# 第三章 问题定义

人们会使用采样设备记录下移动对象的一些位置信息，而位置信息中包含了空间和时间属性。现实世界虽然是一个三维的空间，但是人和车辆的移动可以看做是在二维平面中的移动，并且在轨迹相似性查询中，关于移动对象在二维空间中的移动轨迹的研究最为广泛，因此在空间上本文将问题放在二维空间下研究。除了空间信息外，本文中的位置信息还涉及到时间信息。考虑到人的户外活动一般是以天为单位，因此我们时间信息中不需要包含年月日等信息，只需要用到包含时分秒的时刻信息。所以位置点包含的信息有经度、纬度和时间戳。

实际上移动对象的真实路径包含无穷多个点，这无穷个点共同组成移动对象的一段连续的移动路线，但是GPS的采样策略一般是每隔一定时间或者移动对象每移动一段距离进行一次位置信息的记录，因此最后得到的轨迹数据是由有限个位置点组成。下面介绍与轨迹数据相似性相关概念的定义。

定义3-1（样本点）由采样设备获得移动对象在某一时刻下的地理位置信息，包括空间信息和时间信息，其中空间信息由经度和纬度组成，时间信息是一个时间戳。这些信息组成一个样本点p，或者叫采样点，。

定义3-2（轨迹）通过采样设备获得了一个移动对象一段时间内的样本点，这些样本点以时间顺序组成的序列T叫做轨迹，其中。

定义3-3（轨迹相似性计算函数）输入两条轨迹数据Q和R，使用轨迹相似性计算函数进行计算，输出一个实数值表示轨迹之间的相似程度，该值越大，两条轨迹越相似。

定义3-3（轨迹相似性查询）给定一条查询轨迹Q，以及一个存储了轨迹数据的数据库DB，输出DB中的一条轨迹，使得，有。

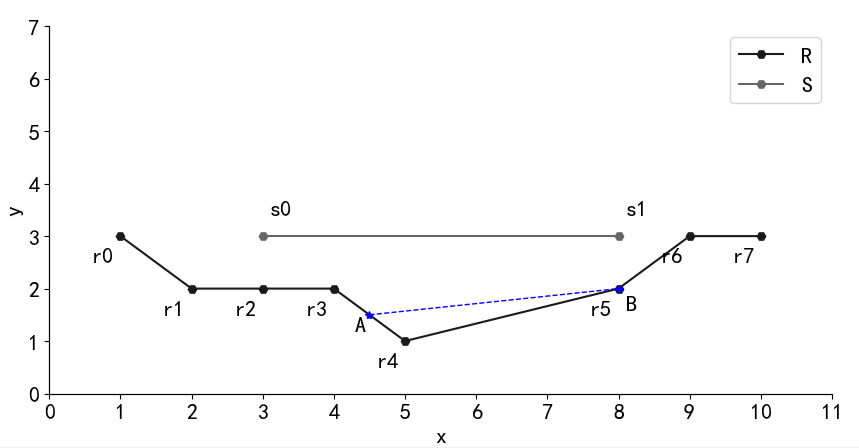
上面定义了一些轨迹相似性计算中用到的一些基本概念，下面再介绍本文中需要用到的一些其他条件。本文在实验中首先使用了兴趣点和转折点提取算法，将能描述轨迹形状的点以及移动对象的兴趣点提取出来，进行进一步分析。因此如果不加解释，本文中使用的样本点全部是经过提取算法处理后的点，而不是原始轨迹数据。

数据库中存储的采样设备采集的轨迹数据通常是一个用户或者一辆汽车一天的移动路线的记录，描述了一个移动对象在很长一段时间内的移动情况。如果是出租车轨迹，可能这辆车一天之内来来回回在城市大小景点、酒店中间要转好几圈，而我们输入的查询轨迹有时会比较短，查询轨迹只需要起点、终点加几个转折点就可以确定，下面以一个例子说明这种情况中存在的问题。为简化起见，我们使用两条长度稍短的轨迹来说明。

假设警察在追踪驾车逃跑的疑犯，但是由于附近道路的监控出现故障，只有A点和B点的监控拍摄到了疑犯的车辆，并由于故障，AB两地的摄像头不能拍清楚过路车辆的车牌号。警察希望通过行车轨迹数据库找到当时路过的车辆，查看行车记录仪是否拍摄到疑犯车内的详细情况以及逃往哪个方向。

假如数据库中存在两条轨迹R和S，如图所示。其中轨迹R在8点的时候从开往处，到达点为8点30分，途径A点和B点。轨迹S在8点10分从点开至点，到达点为8点25分。查询轨迹Q是8点15分，从A点开始，到8点25分，结束于B点的这样一条简单的轨迹。查询要求是希望得到在时间、空间和形状上和查询轨迹最相似的轨迹。由于本例子中查询轨迹过于简单，可忽略形状上的相似性。

根据轨迹R和S的文字描述以及可视化的轨迹展示，发现轨迹R中从到这段子轨迹段很符合查询要求。但是如果使用之前的轨迹相似性查询方法，会直接将整条数据轨迹R和查询轨迹Q计算相似性，带来的结果就是与A和B相距较远的样本点如到、到会对加大轨迹间的距离，不能发现轨迹R的子轨迹段和查询轨迹的相似。



给定了查询轨迹之后，在计算数据库中大量的数据轨迹与查询轨迹的相似性时，会出现这样一个问题，就是

下面介绍本文需要解决的问题。在前人的研究中，一些经典的轨迹相似性查询如DTW、LCSS和EDR不能完全适用于每一个需要使用轨迹相似性查询

高时效性、形状、距离、子轨迹段的概念介绍

高时效性下的轨迹局部相似性查询算法的设计与实现

问题定义：在考虑时间和二维空间的情况下，找出数据轨迹中与查询轨迹最相似的部分子轨迹段，并计算该子轨迹段与查询轨迹段的相似性。在数据库中找出最相似的子轨迹段。