## 3.2时空归一化

### 3.1.1 PTM算法中时空维度结合存在的问题

之前有很多研究人员研究过包含时间和空间两个维度的轨迹相似性查询技术，大多数采取的做法是将时间和空间分割开，分别计算相似程度或者距离，然后使用权值将二者进行结合。最典型的方法就是下面的PTM算法（Personalized trajectory matching）。

PTM算法主要利用轨迹间的时间和空间因素，计算轨迹间的相似程度。首先考虑空间因素，查询轨迹Q上的点和数据轨迹R上的点的空间距离为，这里的空间距离计算的是路网上的距离，基于此空间距离，获得和的空间距离影响因子，如公式（2）所示。距离越大，影响因子越小，当距离大到一定程度时，将该因子置0。然后借用LCSS的思想，基于轨迹上所有样本点的位置关系以及样本点的权重得到空间上最大相似程度，如公式（4）所示，其中权重是基于用户对某一个点的重视程度来确定，也是文章的创新点之一。然后考虑时间因素，查询轨迹Q上的点和数据轨迹R上的点的时间差为，时间影响因子为。然后基于所有点的时间关系得到时间相似因子，其计算方法与空间相似因子的计算方法类似。最后使用参数将空间相似度和时间相似度结合起来，得到轨迹相似度，如公式（6）所示。





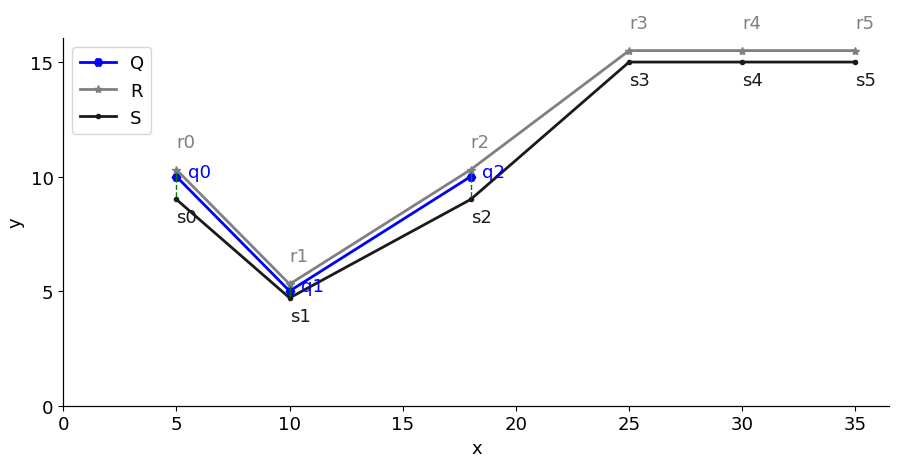


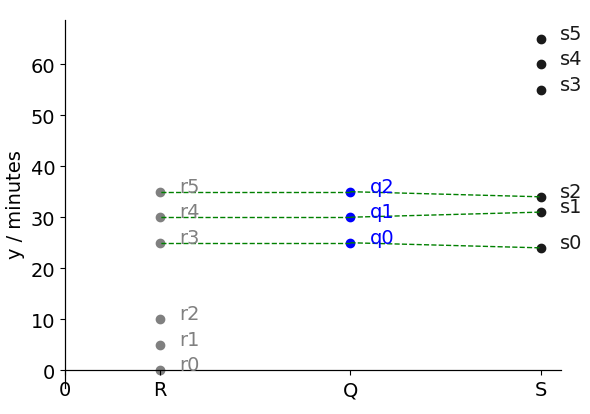


下面使用一个例子来讨论该方法在时间和空间的结合上出现的一些问题。为了减少时间和空间在数值上的影响，这里使用的时间和空间的数值尽量接近。并且为了简化运算，突出主要问题，这里使用欧氏距离代替论文中使用的路网距离。

假设我们有数据轨迹R和S，还有一条稍短的查询轨迹Q。在空间关系上面可以参考图一，x轴和y轴是空间的两个维度，图中可以看出轨迹R中至段和查询轨迹Q很相似，轨迹S中的至也和查询轨迹Q很相似，但是很明显在这一段轨迹R和Q的距离更近一点。关于时间维度可以参考图二，这里需要解释一下，图二的纵轴代表从8点开始的分钟数，比如的y值为0，代表的时间戳为8点整，处的y值为15，代表处的时刻为8点15分。单独从时间维度上看，轨迹R在至内的时间戳和轨迹Q的更加吻合，而轨迹S在至内的时间戳和轨迹Q稍有偏差。

解释完三条轨迹的时间和空间数据后，我们来使用PTM算法分别计算查询轨迹Q与R和S的相似性，关于参数设置如下，时间和空间阈值都取无穷大，每个样本点权重相等，时间和空间的权重一样，。先计算Q与R的PTM相似性，空间相似性为2.22，时间相似性为3，时空结合相似性为2.61。然后计算Q和S的PTM相似性，空间相似性为1.10，时间相似性为1.10，时空结合相似性为1.10。我们由PTM算法可以得出轨迹R和轨迹Q的相似度更大，将R作为查询结果返回。





现在我们分析一下这样计算存在的问题。我们可以回到原始查询请求以及轨迹数据的分析部分。根据查询轨迹，查询的目的应该是想找到一条轨迹，这条轨迹在8点25分左右出现在点，在8点30分左右出现在点，在8点35分出现在点。从描述上看，应该是轨迹S较为符合我们的查询需求。再看一看我们的查询结果R，对应时间下的空间位置与查询目的相去甚远，很显然不符合查询要求。

出现上述问题的原因是在PTM中，将时间和空间作为两个方面去考虑问题，空间上找距离Q最接近的样本点，计算空间相似度，然后在时间上也找距离Q最接近的样本点，计算时间相似度，这两步没有一点关联，导致空间上的样本点对应关系可能和时间上样本点的对应关系完全不同，最终导致了较差的查询结果。

因此，在考虑时间因素和空间因素的情况下分析轨迹数据相似性，应该将二者统一考虑，计算轨迹中的样本点的对应关系，这样查询得到的结果才符合轨迹相似性查询的要求。

### 3.1.2时空归一化方法

基于前面对PTM算法的讨论以及思考，得出了下面这个结论，即如果想同时考虑时间与空间因素对轨迹相似性造成的影响，那么必须在寻找轨迹样本点之间的对应关系的时候，同时考虑时间和空间在同一对样本点上的差异。

由于时间和空间属于两个维度，我们不能直接将时间纳入时空距离的计算中，需要使用一个将空间转化为时间的因子，用于将时间向空间距离做出转换，如公式所示，其中表示时间转化后的空间距离在时空距离计算中所占的比重，表示数据库中所有轨迹的平均速度，即所有轨迹总长度除以轨迹总的时间跨度。然后就可以使用该转化因子将时间t转化为空间中的z。这样结合二维的欧式空间的x和y两个维度，就可以将时间和空间归一化为一个三维的空间，其中x轴和y轴表示的是原本的空间的两个维度，z轴表示的是由时间转化过来的维度。

上述方法即为本文提出的时空归一化方法，可以将轨迹的时间和空间因素进行结合，得到的空间可以称为三维时空。使用该方法可以直接获得同一对样本点上的时间和空间上的差异，我们统称为时空距离。时空距离的定义与普通三维欧式空间的定义完全相同，点和的时空距离如公式所示。后面如果没有指明，则默认表示点和的时空距离。

其中，