1、评委老师好，我叫丁光伟，我的论文题目是“移动对象的时空轨迹相似性查询算法”。

2、这是我要介绍的主要内容。

3、首先介绍一下论文的研究背景和问题定义。

4、行人和车辆的移动信息被采样设备记录下来，得到的就是轨迹数据。通过轨迹数据挖掘，可以服务于LBS应用。论文所研究的“移动对象的时空轨迹相似性查询”是轨迹数据挖掘中的一项关键技术。

5、下面介绍问题定义。这张图介绍的是轨迹相似性查询的过程，首先由查询者输入查询轨迹，在收到查询轨迹之后，服务器会使用相似性查询算法，返回数据库中与查询轨迹相似的轨迹。已有的一些相似性计算方法都存在一系列问题，比如不能很好考虑时间、对应点匹配效果差、对采样策略敏感，后面的工作将解决这些问题。

6、第二部分介绍时空下的对应点匹配算法。

7、在传统方法中经常使用“基于时间信息的轨迹表示模型TTR模型”，包含的信息是空间坐标和时间戳。这个例子中可以看出TTR模型下，时间和空间是分开考虑的。

8、这篇论文中提出了PTM算法，用于查询包含时间信息的轨迹数据的相似性。PTM算法将时间相似性和空间相似性分开考虑，最后使用线性组合的方法得到时空相似性。

9、PTM算法分别考虑时间和空间，会导致时空对应关系的混乱。在这个例子中，给定查询轨迹Q与数据轨迹R和S，Q和R在空间上最接近的部分为r0至r2，在时间上最接近的部分为r3至r5。Q和S时间和空间上接近部分都是s0至s2。很明显，轨迹S在时间和空间上更接近轨迹Q，而轨迹R与Q只是某一段空间上接近，另一段时间上接近。但是分别考虑时间和空间的结果就是Q与R更相似。显然查询结果和查询需求不符合。

10、因此提出时空归一化表示模型SNTR模型，通过时空转化因素Ist将时间转化为三维坐标系的z轴，原来的欧式空间不做改变，作为坐标系xoy平面。SNTR模型的参数Ist根据具体查询需求而定，如果查询者在时间上要求严格，代表需要放大时间差距，Ist就应该是一个较大的值，如果查询者时间上并不是很严格，那么Ist可以小一点。SNTR模型的好处是：在后续的相似性查询过程中，不会产生时空对应关系的混乱，也会方便计算时间和空间距离计算。

11、时空归一化表示模型与TTR模型表示的信息相同，主要的区别在于使用时空坐标替换掉了TTR的空间坐标和时间信息。

12、介绍完SNTR模型，下面介绍一些SNTR模型下的对应点匹配。这幅图描述了对应点的概念。轨迹R和Q的距离是由R上的点到轨迹Q的距离组成，点到轨迹的距离就是点到另一条轨迹上一个点的距离，而另一条轨迹上的那个点，就是该点的对应点。

13、在DTW算法中，使用动态规划思想寻找对应点，优点是保持对应点匹配结果是时间有序的，即一个时间比较早的点的对应点的时间，一定早于一个时间比较晚的点的对应点。在BDS算法中，是寻找另一条轨迹上的最近位置作为对应点，优点对其效果好。

14、但是DTW算法的缺点是受样本点位置的限制，对齐效果差。BDS算法的缺点是没有考虑时间先后，造成匹配结果的时序错乱。

15、本文提出DTW-BDS对应点匹配算法，（读ppt中的小字部分。。。）

16、举个例子，

（1）在获得DTW算法匹配结果之后，首先进行r0的匹配，BDS的下界是q0，上界是q1；

（2）然后是中间点的匹配，优化了r1和r3的匹配结果。

（3）在优化了r3的匹配结果后，发现对应点由q1变成了DTW-BDS（r3），往后移了，因此r2的对应点的上下界变大了，需要以r3新对应点为上界，更新r2对应点。

（4）最后更新完r4和r5。

17、DTW-BDS对应点匹配算法利用了DTW算法保证了匹配结果的时序性，使用DBS算法提升匹配效果。

18、由对应点匹配可以得到对应轨迹段，数据轨迹相邻样本点称为轨迹段，riri+1的对应轨迹段有三种情况，分别是1、2、3

19、在对应点匹配后，得到了对应轨迹段。轨迹段间的距离计算由时空距离和形状影响组成，首先介绍时空距离的计算方法。

20、采样设备的采样策略不尽相同，几分钟采样一次或者几米采样一次会产生不同的轨迹数据，如图所示。而一些算法对采样策略的变化很敏感，相同的移动对象产生的轨迹，由于采样策略的不同，可能会输出完全相反的结果。

21、上述问题的关键在于：之前的算法对样本点过于依赖，忽略了采样策略造成的影响，为了解决这个问题，这里提出断点的概念。在轨迹段riri+1上以固定大小均匀取点，获得断点，固定大小称为断点阈值。在后面计算轨迹段距离时，通过断点可以更细粒度地考虑轨迹段的特征。通过DTW-BDS对应点匹配算法，可以获得断点的对应点。

22、当断点阈值无穷小时，相邻断点到对应点的距离可以看做相等。变大后，消失的断点可以使用周围保留下的断点代替，然后根据替代的断点的个数，赋予每个断点一个权重。轨迹段的时空距离等于轨迹段上端点和断点乘上权重。这种轨迹段距离的计算方法由于考虑了更多轨迹段的信息，受采样策略的影响会更小。

23、下面介绍轨迹段距离的第二部分，形状影响。我们可以从直观上感受到，轨迹段长度相同时，夹角越小，轨迹段之间越相似。夹角相同时，两条轨迹段长度越长就越相似。因此这里使用数学中的余弦距离和投影相关知识，来量化轨迹段的形状相似。

24、形状相似性simshape就是riri+1投影的长度与Q(ri)Q(ri+1)的最小值，论文中还给出了其他两种对应轨迹段情况下的计算方法。使用sigmoid函数的变种，将形状相似性转换为形状影响权值Ishape，变化范围在1以内。通过参数mu调整形状的影响程度。

25、轨迹段距离由时空距离和形状影响权值相乘得到。轨迹段距离描述了对应轨迹段夹的时间距离、空间距离和形状距离，距离越大，轨迹段越不相似。

26、下面介绍轨迹的距离计算流程。首先对Q和R进行DTW-BDS对应点匹配，获得对应轨迹段，然后通过对应轨迹段的时空距离和形状影响权值，得到轨迹段间的距离，最后所有轨迹段距离之和就是Q和R的轨迹距离。

27、根据前面提出的SNTR模型，和刚才的轨迹距离计算方法，这里提出STS时空轨迹相似性查询算法。首先进行SNTR模型构建，然后遍历计算数据库中的轨迹数据，当与查询轨迹的距离小于阈值，就认为与查询轨迹相似。与前任的算法相比，STS算法包含了更多的优势，比如可以抵抗采样策略，匹配具有时序性等。

28、下面是实验与分析部分。

29、这是实验环境的描述，使用了两个轨迹数据集，一个是微软的真实轨迹数据集GL，一个是在北美路网上合成的轨迹数据集NARN。

30、首先对各个参数的影响进行实验。这里使用与查询轨迹距离最小的top-k条轨迹作为查询结果返回，纵轴代表算法的查准率，就是查询结果中，真正与查询轨迹相似的比例。SNTR模型的时空转化参数的大小代表时间的重要程度，GL数据集中，Ist在5-20范围内效果较好，NARN数据集中，Ist在10-30范围内，效果较好。因为不同轨迹数据的平均速度不同，当速度快时，Ist就应该相应增大。还对断点阈值，形状敏感度参数，以及轨迹距离阈值进行了实验，证明了提出的SNTR模型，时空距离计算等是有效的。

32、然后进行了不同算法之间的对比实验。证明了在查询轨迹不同长度的情况下，STS算法的查准率适中较高，并且较稳定。

33、为了验证SNTR模型的有效性，在DTW算法中使用SNTR模型，发现查准率有明显提升。

34、最后测试了不同算法的抗噪声干扰能力，结果表明STS算法比较稳定，受噪音影响较小。

36、总结