的类。用经纬度标识一个POI。

成员变量：

lon：经度

lat：纬度

## Cluster

作用：

描述簇信息的类。

成员变量：

POIs：POI对象序列，表示构成该簇的点集。

left,right：该簇在层次聚类树上的左右子簇

成员函数：

combine：静态方法，合并连个簇，生成一个新簇

## HierarchicalClustering

作用：

求层次聚类

成员变量：

dist：list<tuple(Cluster,Cluster)>，记录簇两两之间相似度。

clusters：当前阶段的所有簇。

成员函数：

\_\_init\_\_：构造函数。将每个点视为一个簇构建Cluster对象，并求出这些簇之间的相似度，存储在dist中。

run：求层次聚类的函数。根据clusters求层次聚类，直至只剩一个簇。返回层次聚类树的根节点。

getCombinedDist：

更新两个簇合并之后的dist

calcDist：

计算两个簇之间的相似度

## MinHierarchicalClustering

作用：HierarchicalClustering的子类，用簇之间的最小距离来表示簇之间的相似度。（最小距离指两个簇之间距离最近的两个点之间的曼哈顿距离）

成员变量，成员函数同HierarchicalClustering

# 算法介绍

## HierarchicalClustering.run

循环取出簇集合中相似度最大的两个簇，调用Cluster.combine函数将这两个簇合并，生成一个新的簇，更新层次聚类树，并调用getCombinedDist函数更新相似度表。直至仅剩一个簇，算法结束，返回树的根节点。

## MinHierarchicalClustering. getCombinedDist

根据簇合并的性质，可以通过递推而非计算的方式得到新生成的簇new与其他簇之间的相似度，即遍历其他簇i，将合并的两个簇a,b，取dist(new,i)=min(dist(a,i),dist(b,i))。优化了时间复杂度。

## MinHierarchicalClustering. calcDist

暴力计算两个簇之间的相似度，对于MIN方法，枚举两个簇的点，计算两点之间的曼哈顿距离，取最短的距离为两个簇之间的相似度。该函数主要用于初始化阶段，即每个簇只有一个点的情况下，求出相似度矩阵。

## writeFile

调用dfs函数遍历层次聚类树，将聚类信息已json格式写入规定的文件

## dfs

用深度优先搜索遍历层次聚类树