2020 国际大数据竞赛数据说明

Data Description of 2020 International Big Data Competition

1、城市区域背景(Urban context):

每个城市文件夹下的density.csv、migration.csv、transfer.csv中的坐标均属于该城市;grid_attr.csv为该城市重点区域覆盖的坐标对应关系，重点区域未覆盖城市所有网格坐标。

The coordinates of density.csv, migration.csv and transfer.csv in each city folder belong to the city; grid_ attr.csv is the coordinate correspondence of the key urban regions, which do not cover all the grid coordinates.

2、transfer.csv中有某小时网格多值情况(In transfer.csv, there is hourly grid multi value situation):

每小时城市与城市之间的迁移数据，如某小时有多路迁移，建议求和计算

It includes hourly urban-to-urban migration data. If there is multiple migration in an hour, it is recommended to sum the data.

3、grid_attr.csv坐标解释(grid_ attr.csv coordinate description):

城市区域中坐标为该城市重点区域覆盖的坐标;坐标没有按照顺序排列;坐标不是区域的外围坐标点集合，是属于区域中的点。

The coordinates in the urban region are the coordinates covered by the key urban regions; the coordinates are not arranged in order; the coordinates are not a collection of peripheral coordinate points of the region, but belong to the coordinate points within the region.

4、指数解释（Index description）：

1）migration.csv城市间迁徙指数： 反映城市之间的流入或流出人口规模，城市间可横向对比

migration.csv (Intercity migration index): it represents the population size of inflow or outflow between cities, which can be compared horizontally.

2）transfer.csv中网格关联强度指数： 反映网格之间分小时的出行量规模

transfer.csv (Medium grid association intensity index): it represents the hourly travel volume between grids.

3）density.csv网格人流量指数： 反映网格内分小时的人流量规模

density.csv (Grid human flow index): it represents the scale of hourly human traffic in the grid.

注意：不同指标的计算策略不同，指标之间不能直接相互对比

Note: the calculation strategies of various indicators are different, and the indicators cannot be directly compared with each other.