

训练数据简述

```
>>> raw.head(5)
   date      time  area  upload_quantity_GB  download_quantity_gb
0  2018/3/1  0:00:00    1         0.324710          3.560744
1  2018/3/1  0:00:00    2         0.072757          0.830950
2  2018/3/1  0:00:00    3         0.137215          0.536776
3  2018/3/1  0:00:00    4         0.039931          0.265031
4  2018/3/1  0:00:00    5         0.082546          1.537391
```

总行数 1 4413 8200

总区域数 13 2279

每个区域下的数据行数量

```
count    132279.000000
mean      1089.652931
std       250.277985
min        1.000000
25%      1150.000000
50%      1193.000000
75%      1194.000000
max      2388.000000
```

目标

特定区域 特定时间段 预测流量

解决方案

对所有区域根据历史行为进行聚类
对于同一类区域，选择这些区域的历史行为训练模型
序列模型：针对历史序列数据推测未来序列

咨询

可能存在哪几种流量模式，怎么量化这些模式