

# 移动对象的 时空轨迹相似性查询算法

姓名： 丁光伟  
导师： 杨晓春 教授  
学号： 1600894  
专业： 计算机系统结构

# 目录

- > 研究背景介绍
- > 三维时空下样本点匹配算法
- > 时空轨迹相似性查询算法
- > 实验与分析
- > 总结

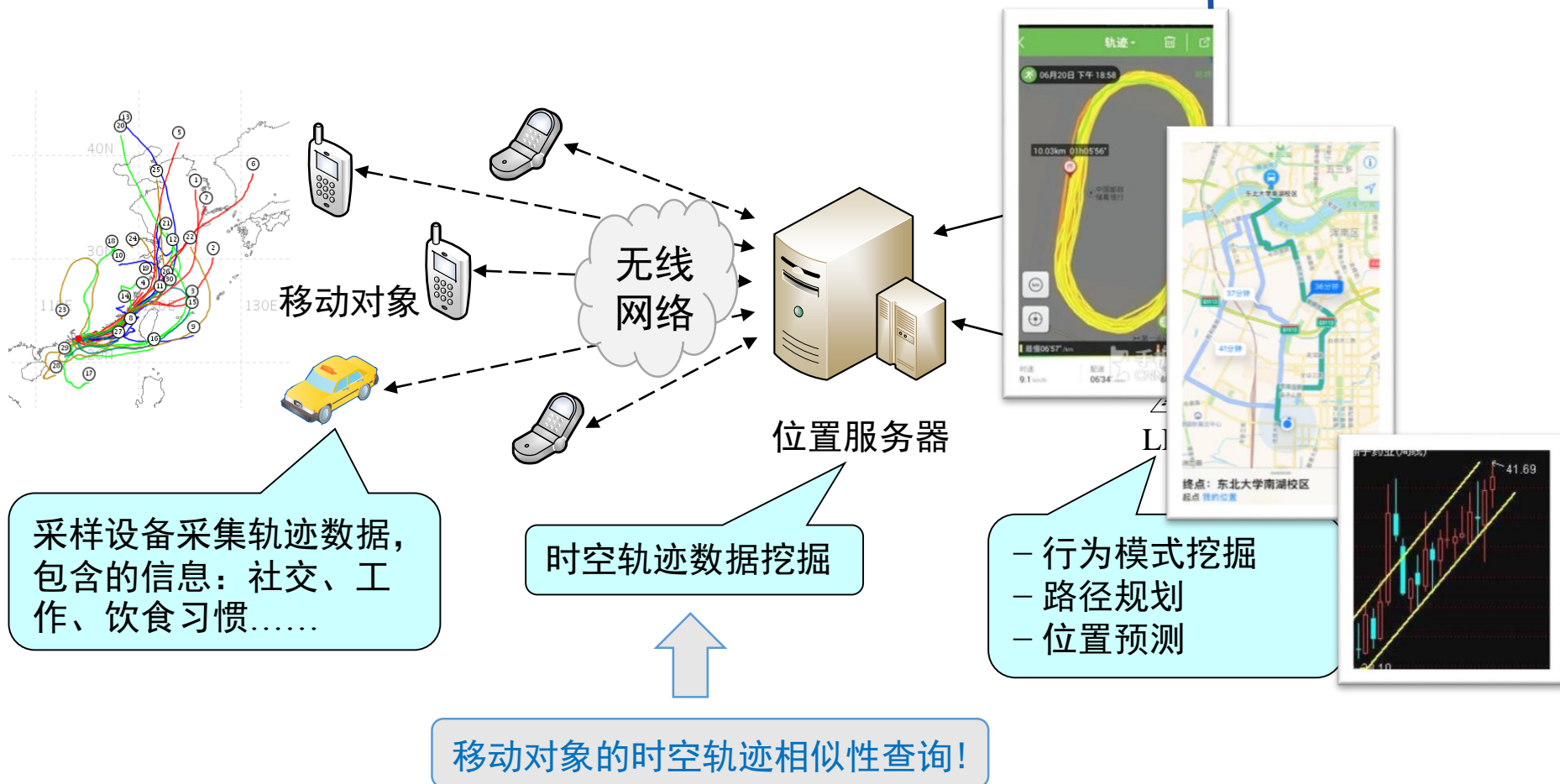
01

# 研究背景介绍

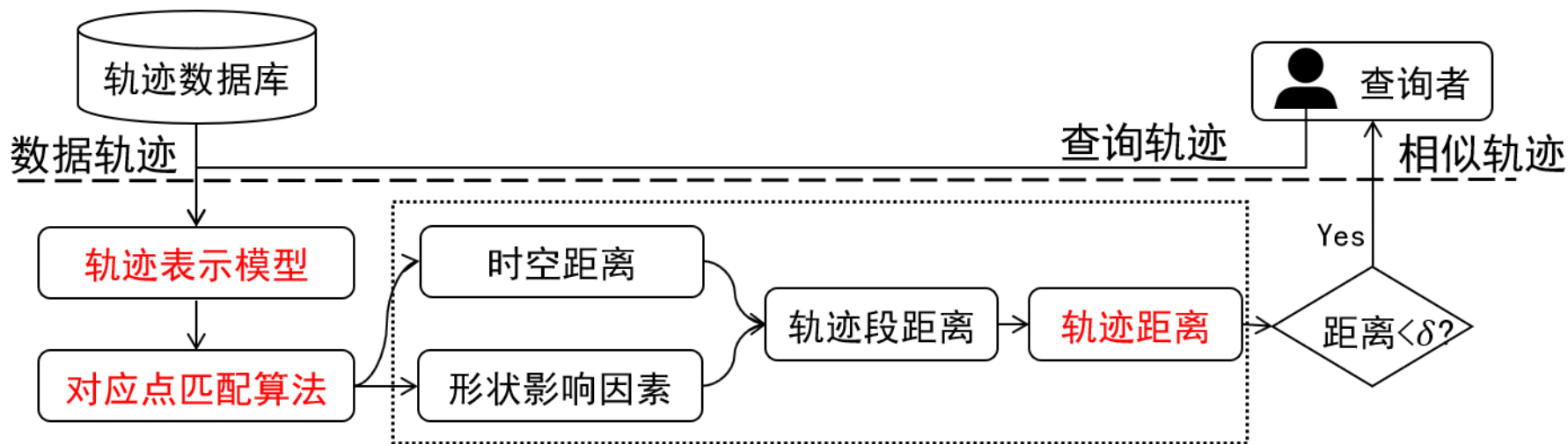
研究背景

问题定义

# 研究背景



# 问题定义



- 1、时空混乱
- 2、对应点匹配效果差
- 3、对采样策略敏感

已有的轨迹  
相似性计算方法

	空间	时间+空间
欧式空间	DTW、BDS、EDR	SDTW
路网	DMS、DMSS	PTM、DMT

02

# 时空下对应点 匹配算法

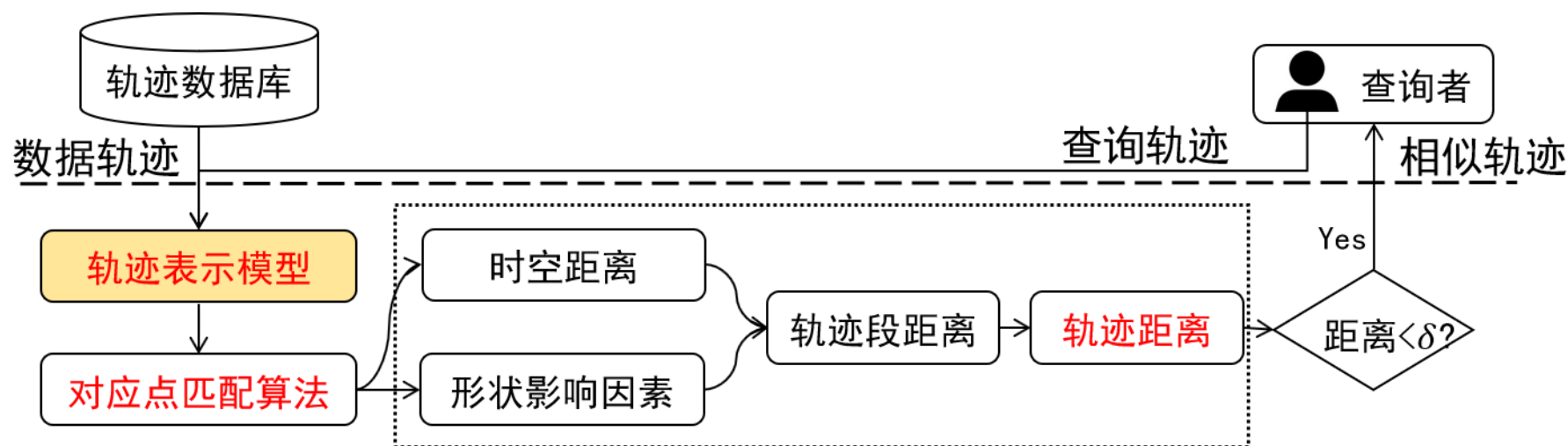
时空归一化表示模型

DTW-BDS样本点匹配

对应点

对应轨迹段

# 时空归一化表示模型



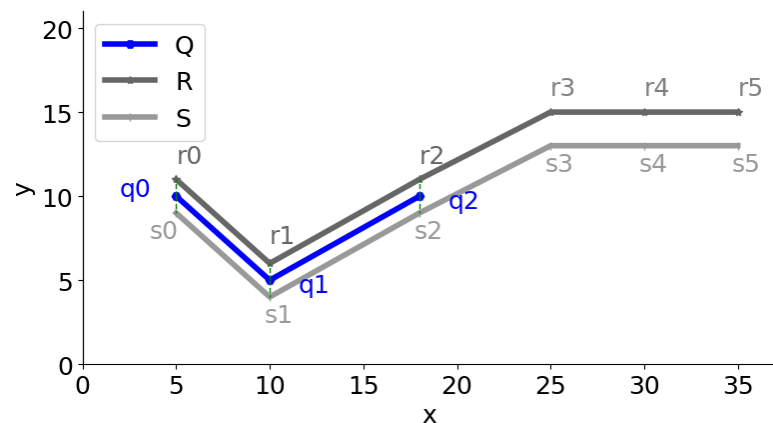
# 时空归一化表示模型

基于时间信息的  
轨迹表示模型 (TTR)

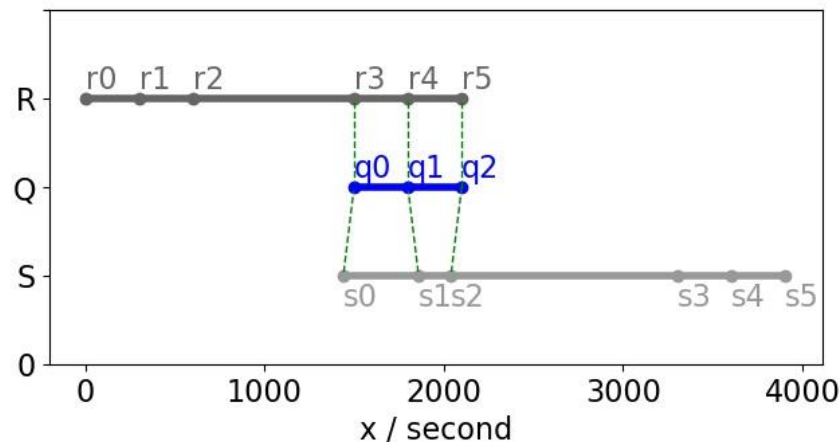
序号	时间戳	空间坐标
0	$t_0$	$(lat_0, lng_0)$
1	$t_1$	$(lat_1, lng_1)$
2	$t_2$	$(lat_2, lng_2)$
3	$t_3$	$(lat_3, lng_3)$
4	$t_4$	$(lat_4, lng_4)$
5	$t_5$	$(lat_5, lng_5)$



空间

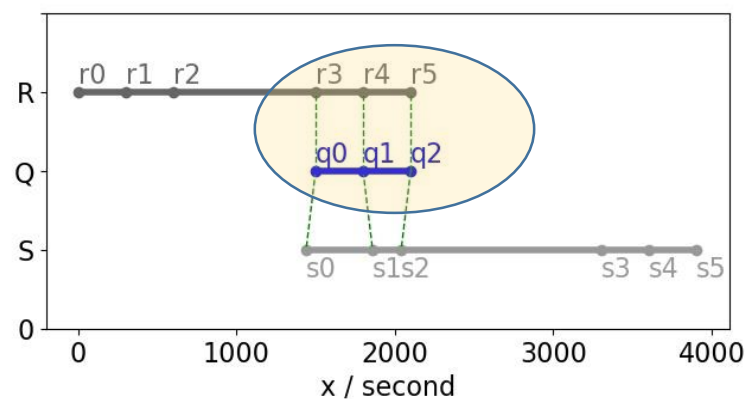
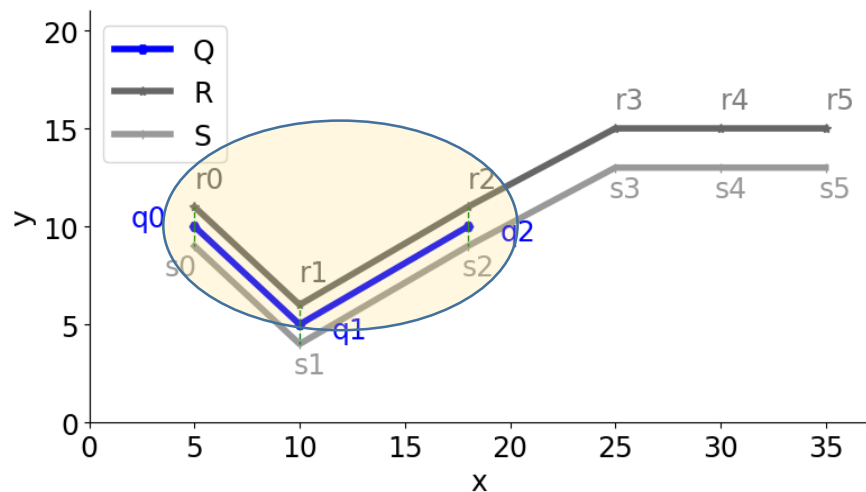
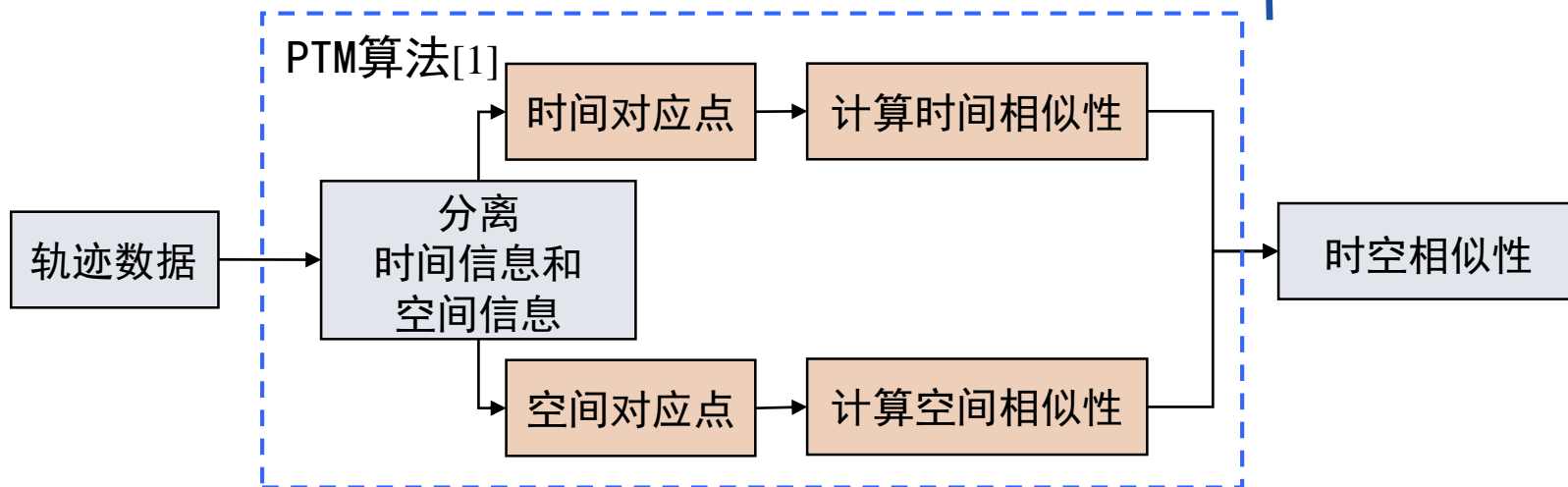


时间





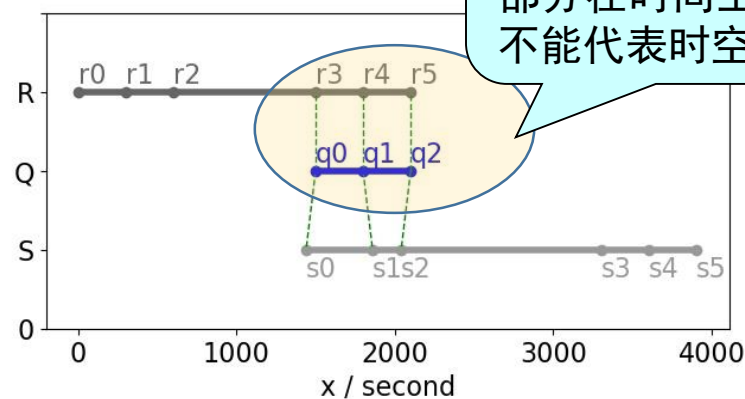
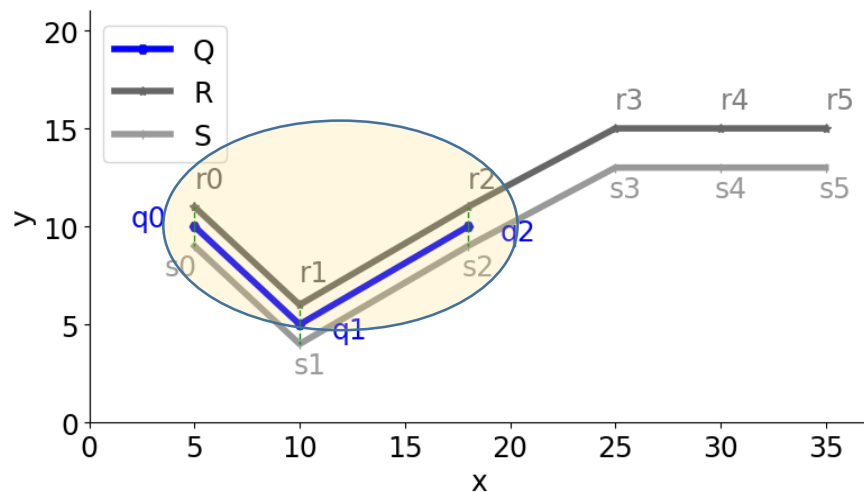
# 时空归一化表示模型



[1] Shang S, Ding R, Zheng K, et al. Personalized trajectory matching in spatial networks[J]. Vldb Journal, 2014, 23(3):449-468.

# 时空归一化表示模型

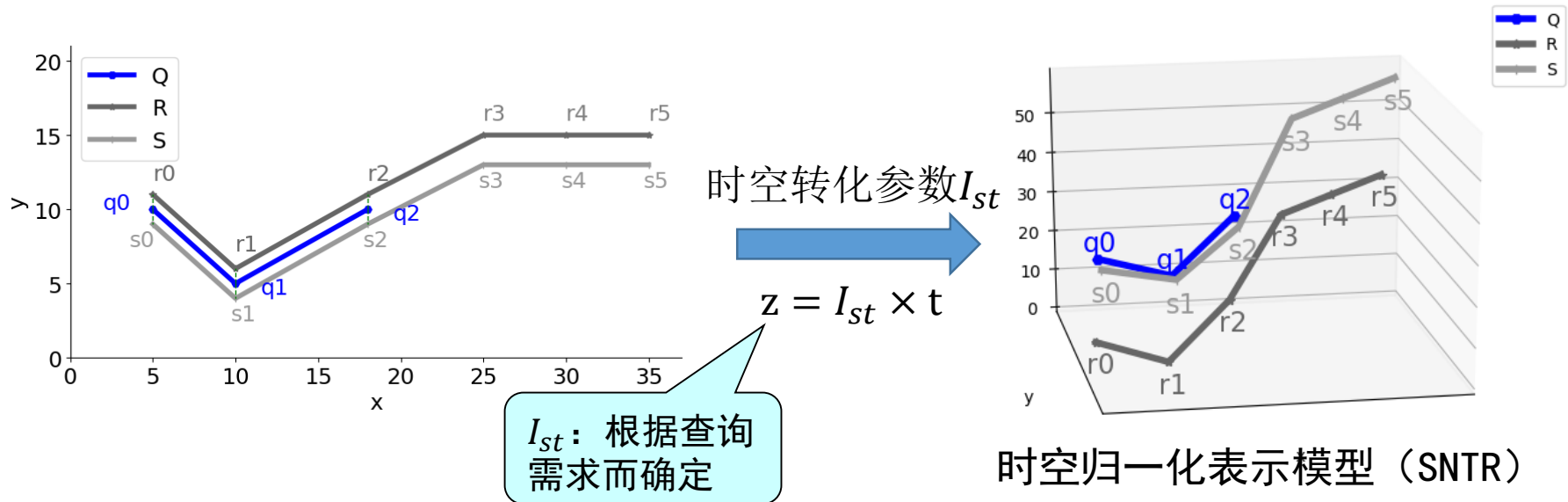
## PTM算法[1]



Q与R的一部分在空间上相似，与R的另一部分在时间上相似，不能代表时空相似

[1] Shang S, Ding R, Zheng K, et al. Personalized trajectory matching in spatial networks[J]. Vldb Journal, 2014, 23(3):449-468.

# 时空归一化表示模型



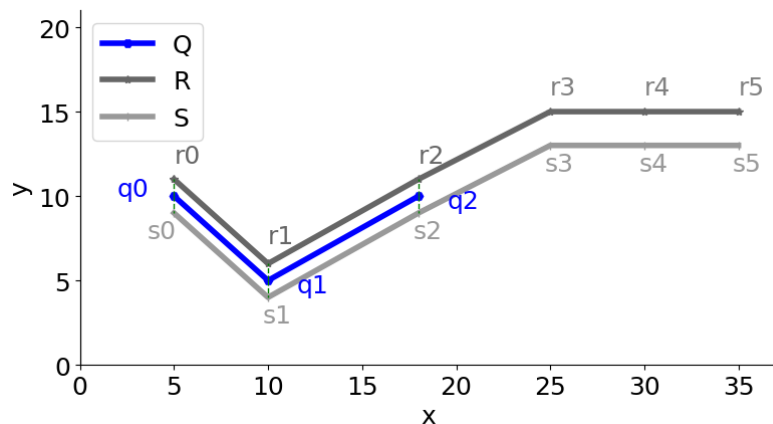
例1: 根据小偷逃跑所驾驶车辆轨迹, 查询是否有过往车辆拍摄到小偷抛弃赃物, 城市道路环境下, 摄像头最远拍清200m, 且以小偷的行车速度12.5s内可领先200m, 则  $I_{st}=16$

例2: 利用用户A平时上班轨迹寻找顺风车, A允许与自己空间距离相距500m, 时间上允许相差5分钟, 则  $I_{st}=1.67$

## SNTR模型的好处(与TTR模型相比):

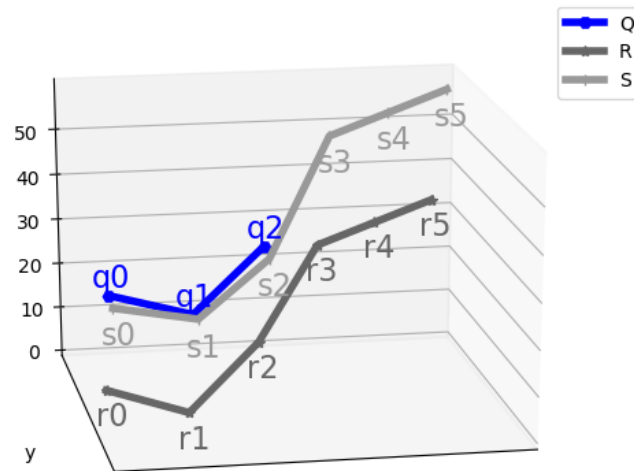
- 相似性查询过程中, 不会产生时空对应关系混乱
- 便于时空距离计算

# 时空归一化表示模型



时空转化参数  $I_{st}$

$$z = I_{st} \times t$$

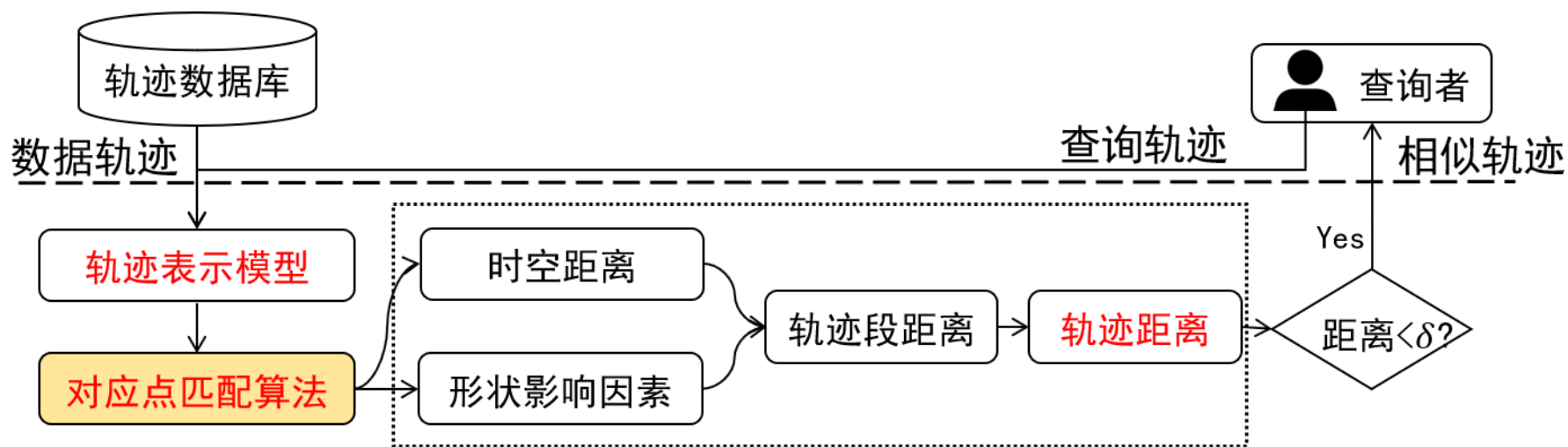


时空归一化表示模型 (SNTR)

表1 轨迹表示模型对比

表示模型	空间坐标	时间信息	语义信息	网格标号	时空坐标
STR	✓	—	—	—	—
TTR	✓	✓	—	—	—
TETR	✓	—	✓	—	—
CTR	—	✓	—	✓	—
SNTR	—	—	—	—	✓

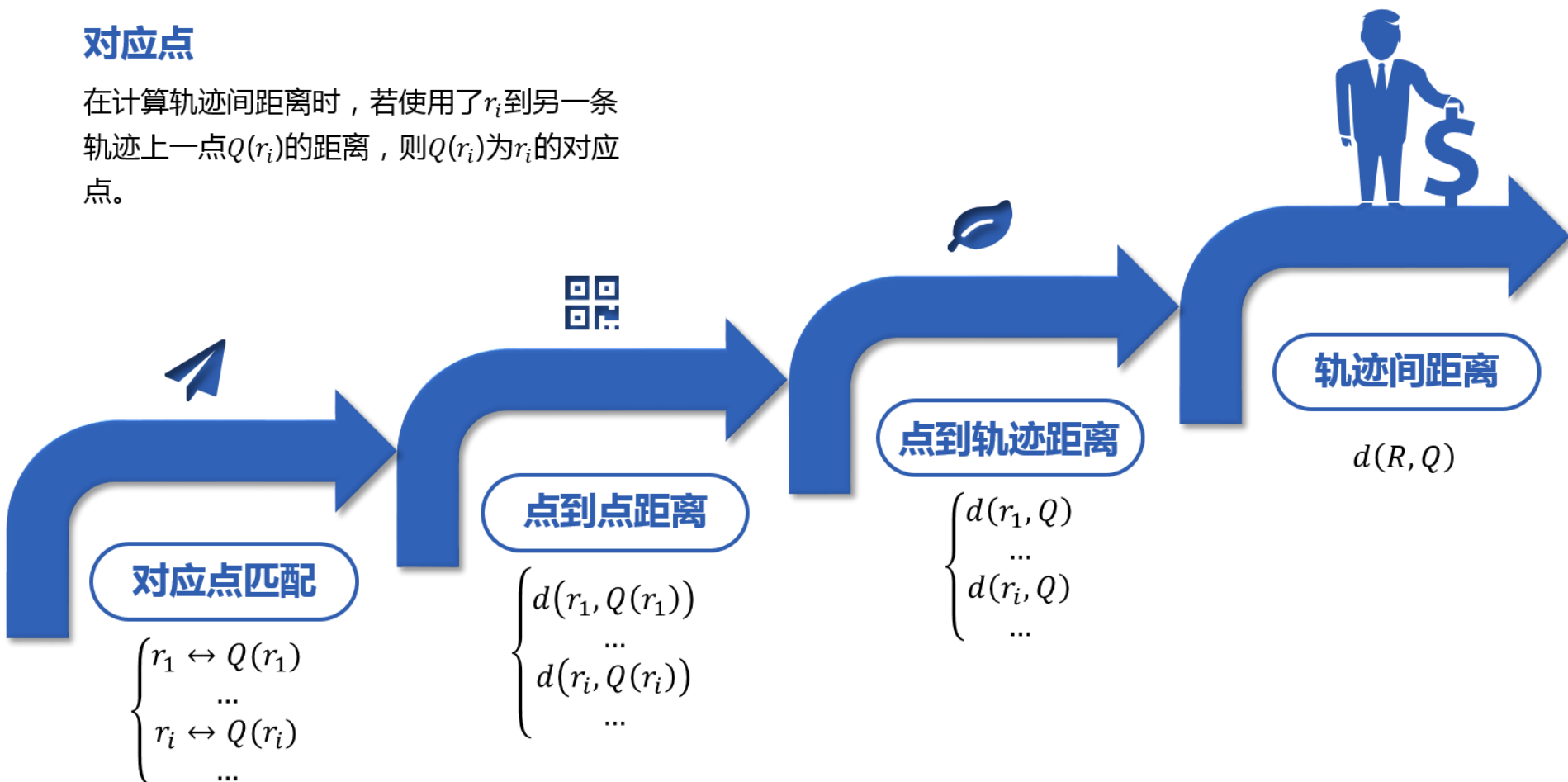
# DTW-BDS对应点匹配



# DTW-BDS对应点匹配

## 对应点

在计算轨迹间距离时，若使用了 $r_i$ 到另一条轨迹上一点 $Q(r_i)$ 的距离，则 $Q(r_i)$ 为 $r_i$ 的对应点。

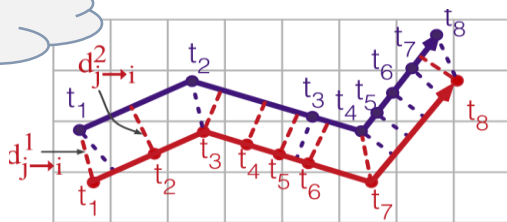


# DTW-BDS对应点匹配

Yi B K et  
al. ICDE'1998

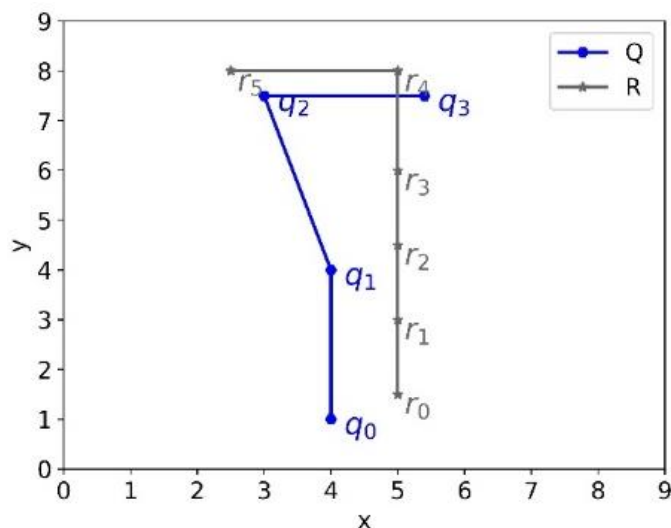


Na T et  
al. TKDE'2017

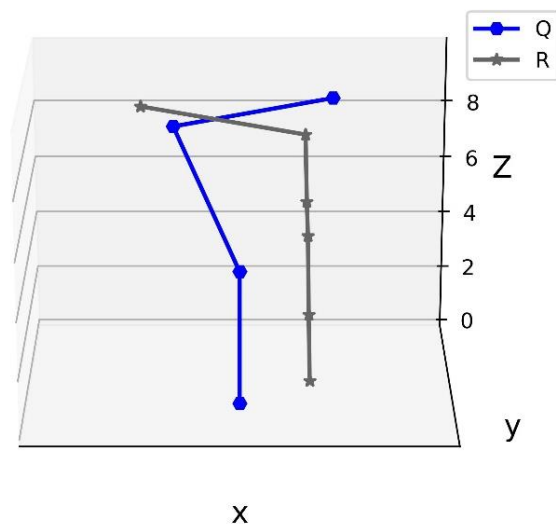


DTW: 动态规划思想寻找整体最优的匹配方案  
优势: 保持时间有序

BDS: 寻找另一条轨迹上最近位置作为对应点  
优势: 样本点更好地对齐



时空  
归一化



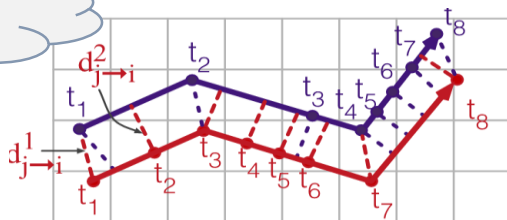
# DTW-BDS对应点匹配

Yi B K et  
al. ICDE'1998

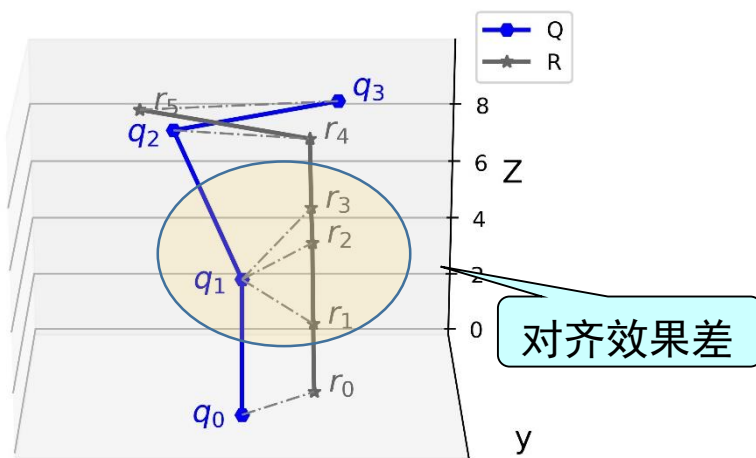


DTW: 动态规划思想寻找整体最优的匹配方案  
优势: 保持时间有序  
缺点: 只能与样本点匹配, 对齐效果差

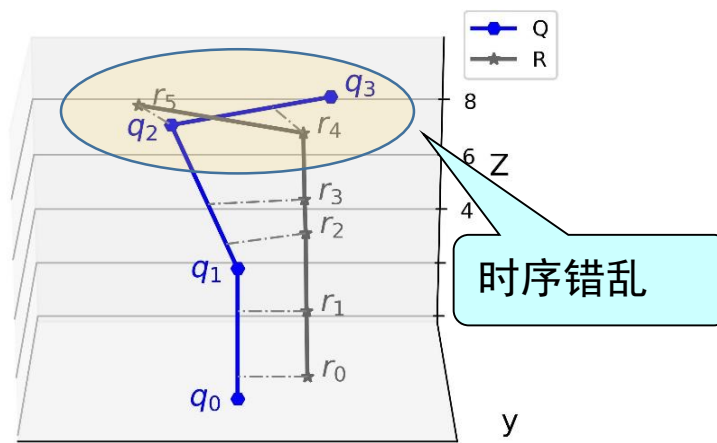
Na T et  
al. TKDE'2017



BDS: 寻找另一条轨迹上最近位置作为对应点  
优势: 样本点更好地对齐  
缺点: 未考虑时间, 对应点匹配结果时序错乱



DTW匹配结果

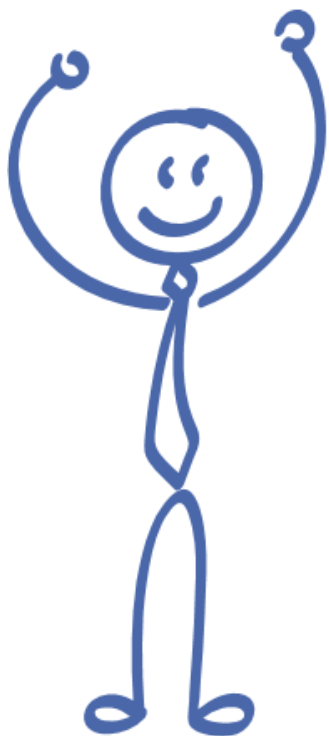


BDS匹配结果



# DTW-BDS对应点匹配

## DTW-BDS对应点匹配



### 获取DTW对应点

首先使用DTW算法，获取所有样本点之间的匹配关系



### BDS算法更新对应点

以相邻样本点的对应点为BDS算法上下界，按序更新DTW对应点

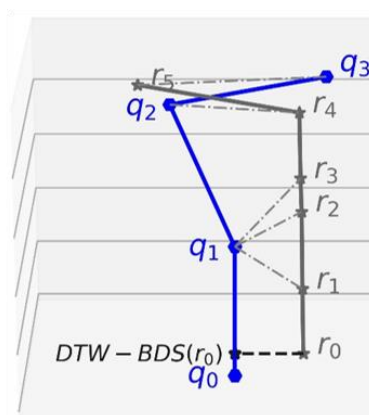


### 前向更新

若新对应点时间戳后移，则需要向前向更新

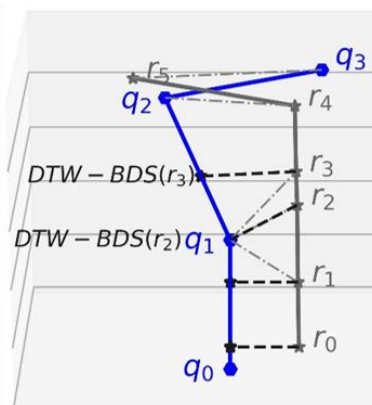
# DTW-BDS对应点匹配

## DTW-BDS对应点匹配过程



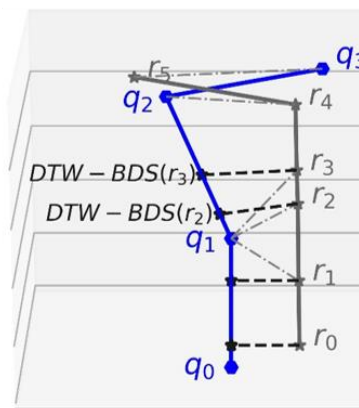
第一个点匹配

$DTW-BDS(r_0) =$   
 $BDS(r_0, q_0, DTW(r_1).first)$



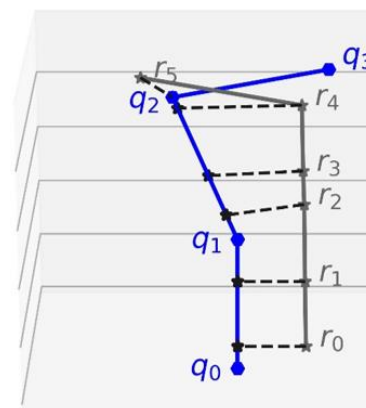
中间点匹配

$DTW-BDS(r_i) =$   
 $BDS(r_i, DTW-BDS(r_{i-1}),$   
 $DTW(r_{i+1}).first)$



前向更新

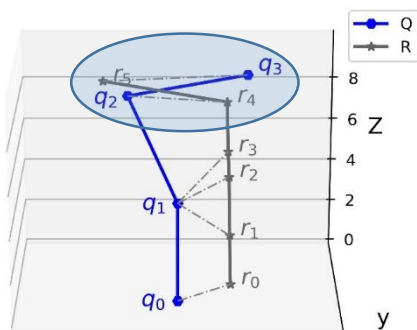
$r_3$ 的对应点更新后，时间戳后移。需要使用DTW-BDS( $r_3$ )作为新上界，更新 $r_2$ 的对应点



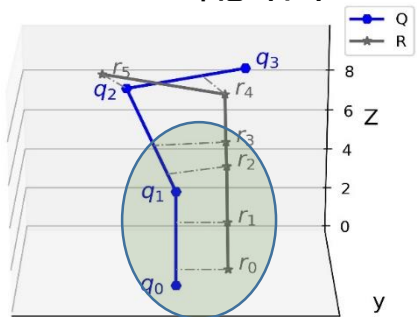
保持时序性

逆向部分的匹配结果也能保证时序性

# DTW-BDS对应点匹配

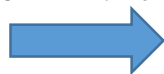


DTW匹配结果

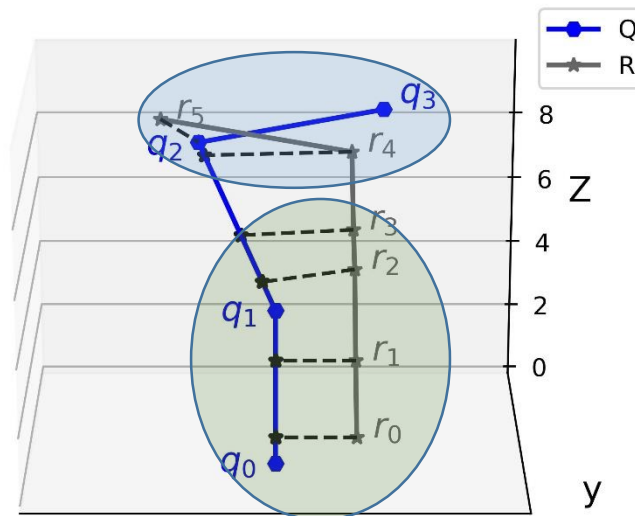


BDS匹配结果

保证时序性



提升对齐效果



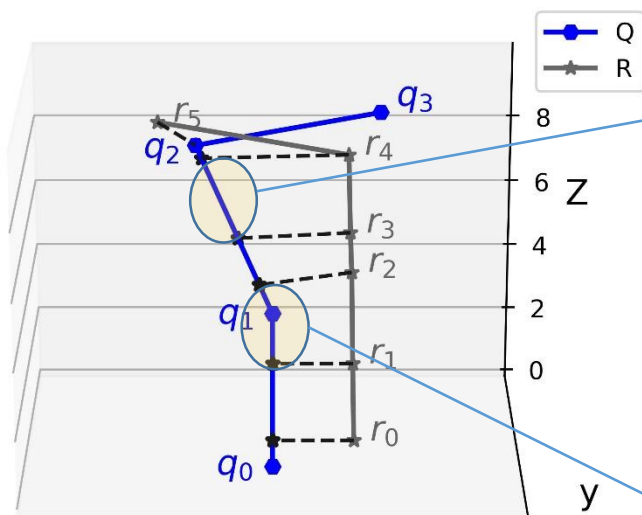
DTW-BDS匹配结果

DTW-BDS算法优势:

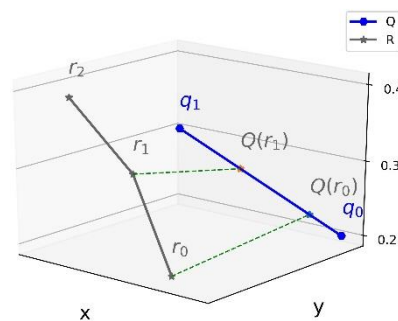
- 匹配结果保持时序性
- 对齐效果好

# DTW-BDS对应点匹配

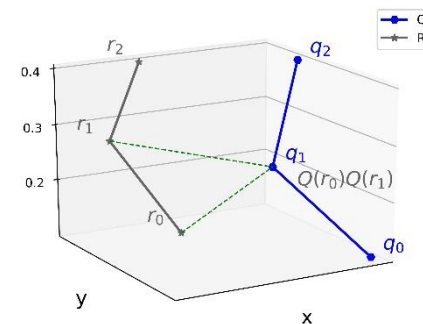
## 对应轨迹段的三种情况



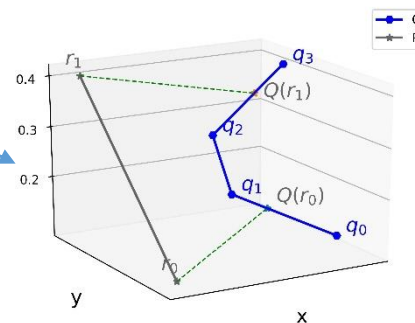
DTW-BDS匹配结果



1、不包含样本点(直线段)



3、单个点



2、包含样本点(折线段)

03

# 时空轨迹 相似性查询算法

时空距离

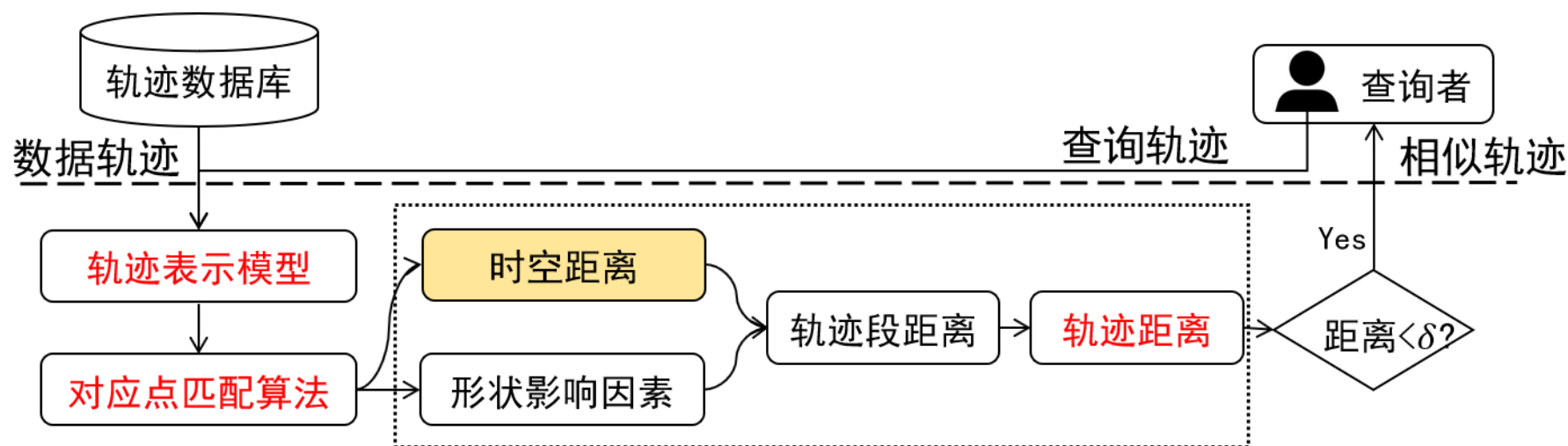
形状影响

轨迹段距离

轨迹距离

轨迹相似性查询

# 时空距离



# 时空距离

Yi B K et al. ICDE'1998

## DTW相似性计算矩阵

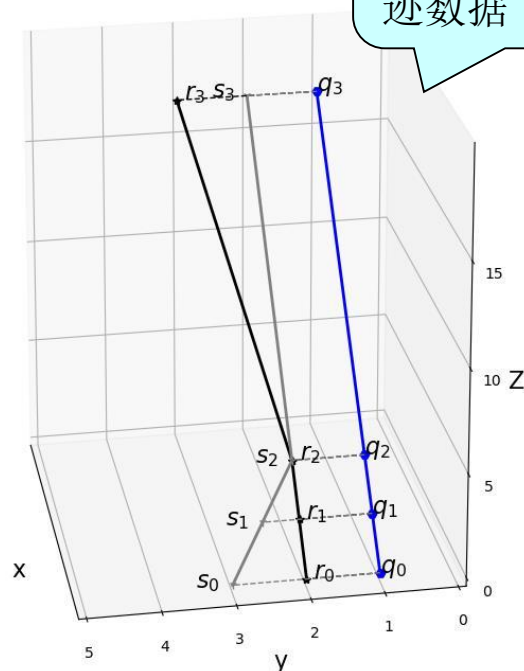
Q & R

	0	$q_0$	$q_1$	$q_2$	$q_3$
0	0	$+\infty$	$+\infty$	$+\infty$	$+\infty$
$r_0$	$+\infty$	1	3.24	7.36	24.48
$r_1$	$+\infty$	3.24	2	4.24	19.37
$r_2$	$+\infty$	7.36	4.24	3	16.15
$r_3$	$+\infty$	24.39	19.27	16.04	5

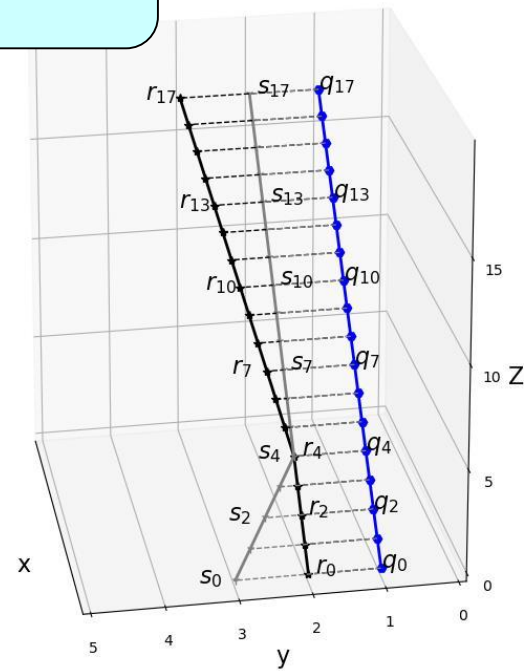
Q & S

	0	$q_0$	$q_1$	$q_2$	$q_3$
0	0	$+\infty$	$+\infty$	$+\infty$	$+\infty$
$s_0$	$+\infty$	2	4.5	8.62	25.65
$s_1$	$+\infty$	4.83	3.5	5.74	20.77
$s_2$	$+\infty$	9.30	6	4.5	17.54
$s_3$	$+\infty$	26.42	21.07	17.54	5.5

采样策略不同会记录下不同的轨迹数据



(a) 稀疏样本点



(b) 密集样本点

$DTW(Q, R) < DTW(Q, S)$

$\neq$

$DTW(Q, R) > DTW(Q, S)$

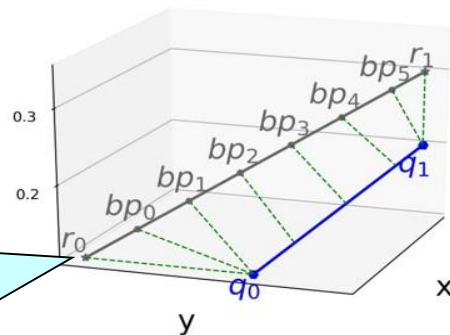
轨迹距离计算存在问题:

只依赖样本点的算法对采样策略较敏感, 相同的移动会导致不同查询结果

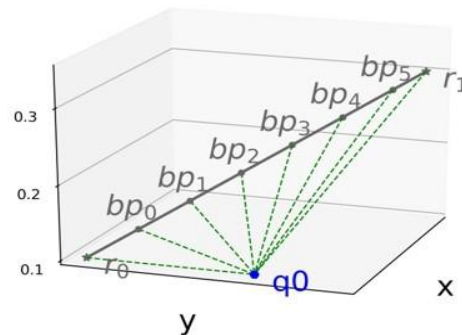
# 时空距离

断点 (break point)

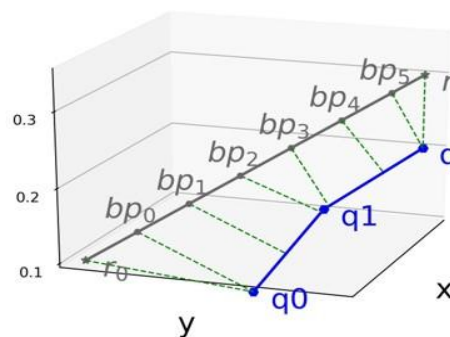
在轨迹段  $r_i r_{i+1}$  上以固定的断点阈值  $\eta$  均匀取点, 获得断点



(a)



(b)



(c)

断点的对应点

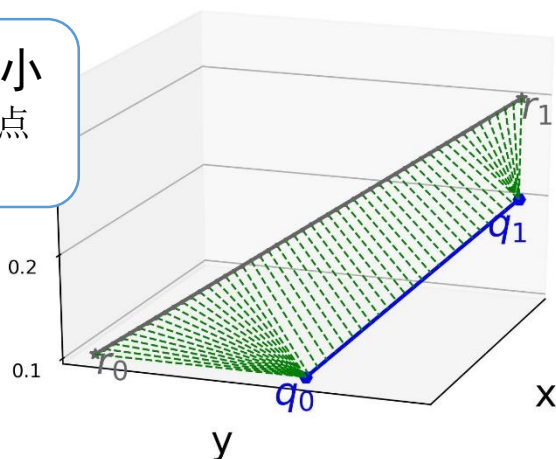
断点的作用:

通过人为增加采样点, 更细粒度地考虑轨迹段  $r_i r_{i+1}$  的时空特征, 使计算结果更准确。



# 时空距离

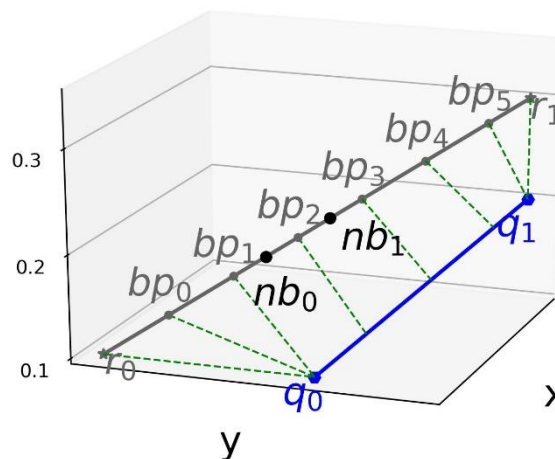
$\eta$ 趋向于无穷小  
相邻断点到对应点的  
距离近似相等



点	$r_0$	$bp_0$	$bp_i$	$bp_5$	$r_1$
长度		$\eta$	$\eta$	$\eta$	$d(bp_5, r_1)$
权重	1/2	1	1	$\frac{1}{2} + \frac{d(bp_5, r_1)}{2\eta}$	$\frac{d(bp_5, r_1)}{2\eta}$

$\eta$ 变大后

在一定精度范围内,  
使用 $bp_2$ 代替从 $nb_0$   
到 $nb_1$ 间的所有消  
失的断点



断点权重与代  
替的对应点个  
数成正比

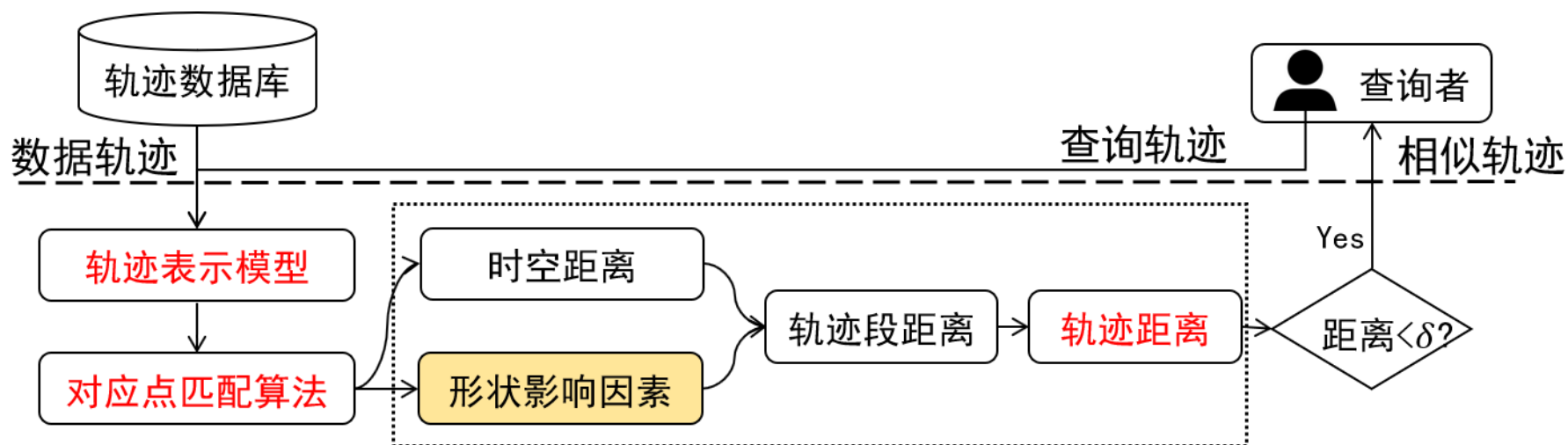
时空距离 =  $\sum$  (点与对应点间距离  $\times$  权重)

$$d_{st}(r_i r_{i+1}, Q(r_i) Q(r_{i+1})) = \sum_{v_j \in \{r_i, bp_0 \dots r_{i+1}\}} w_j \times d(v_j, Q(v_j))$$

轨迹段时空距离:

使用断点考虑了更多轨迹段信息, 受采样策略影响更小

# 形状影响因素



# 形状影响因素

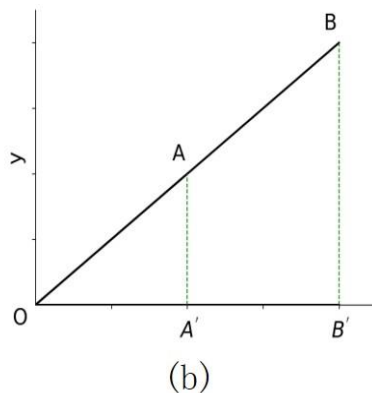
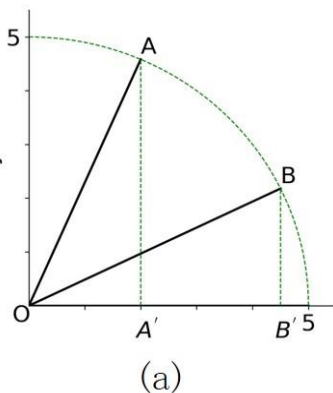
- 轨迹段长度相同，对应轨迹段间夹角越小形状越相似
- 相同夹角，轨迹段长度越长形状越相似

余弦距离:  $\cos(\theta) = \frac{\vec{a} \cdot \vec{b}}{|\vec{a}| |\vec{b}|}, \theta \in [0, \pi]$

投影:  $|a'| = |\vec{a}| \cos(\theta) = \frac{\vec{a} \cdot \vec{b}}{|\vec{b}|}$

数学知识

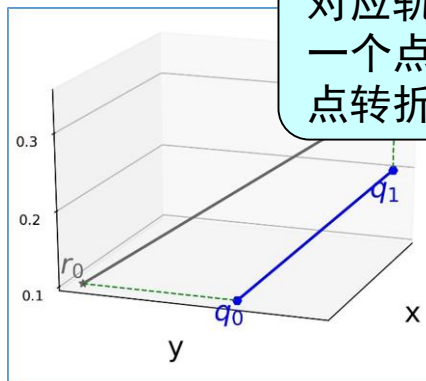
从投影的值可以看出轨迹段间形状相似程度



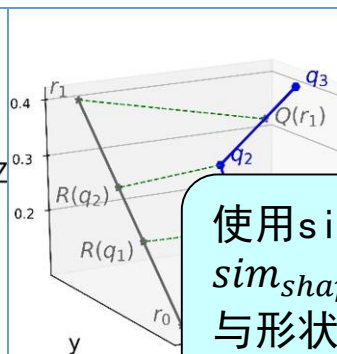
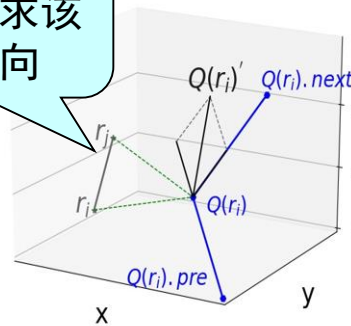
# 形状影响因素

## 形状相似性

$$sim_{shape}(r_i r_{i+1}, Q(r_i)Q(r_{i+1})) = \min \begin{cases} d(r_i, r_{i+1}) * \cos(\overrightarrow{r_i r_{i+1}}, \overrightarrow{Q(r_i)Q(r_{i+1})}) \\ d(Q(r_i), Q(r_{i+1})) \end{cases}$$



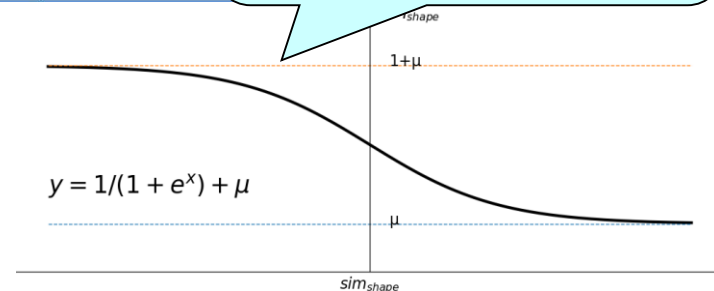
对应轨迹段为一个点，求该点转折方向



对应轨迹段为折线段，分别求局部对应轨迹段

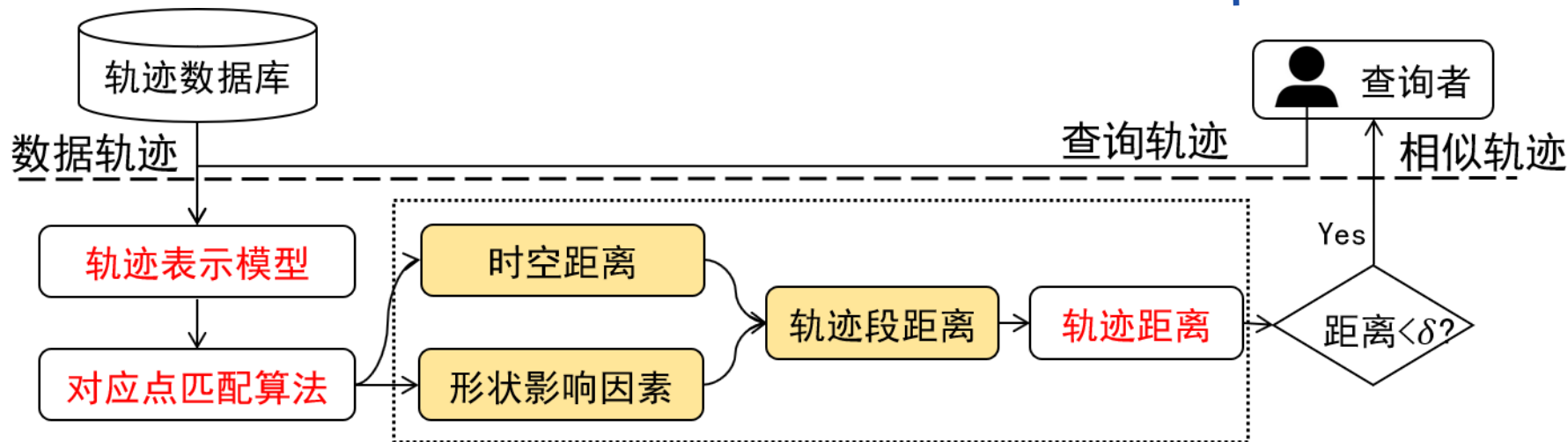
使用sigmoid阈值函数将  $sim_{shape}$  转化为  $I_{shape}$ ，与形状相似呈负相关

## 形状影响权值



$$I_{shape}(r_i r_{i+1}, Q(r_i)Q(r_{i+1})) = \frac{1}{1 + e^{sim_{shape}(r_i r_{i+1}, Q(r_i)Q(r_{i+1}))}} + \mu$$

# 轨迹段距离



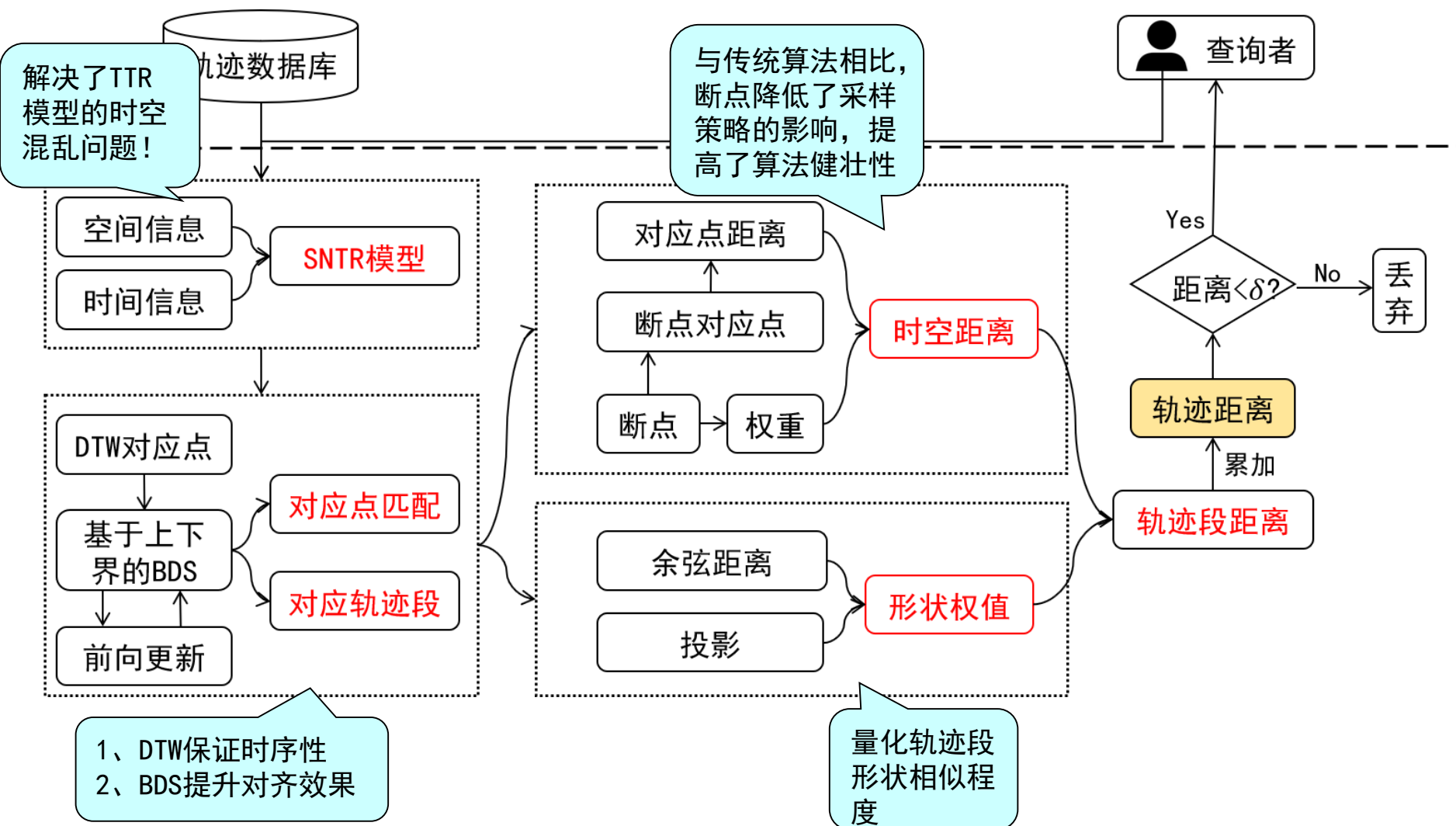
轨迹段距离 = 时空距离 × 形状影响权值

$$d_{segment}(r_i r_{i+1}, Q(r_i) Q(r_{i+1})) = I_{shape}(r_i r_{i+1}, Q(r_i) Q(r_{i+1})) \times d_{st}(r_i r_{i+1}, Q(r_i) Q(r_{i+1}))$$

$d_{segment}$  描述了对应轨迹段间:

- 时空距离, 距离越远,  $d_{segment}$  越大
- 形状上的相似程度, 形状越相似,  $d_{segment}$  越小

# STS轨迹相似性查询算法



04

# 实验与分析

参数的影响

查询轨迹长度的影响

三维时空有效性研究

噪音的影响

# 实验设置

## 实验环境描述

类别	描述
CPU	Intel (R) Core(TM) i7-6700 3.40 GHz
硬盘	8.00 GB
内存	1T
操作系统	Microsoft Windows 7(64位)
IDE	JetBrains PyCharm
编程语言	python
相关开发包	numpy, matplotlib, mpl_toolkits

## 轨迹数据集描述

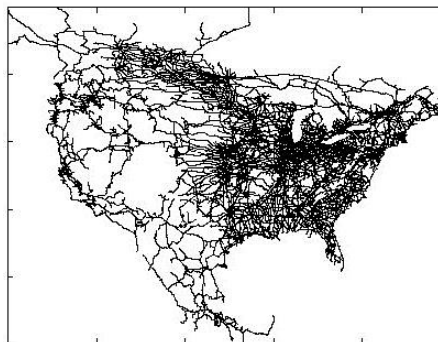
轨迹数据集	轨迹条数	平均轨迹点数量
GeoLife (GL)	17621	843.1
North America Road Network (NARN)	20000	61.6 (10-150)

### 真实轨迹

北京行人车辆轨迹  
微软亚研院  
182个志愿者  
5年收集时间



(a) GL数据集



(b) NARN数据集

### 软件合成轨迹:

<http://www.cs.utah.edu/~li feifei/SpatialDataset.htm>



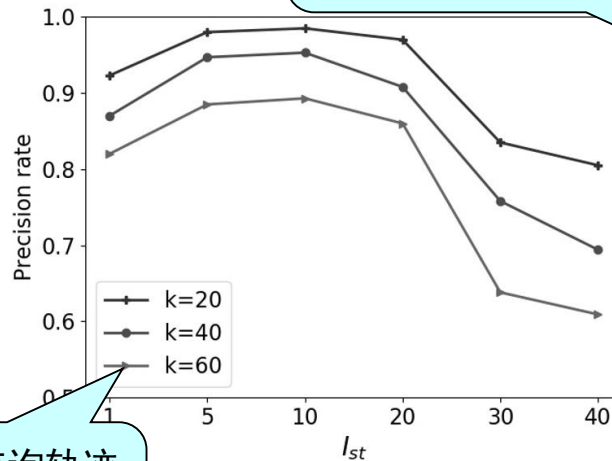
# 参数的影响

纵轴代表查准率，即查询结果中，真正与查询轨迹相似的比例。

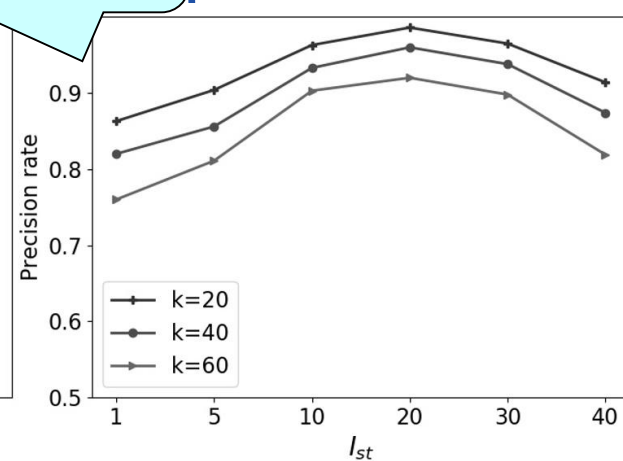
## (1) SNTR模型中时空转化参数对查询结果的影响

时空转化参数代表时间的重要程度。太小会忽视时间差距的影响，太大会放大时间的重要性，忽略空间上的差距。因此需要适中的数值。

$k$ 表示与查询轨迹距离最小的top- $k$ 条轨迹。



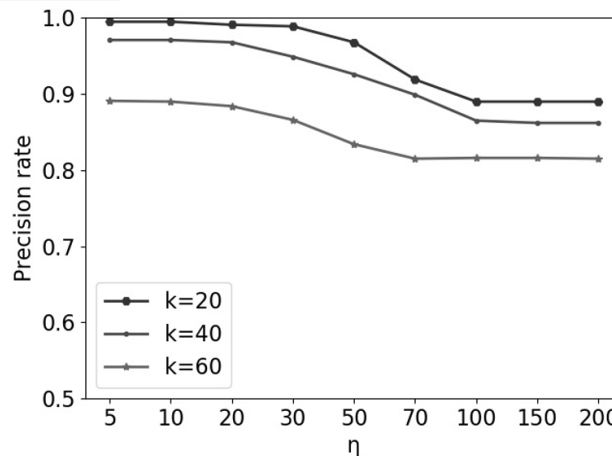
(a) GL数据集



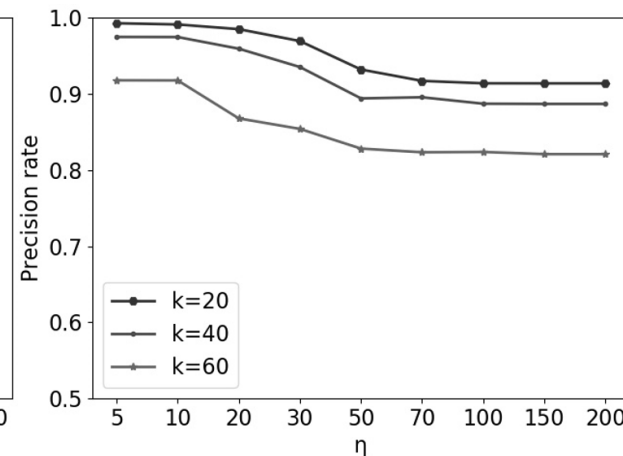
(b) NARN数据集

## (2) 断点阈值对查询结果的影响

断点阈值越大，轨迹段中断点数目越少，断点的作用越小，查准率越低



(a) GL数据集

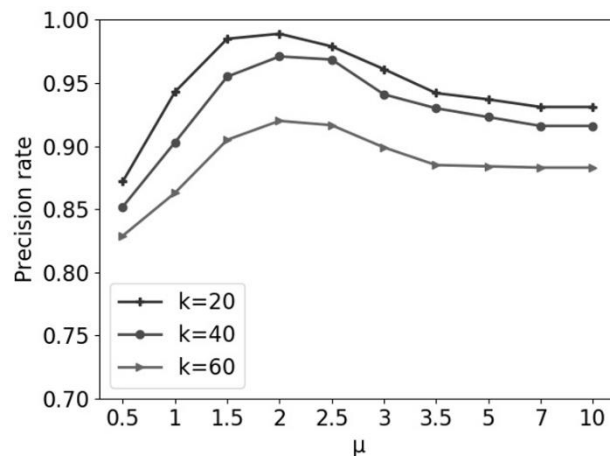


(b) NARN数据集

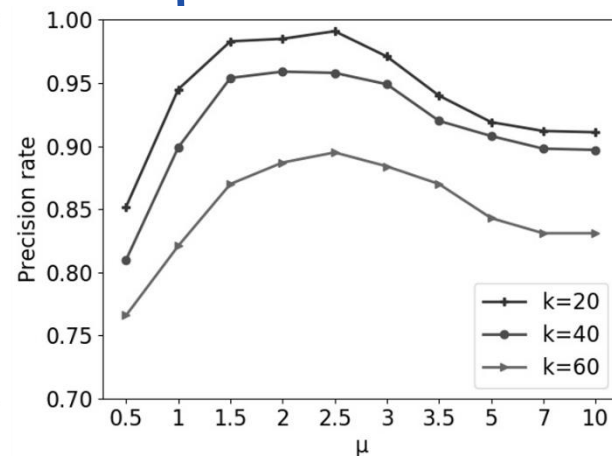
# 参数的影响

## (3) 形状敏感度参数对查询结果的影响

形状敏感度参数太小会导致形状因素占比重过大，太大会导致形状不起作用



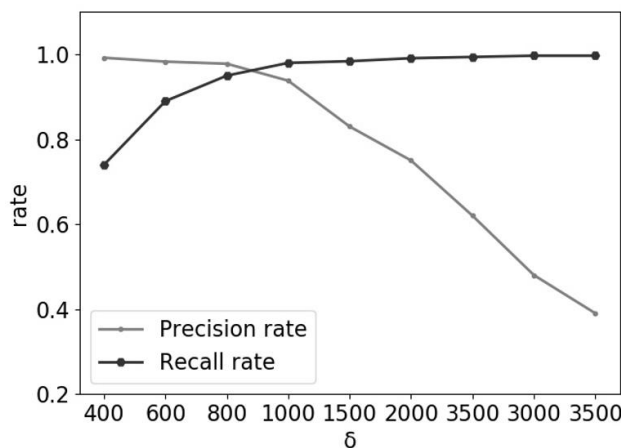
(a) GL数据集



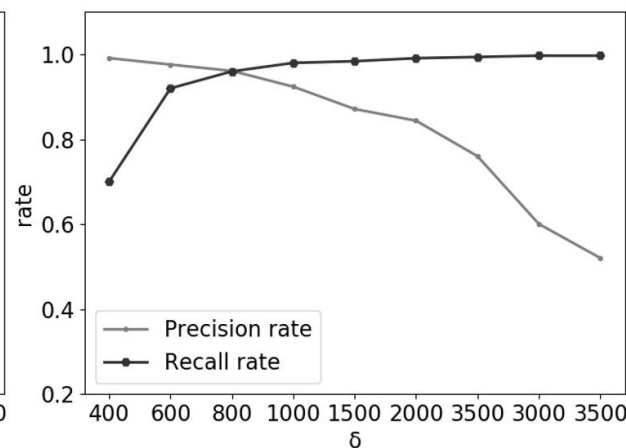
(b) NARN数据集

## (4) 轨迹距离阈值对查询结果的影响

轨迹距离阈值是轨迹相似与不相似的分界线。太小导致查全率较低，过大导致查准率较低



(a) GL数据集

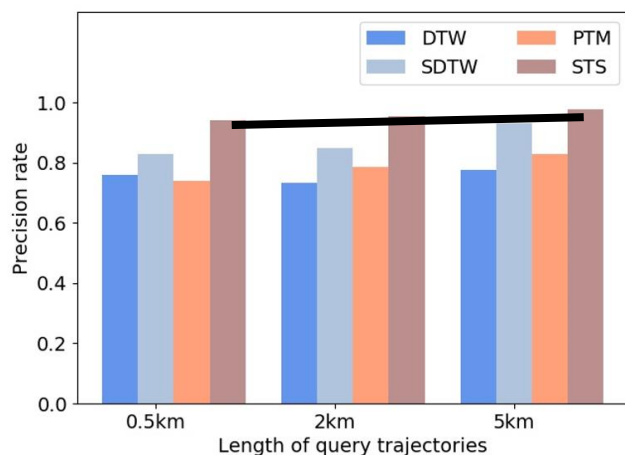


(b) NARN数据集

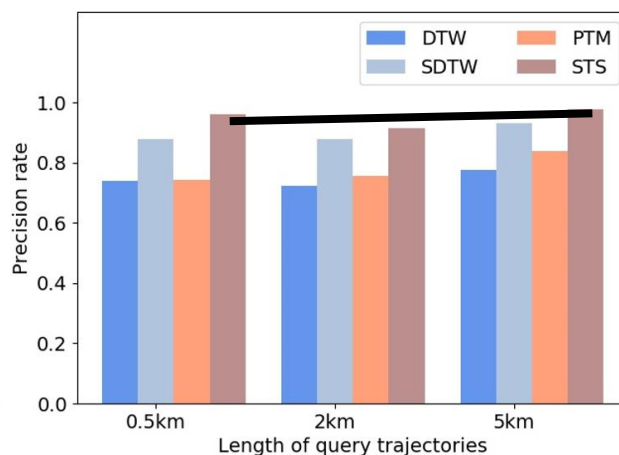
# 查询轨迹长度的影响

参与比较的算法：DTW、SDTW、PTM、STS

变量为不同的查询轨迹的长度，通过调节算法参数，使用获得的最大的查准率作为算法效果



(a) GL数据集



(b) NARN数据集

## 结论：

DTW未考虑时间因素，查准率较低

SDTW受采样策略影响大，波动较大

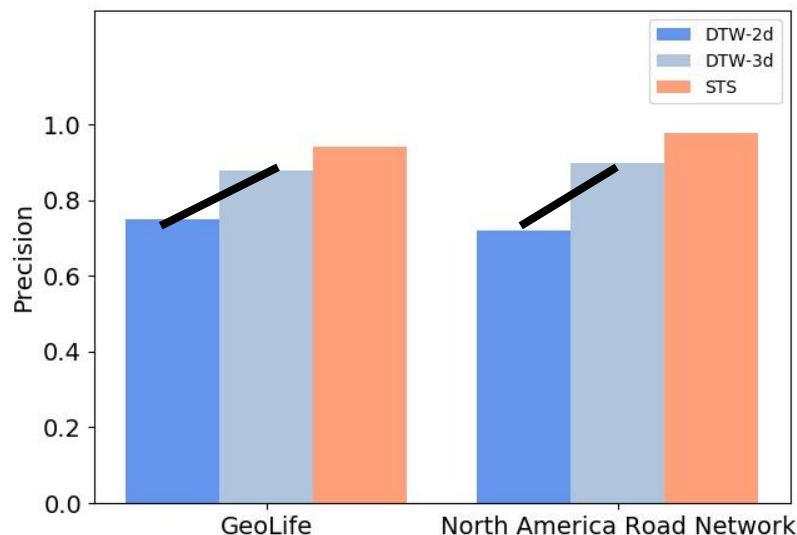
PTM易造成时空混乱，查准率较低

STS查准率较高，较稳定，轨迹长度的变化未带来较大影响

# 验证SNTR模型有效性

参与比较的算法：DTW-2d、DTW-3d、STS

分别在欧式空间和三维时空下使用DTW算法，使用查准率的变化验证三维时空的有效性



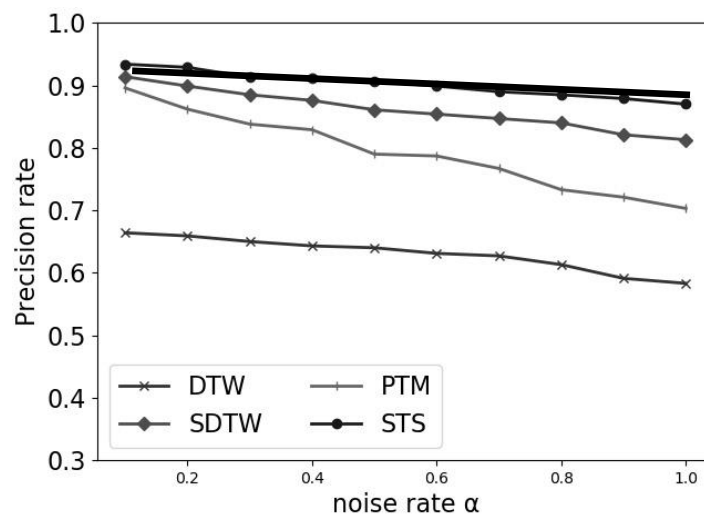
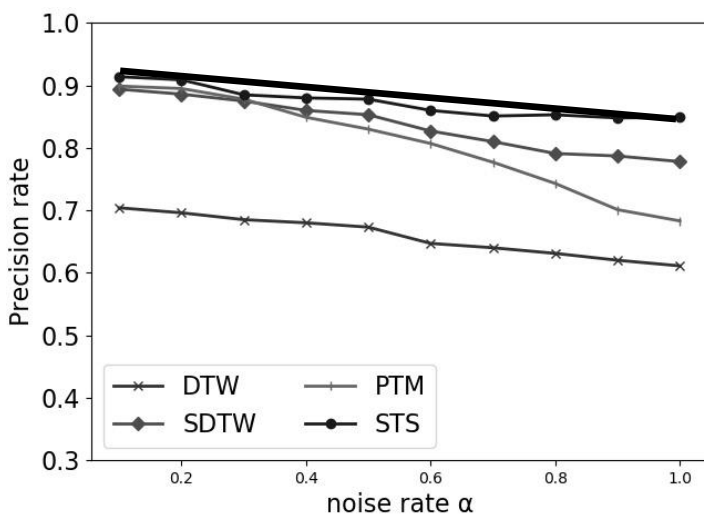
结论：

SNTR模型对DTW-3d的查准率有较大提升，因此SNTR模型是有效的

# 噪音的影响

参与比较的算法：DTW、SDTW、PTM、STS

使用均匀分布的随机数添加噪音，噪音率 $\alpha$ 来表示不同程度的噪音



结论：

DTW查准率较低，PTM波动较大  
SDTW与STS查准率较高，较稳定

05

# 总结

研究内容

贡献点

# 总结



- 提出了SNTR模型，解决相似性查询算法中时空混乱的问题。
- 提出了DTW-BDS对应点匹配算法，提升了对应点匹配效果。
- 设计了更加准确的轨迹距离计算方法，降低采样策略的影响。
- 基于以上工作，提出STS查询算法。
- 实验结果反映了本文提出的模型和算法的有效性和准确性。

# 谢谢

请各位老师批评指正