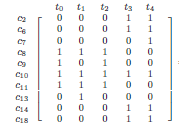
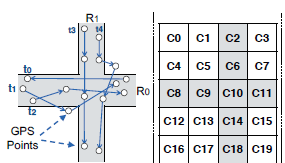
# 软件学院 数据分析/挖掘课程作业之1

1. 利用提供的1000条轨迹数据(位于上海地区), 将这些轨迹数据投影到百度地图以进行可视化;
2. 使用Pythong lshash 0.0.4dev包 (<https://pypi.python.org/pypi/lshash/0.0.4dev>)处理上述轨迹数据

2.1) 将上海地区进行栅格化处理, 形成轨迹-栅格的0./1矩阵



如上图所示, 左图包括t0…t4五个轨迹数据, 位于R0和R1两个道路,中间图将该区域划分为5\*4的栅格, 右图为轨迹-栅格的0/1, 其中一个轨迹例如t0经过栅格c8,则对应的矩阵元素为1,否则为0(t0不经过栅格c2, 则对应的矩阵元素为0).

2.2) 完成栅格化处理之后, 每个轨迹对应一个0/1向量, 利用Pythong lshash进行处理, hash\_size为栅格总数的0.01, 0.05, 0.1, 0.15, 0.20, 0.25, 0.3, 任意选择一个轨迹, 然后高亮所选择的轨迹及其knn的轨迹 (k = 1, 2, 3, 4, 5).

1. 提交日期：2017/03/25日 23：59PM

附录信息:

a）服务器地址及帐号信息

实验服务器地址 Sever: 101.37.20.204 Username: student Password: dataming

实验数据存放目录：/home/student/DataMining/Traj（只许读，严禁写入修改删除移动！ ）

推荐安装 Xftp+Xshell5 访问和获取数据

建议在本机配置相关实验环境：推荐 linux+Python+Anaconda，将数据下载到本地 PC 进行实 验。如需在实验服务器上操作，请在/home/student/下建立自己姓名的目录，如有非法目录建立，会定期清理，请大家注意。

推荐机器学习 python 算法包：scikit-learn 安装 Anaconda 会自动包含 scikit-learn

相关 API 文档，参考 <http://scikit-learn.org>

b）轨迹数据格式:

每一行为一条轨迹的GPS点数据信息，具体包括4个字段信息：Tid, Time, X, Y

其中Tid是轨迹编号，Time是这个轨迹点的时间，轨迹坐标以UTM码的形式给出，其中X,Y是UTM码的两个坐标，UTM码的单位是米，处理过程中可以直接用UTM码进行格子的切割。轨迹所在范围为（346000,3448600）到（362800,3463800）的矩形内。栅格可按20米乘20米的大小切割，有轨迹点落下的栅格共有44107个栅格，可把这44107个栅格作为LSH运算中每条轨迹的属性。