# 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。

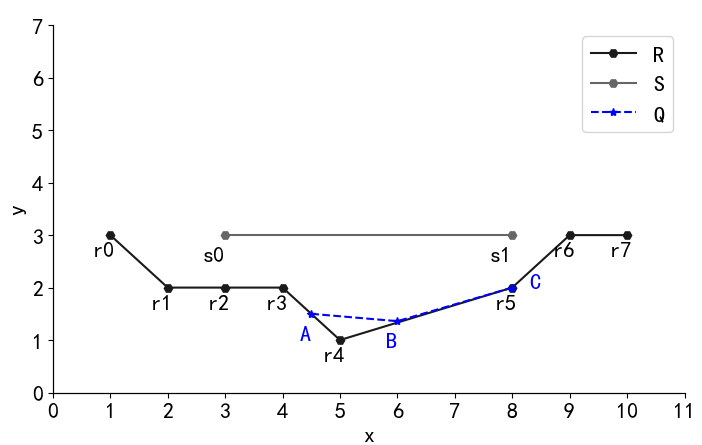
定义3-3（轨迹相似性查询）给定一条查询轨迹Q，以及一个存储了轨迹数据的数据库DB，输出DB中的一条轨迹，使得，有。

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。

假设用户在A点看到一辆套用自己车牌的车在行驶，想截住他，于是从A点开始驾车追，途径B点，但是在C点时跟丢了。由于用户车内摄像头坏了，为了提供更多信息给交警，用户希望通过自己当时的行车轨迹，在轨迹数据库中查找有相似轨迹的车辆，联系车主查看行车记录仪，以提供更多信息。

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即用户的行车轨迹，是8点15分从A点开始，途径B点，到8点25分结束于C点。查询要求是希望得到和查询轨迹最接近的轨迹。



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，会直接计算整条数据轨迹R和查询轨迹Q的相似性，其结果就是轨迹R中轨迹Q相距较远的样本点如到、到会对加大轨迹间的距离，掩盖了轨迹R中到子轨迹段和查询轨迹的相似性，使得与Q相距较远的轨迹S成为了搜索结果，很明显S不是我们希望得到的查询结果。

为了解决这个问题，需要用到子轨迹的定义。

定义3-4（子轨迹）给定一条轨迹数据，T的子轨迹是T的一部分，其中，i<=j，且包含了轨迹数据T从到的所有样本点。

下面给出本文的问题定义。给定一条查询轨迹Q，以及一个存储轨迹的数据库DB，在考虑时间、空间和轨迹形状的情况下，希望查询到DB中与Q最相似的轨迹，并允许查询结果是某条轨迹的子轨迹。本文称该问题为轨迹局部相似性查询问题。