基站GPS数据分析与可视化

王延清 贺子昊 翟云鹏 杨泽远 张悦枫

2017年6月

基站GPS数据分析与可视化

工程介绍

原始数据

天津市11天的基站数据

基站号		经度 & 纬度		起止时间 & 时长		
14054	11353	117.248450	39.080030	20140506000028	20140506002106	1238
14081	24363	117.361440	39.015360	20140506002546	20140506003652	666
1317	58623	117.375681	38.986116	20140506003809	20140506003852	43
14081	34201	117.370760	38.981610	20140506010626	20140506073433	23287
14081	24242	117.363340	38.992200	20140506075817	20140506110823	11406
14081	24201	117.370760	38.981610	20140506112542	20140506120039	2097
14081	34242	117.363340	38.992200	20140506120501	20140506122703	1322
14081	24242	117.363340	38.992200	20140506123215	20140506153705	11090
14081	24193	117.444470	38.986560	20140506160440	20140506162648	1328
14081	14193	117.444470	38.986560	20140506164906	20140506170843	1177
14054	20351	117.349130	39.027450	20140506171633	20140506172551	558
14057	11211	117.234657	39.091129	20140506172646	20140506172646	0
14054	11083	117.239111	39.092049	20140506172705	20140506172726	21
14057	11372	117.233250	39.096220	20140506172808	20140506172812	4
14057	11373	117.233250	39.096220	20140506172838	20140506211026	13308
14057	21103	117.221992	39.100487	20140506211943	20140506225901	5958
14054	11083	117.239111	39.092049	20140506225932	20140506230138	126
14054	11402	117.240300	39.083750	20140506230206	20140506231026	500
14081	24421	117.364920	39.022460	20140506231202	20140506231215	13
14081	34183	117.404200	39.002840	20140506231442	20140506232716	754
14081	24242	117.363340	38.992200	20140506233422	20140506235255	1113
1317	50331	117.365193	38.983627	20140506235332	20140506235357	25

基站GPS数据分析与可视化

概览

1. 基于关键点提取和时序分析的轨迹语义化
2. 基于活动密集度的轨迹语义化
3. 用户行程推荐与可视化

基站GPS数据分析与可视化

概览

1. 基于关键点提取和时序分析的轨迹语义化
2. 基于活动密集度的轨迹语义化
3. 用户行程推荐与可视化

基于关键点提取和时序分析的轨迹语义化

项目介绍

用户轨迹语义化是指：根据用户GPS位置及停留时间等数据，将用户轨迹模式和用户目的进行匹配。

Where and Why

基于关键点提取和时序分析的轨迹语义化

项目意义

1. 合理解释轨迹模式

2. 为轨迹周期性探究提供语义化解释

3. 发挥轨迹数据的商业价值，是用户推荐系统的基础。

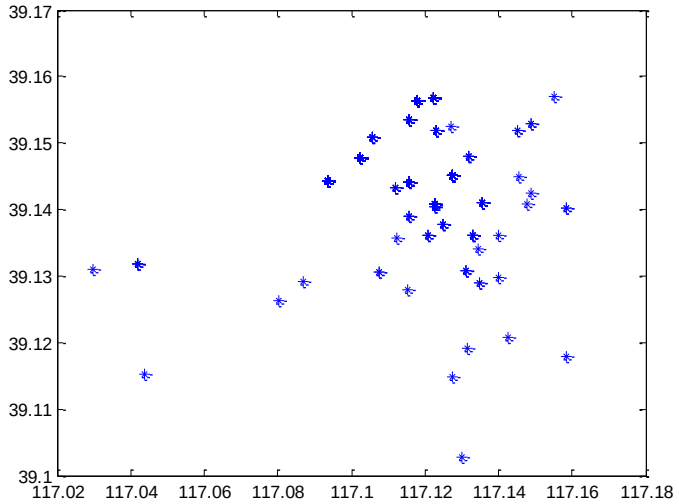
基于关键点提取和时序分析的轨迹语义化

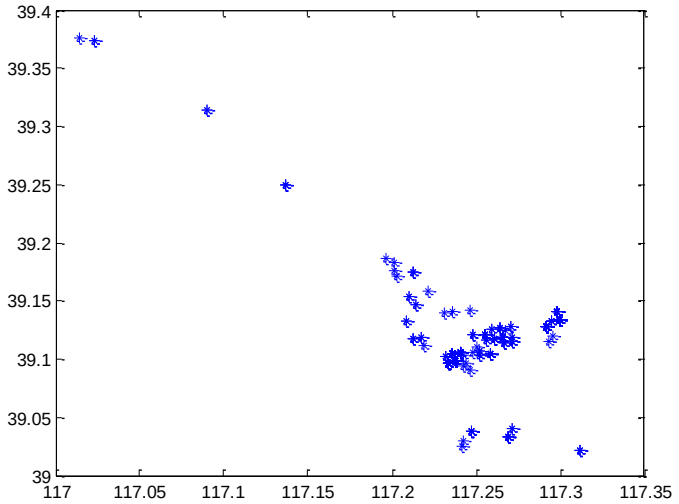
要解决的问题

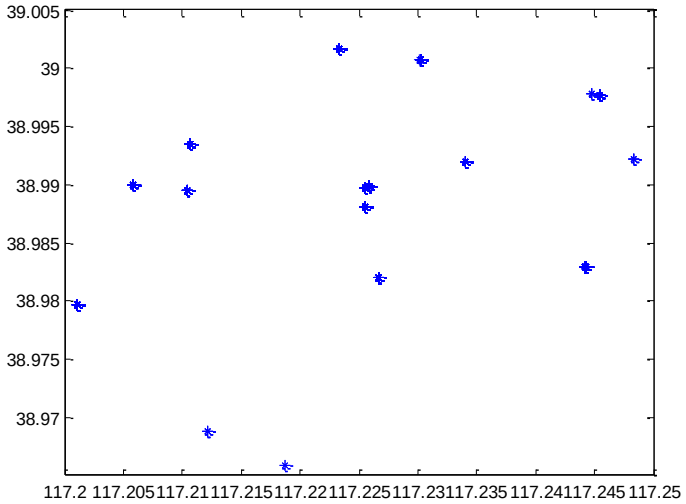
用户数据
Raw Data

GPS　　起止时间　　停留时间

44055	30262	117.225910	38.989820	20140504000054	20140504004352	2578
44055	20262	117.225910	38.989820	20140504004400	20140504004427	27
44055	30262	117.225910	38.989820	20140504004439	20140504004609	90
44055	10471	117.225500	38.988070	20140504004623	20140504004635	12
44055	30262	117.225910	38.989820	20140504004646	20140504004646	0
1318	59457	117.244207	38.982928	20140504004713	20140504010123	850
1318	56053	117.245423	38.997708	20140504010144	20140504010144	0

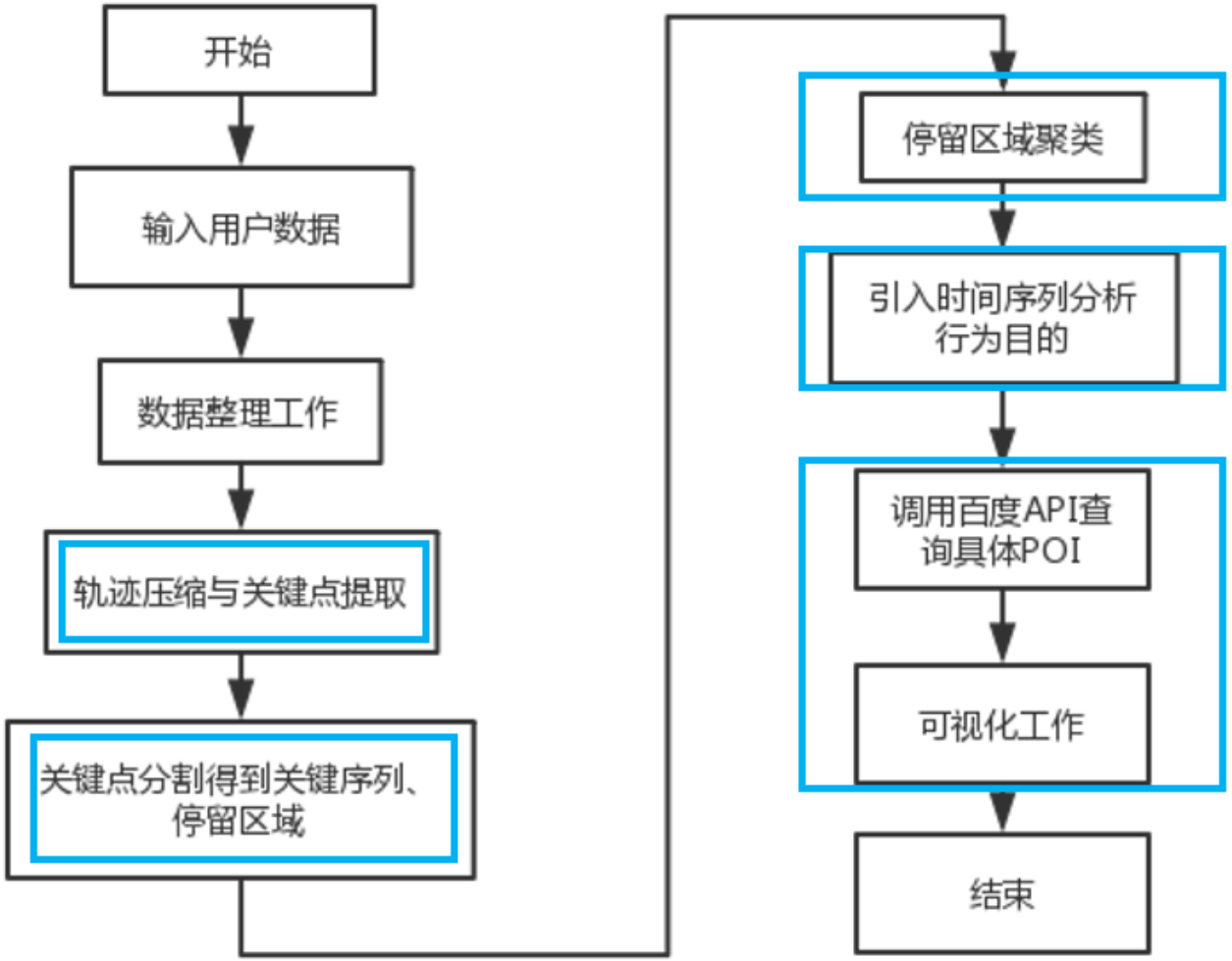






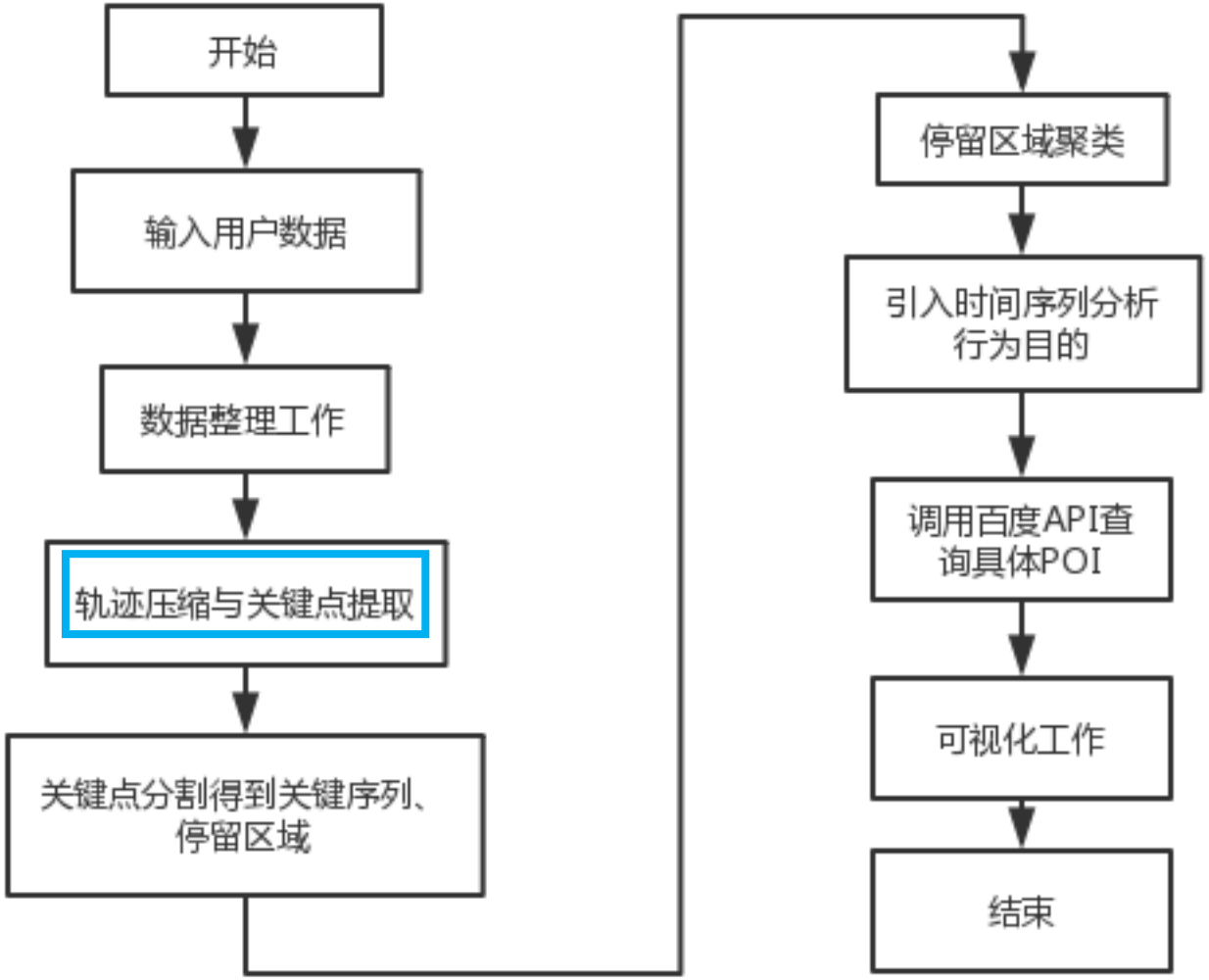
基于关键点提取和时序分析的轨迹语义化

项目流程



基于关键点提取和时序分析的轨迹语义化

轨迹压缩与关键点提取



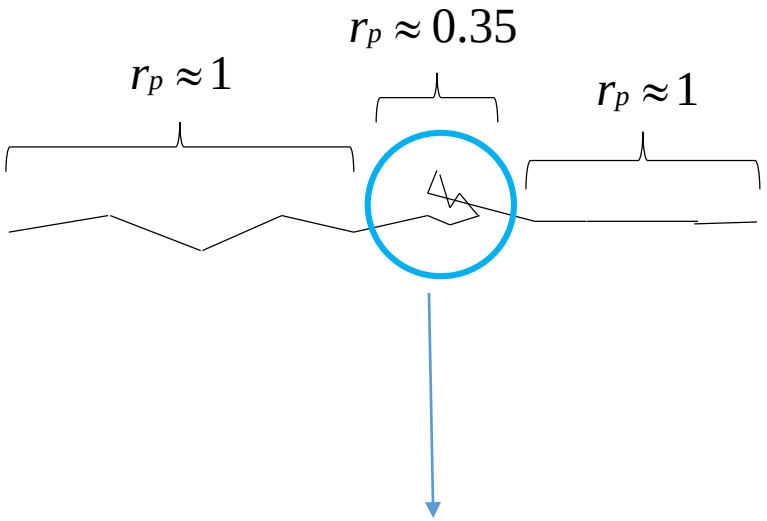
基于关键点提取和时序分析的轨迹语义化

轨迹压缩与关键点提取

准则：相关系数+停留时间

皮尔逊相关系数：

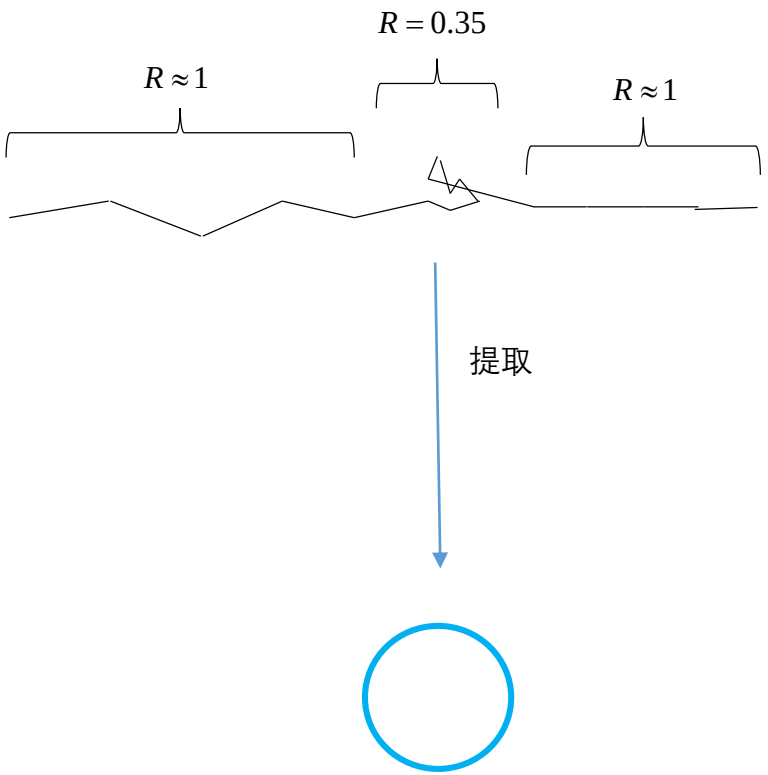
$$r_p(X,Y)=\frac{\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})}{[\sum_{i=1}^{n}(X_i-\overline{X})^2\sum_{i=1}^{n}(Y_i-\overline{Y})^2]^{\frac{1}{2}}}$$

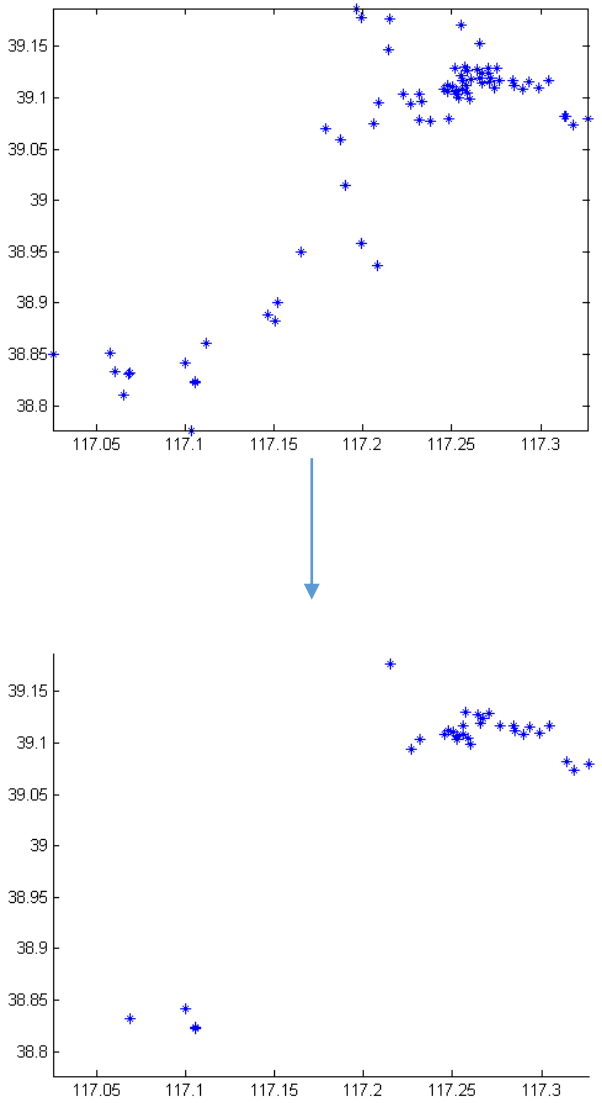


如果满足：time>=time_threshold 或者r<=r_threshold
此点存为关键点

基于关键点提取和时序分析的轨迹语义化

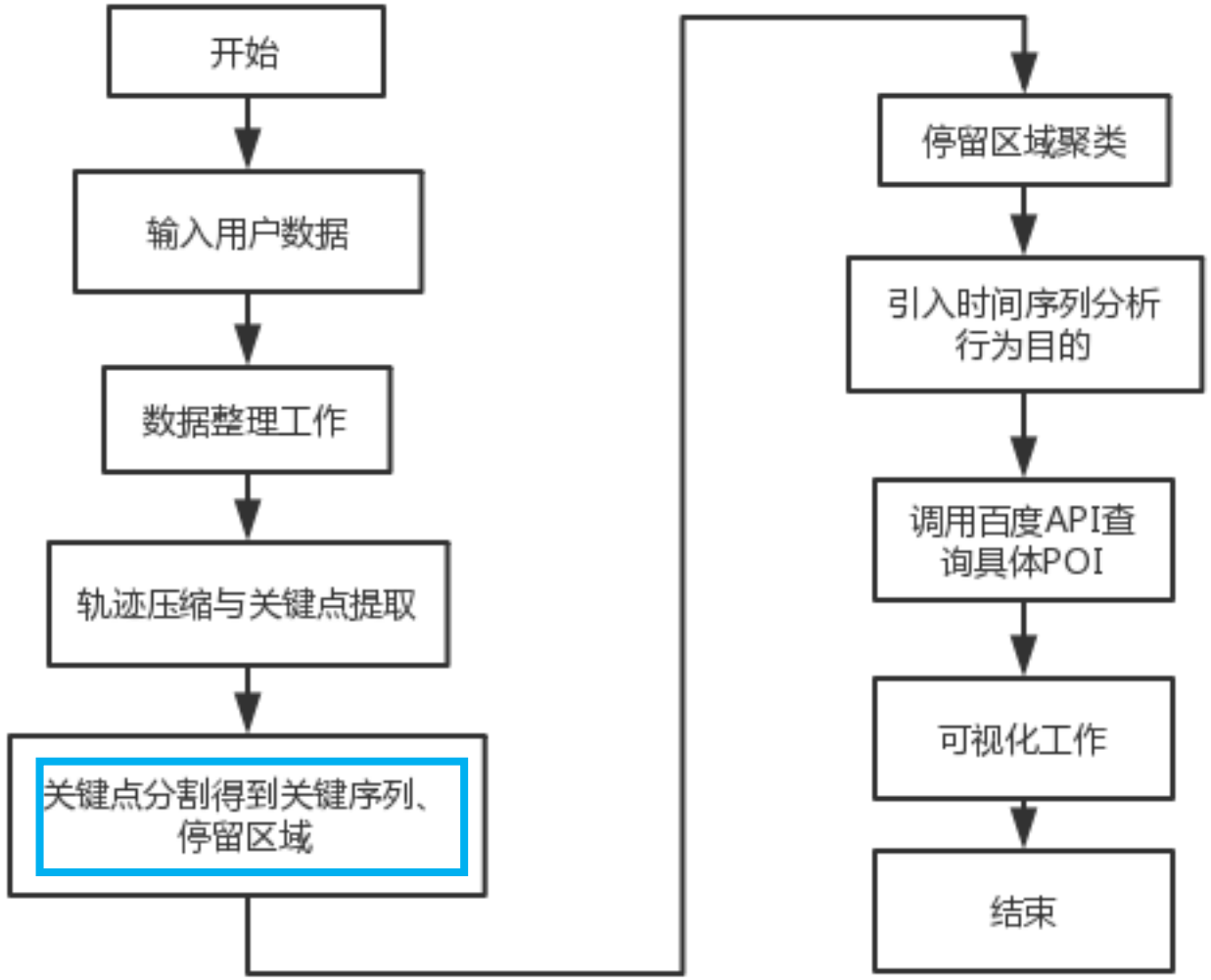
轨迹压缩与关键点提取





基于关键点提取和时序分析的轨迹语义化

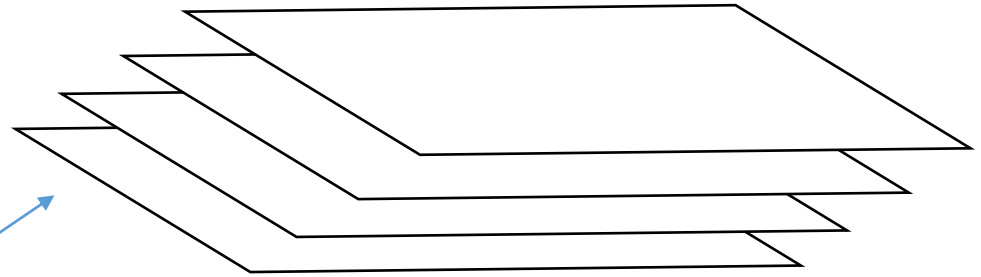
关键点序列分割得到停留区域



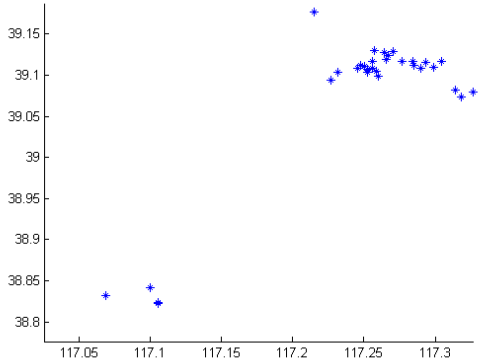
基于关键点提取和时序分析的轨迹语义化

关键点序列分割得到停留区域

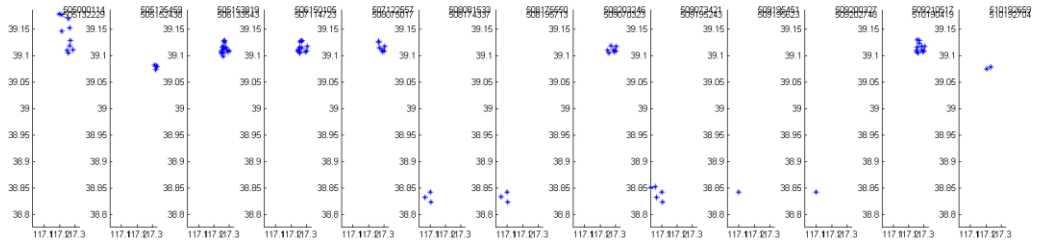
引入时间序列，将轨迹分层

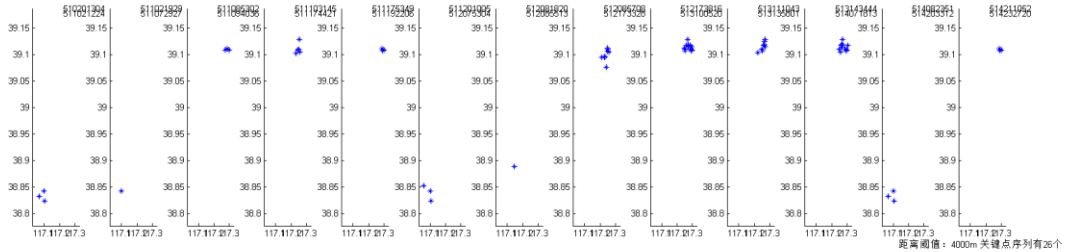


引入时序



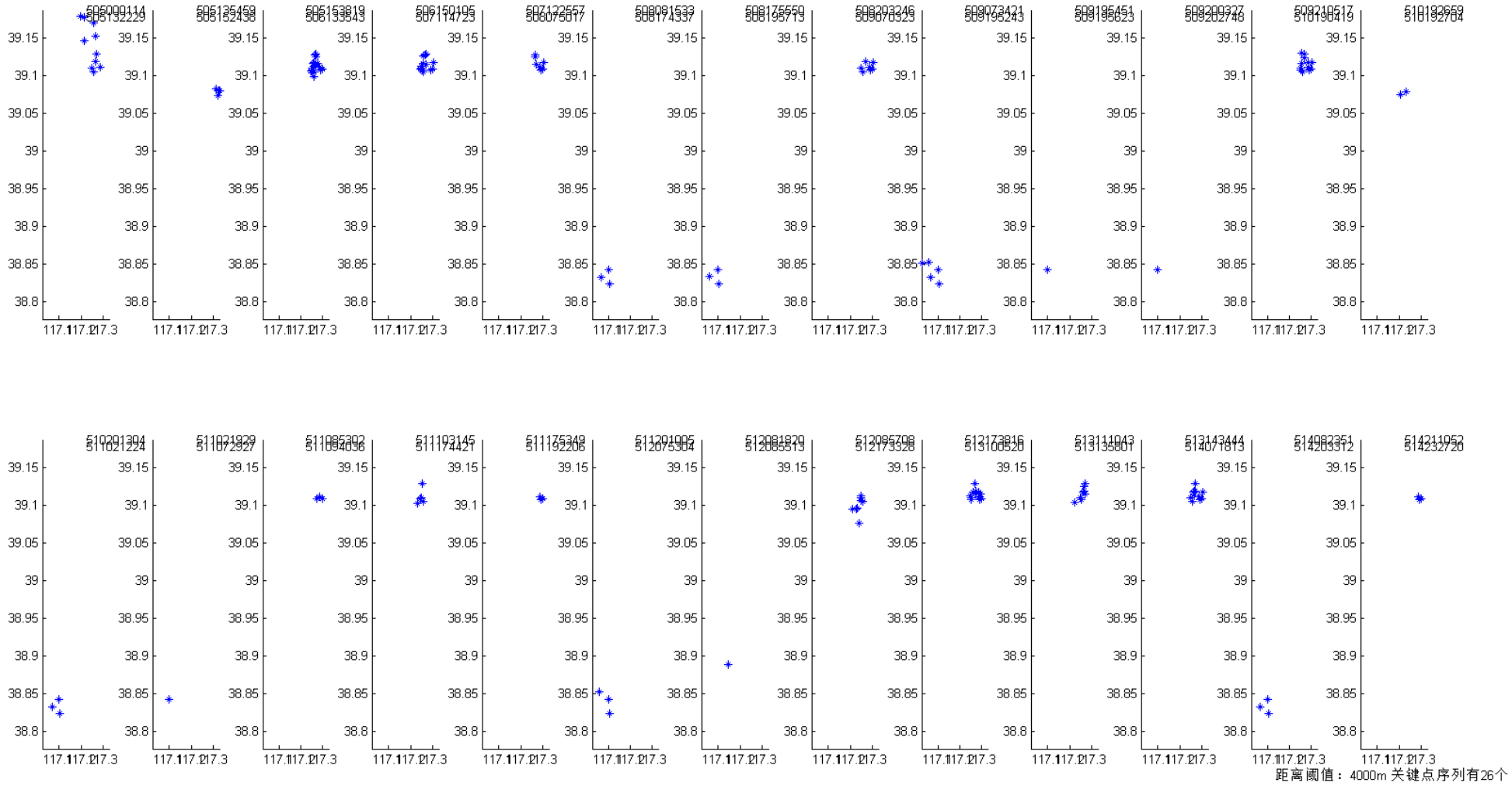
距离阈值分割





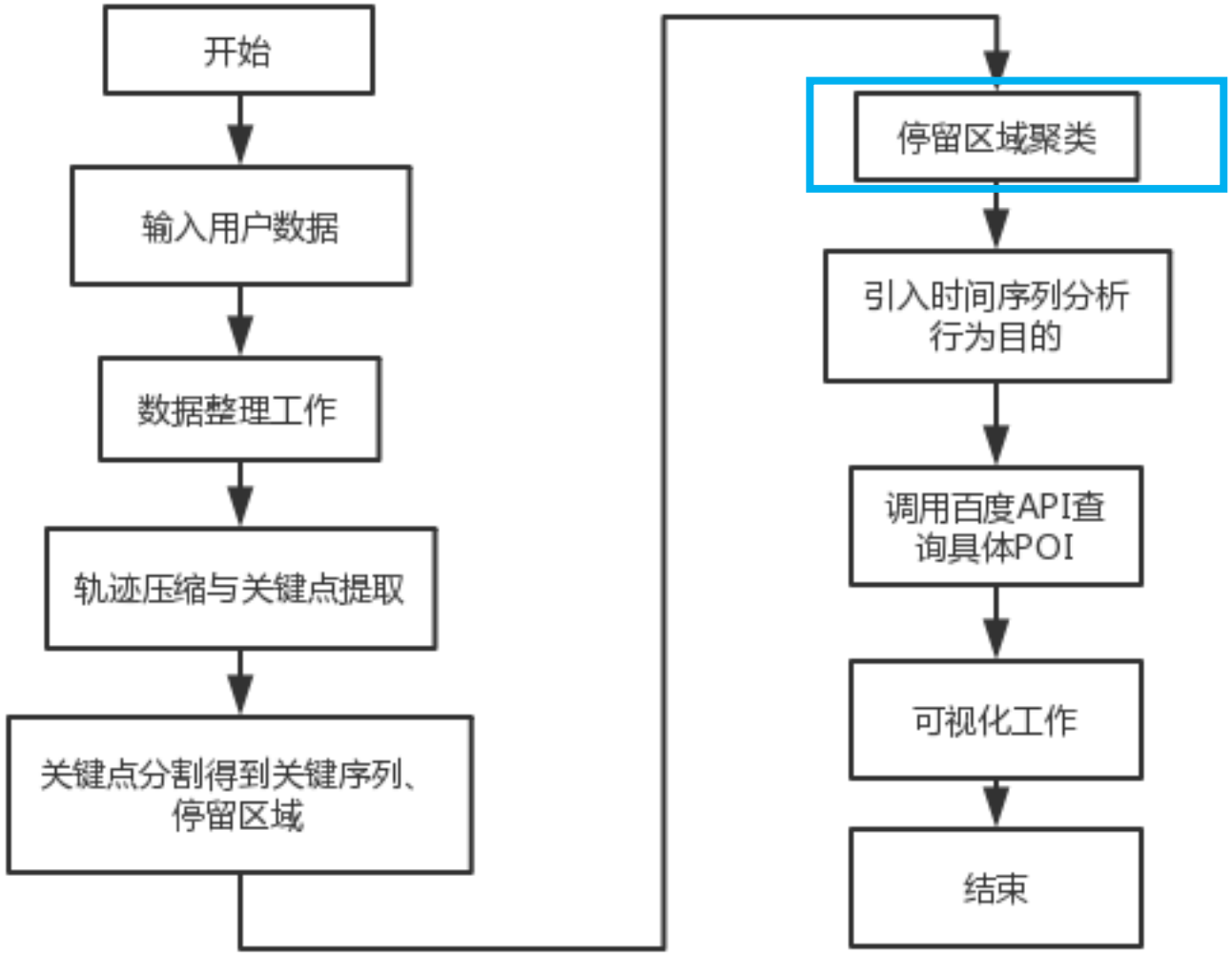
基于关键点提取和时序分析的轨迹语义化

关键点序列分割得到停留区域



基于关键点提取和时序分析的轨迹语义化

停留区域聚类：改进的K-means



基于关键点提取和时序分析的轨迹语义化

停留区域聚类：改进的K-means

将K-means类数N限制在[2;3;4;5]中，使用类内距离和类间距离作为评价准则。

类内距离越小越好，类间距离越大越好。所以使用比值表示。

类内距离：
$$S_w = \sum_{i=1}^{n} \sum_{x_j \in \chi} (x_j - m_i)^2$$

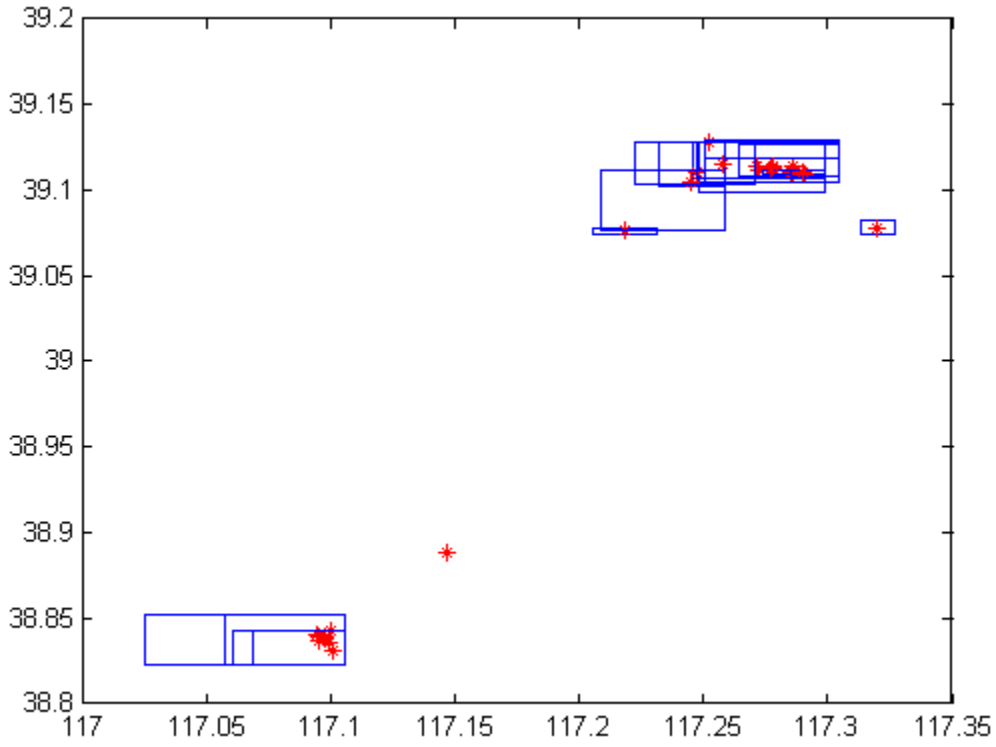
类间距离：
$$S_b = \sum_{i=1}^{N} \left[\sum_{j=1,\cdot\cdot,i-1,i+1,\cdot\cdot N} (m_i - m_j)^2 \right]$$

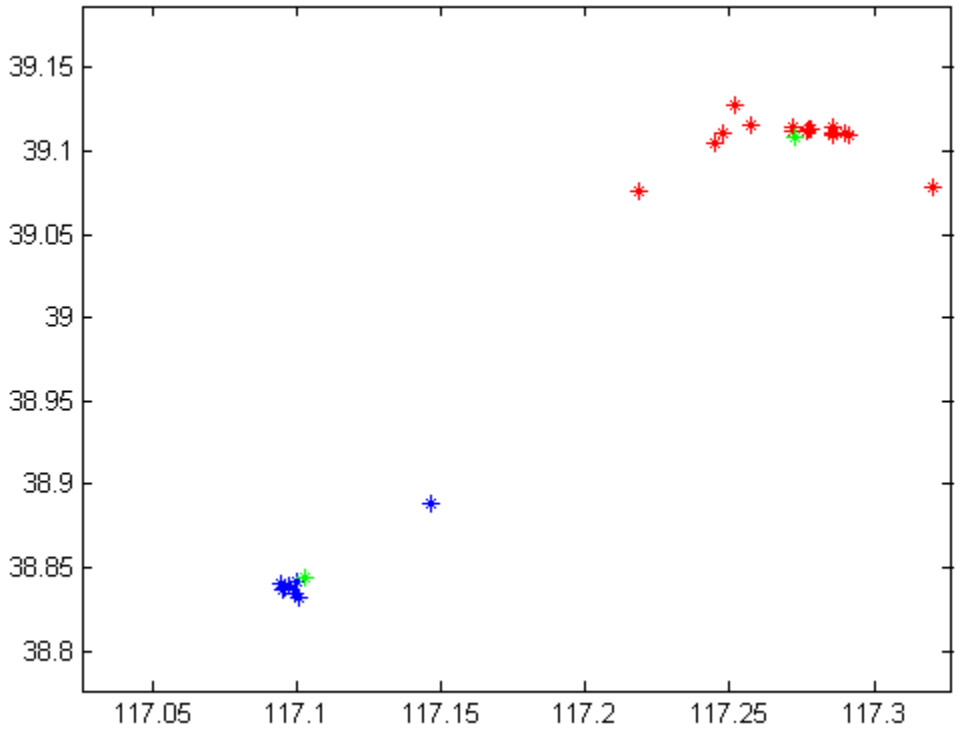
$$N = \arg\max J(N) = \frac{S_b}{S_w}$$

一般会聚成两类或三类

基于关键点提取和时序分析的轨迹语义化

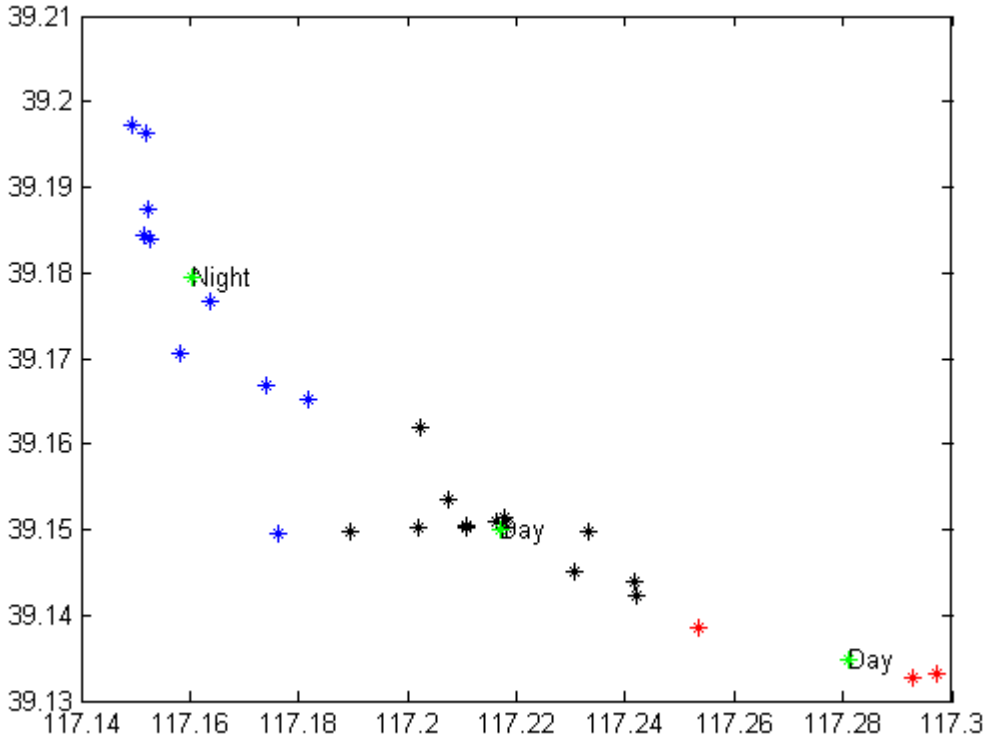
停留区域聚类：改进的K-means





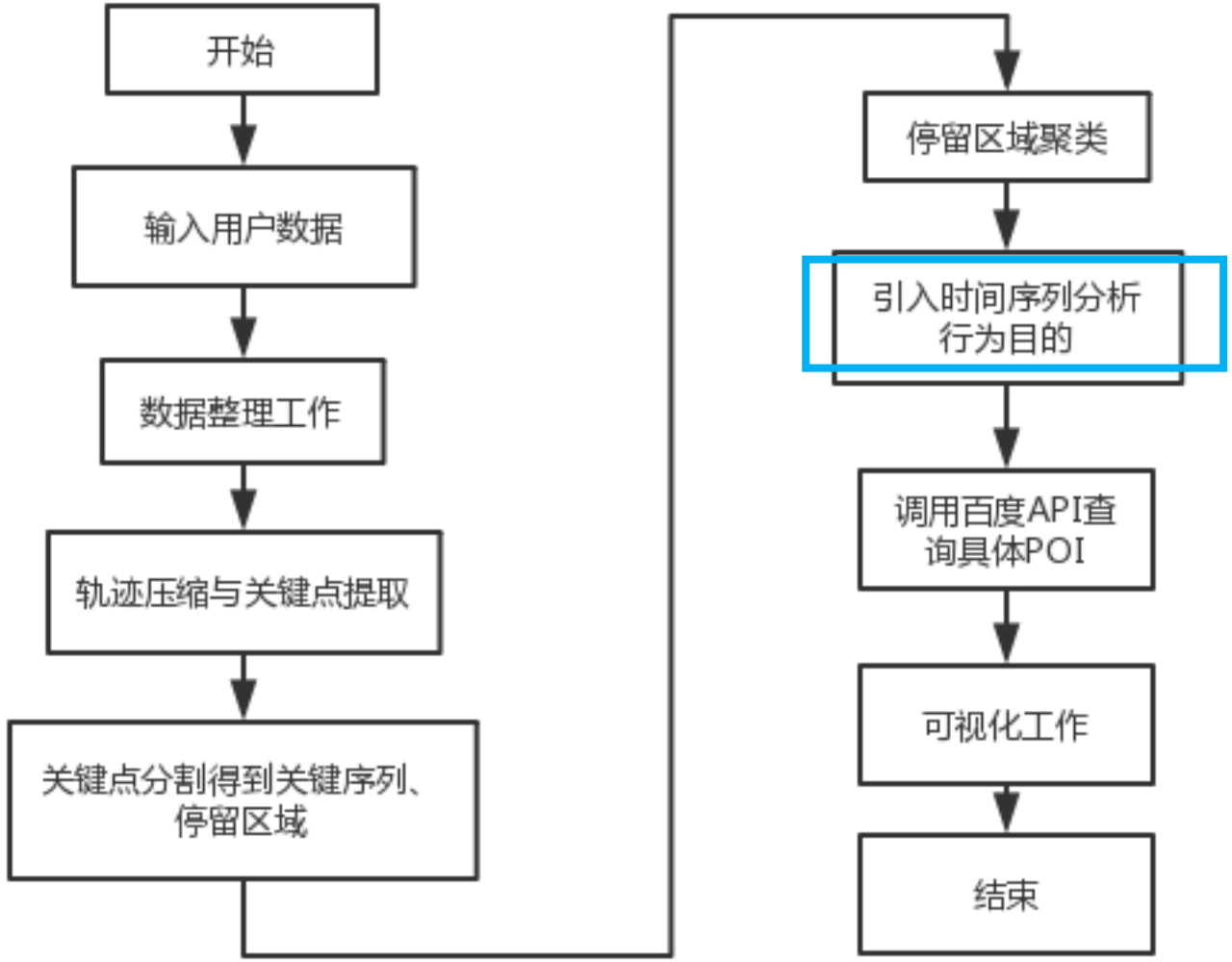
基于关键点提取和时序分析的轨迹语义化

停留区域聚类：改进的K-means



基于关键点提取和时序分析的轨迹语义化

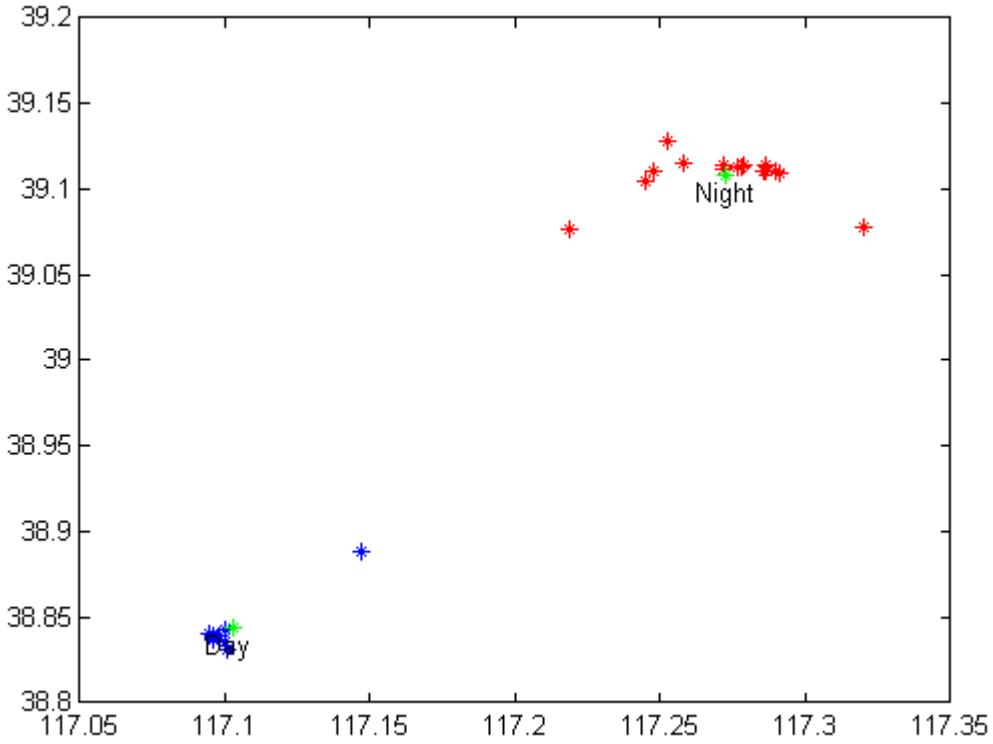
引入时间序列分析用户轨迹语义



基于关键点提取和时序分析的轨迹语义化

引入时间序列分析用户轨迹语义

一般两点型的轨迹分析比较简单，时间区分也比较明显。一般两区域对应为白天和晚上。标签结果如图：



基于关键点提取和时序分析的轨迹语义化

项目亮点

1. 改进了CCM算法，将计算全局相关系数优化为某点前后范围内的相关系数。

2. 改进的K-means聚类，引入评价准则，选取最优的聚类方案，避免聚类数的人工赋值。

3. 人工因素比较少，关键点提取，停留区域提取、停留区域聚类和POI查询等过程不需人工介入，方便海量用户的分析。

基于关键点提取和时序分析的轨迹语义化

后期改进

多模式的探究，三或四类

城市功能区的概念的划分

基站GPS数据分析与可视化

概览

1. 基于关键点提取和时序分析的轨迹语义化
2. 基于活动密集度的轨迹语义化
3. 用户行程推荐与可视化

基于活动密集度的轨迹语义化

概览

群体活动密集度分析

城市人口分布

个体活动度分析

活动区域挖掘　活动周期挖掘　活动模式挖掘

群体活动密集度分析

分析工具

双变量正态密度核函数

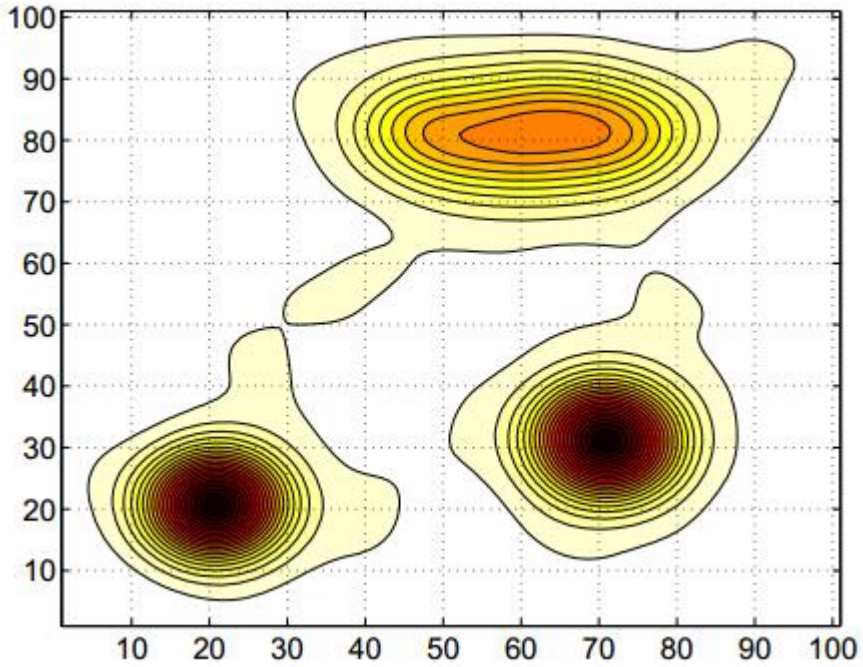
$$f(c) = \frac{1}{n\gamma^2} \sum_{i=1}^{n} \frac{1}{2\pi} \exp\left(-\frac{|c - loc_i|^2}{2\gamma^2}\right)$$

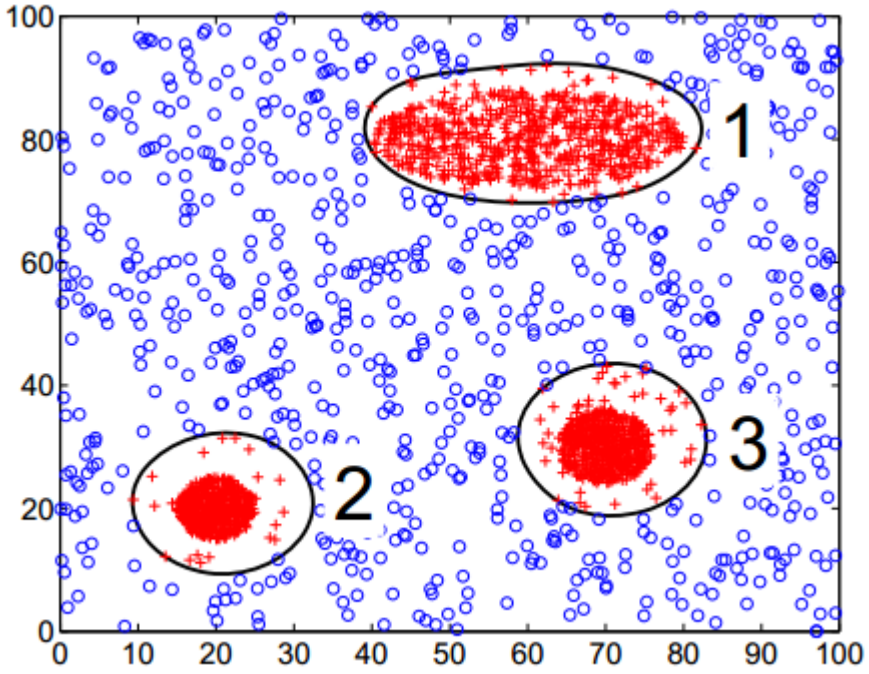
其中 $$\gamma = \frac{1}{2}(\sigma_x^2 + \sigma_y^2)^{\frac{1}{2}} n^{-\frac{1}{6}}$$

B. J. Worton. Kernel methods for estimating the utilization distribution in home-range studies. In Ecology, volume 70,1989

群体活动密集度分析

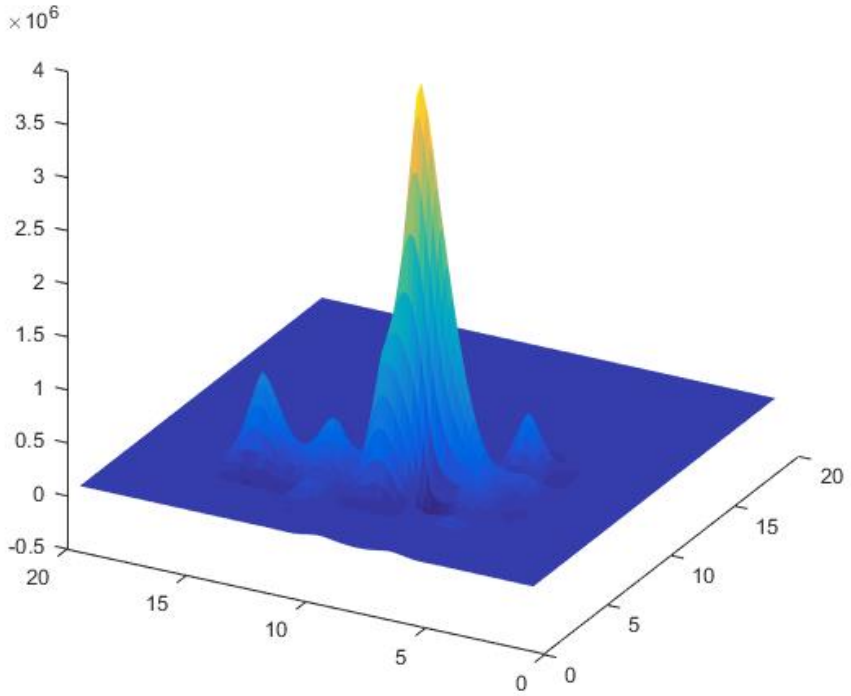
分析工具

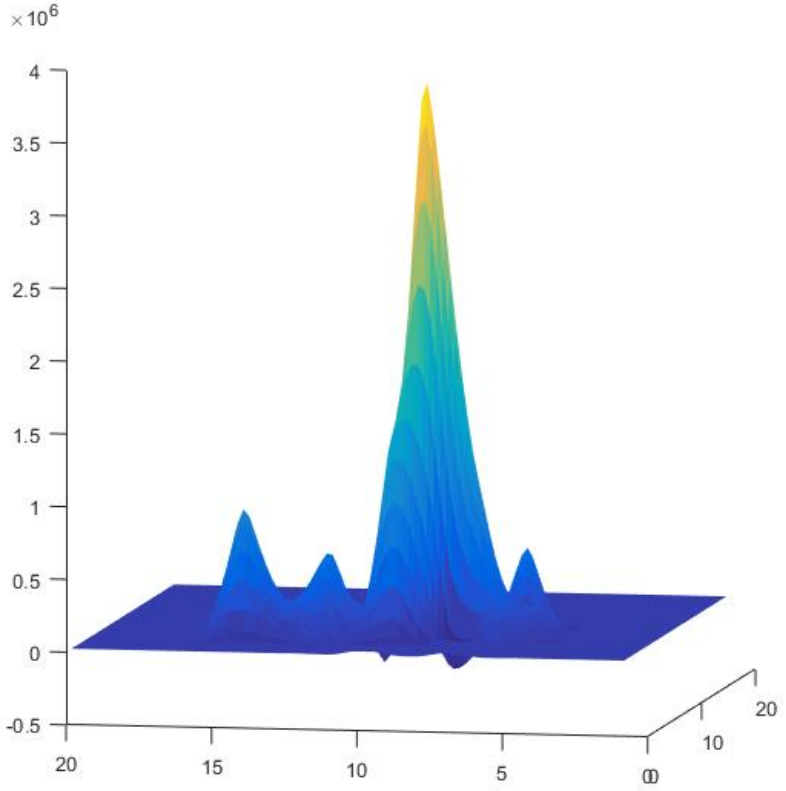




群体活动密集度分析

三维曲面图





群体活动密集度分析

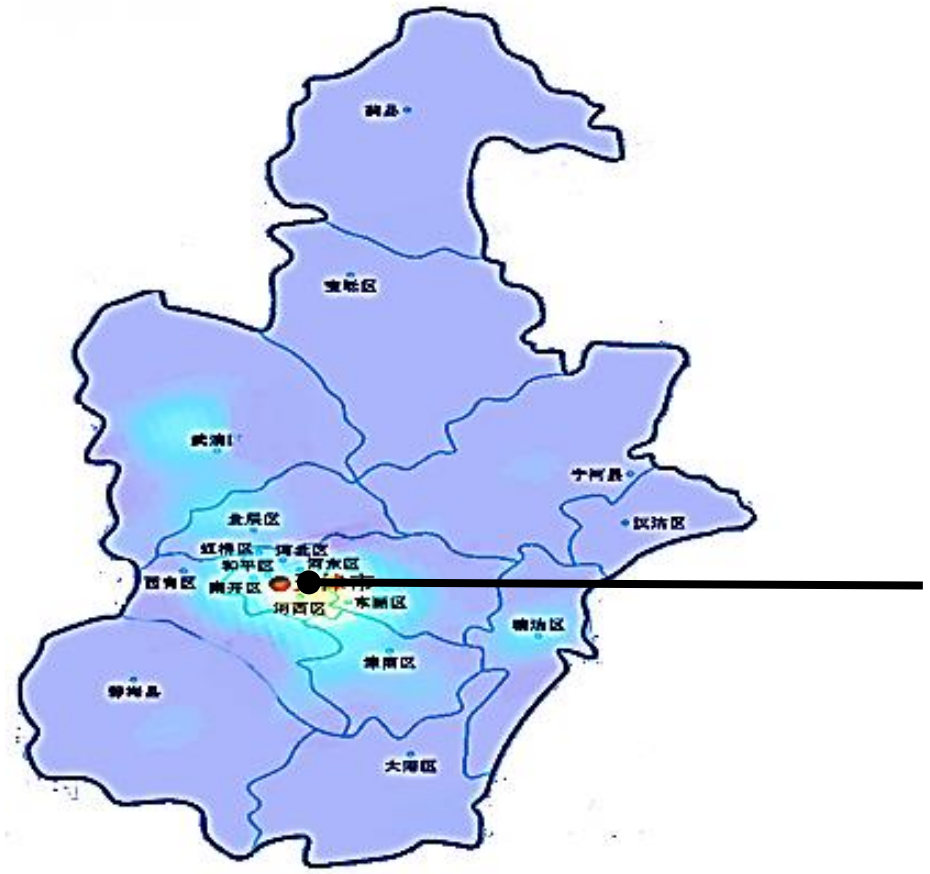
伪彩色图 + 真实地图

将伪彩色图嵌入天津市地图

色温高的区域密集度大

群体活动密集度分析

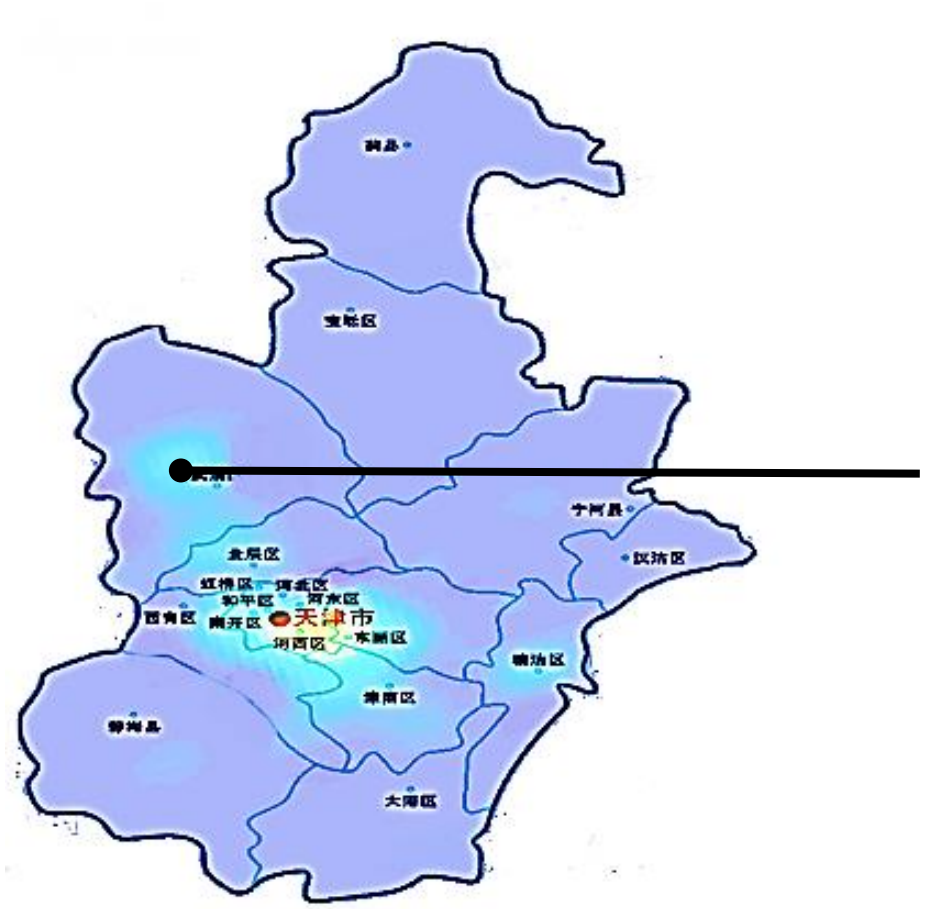
结果分析



天津市中心

群体活动密集度分析

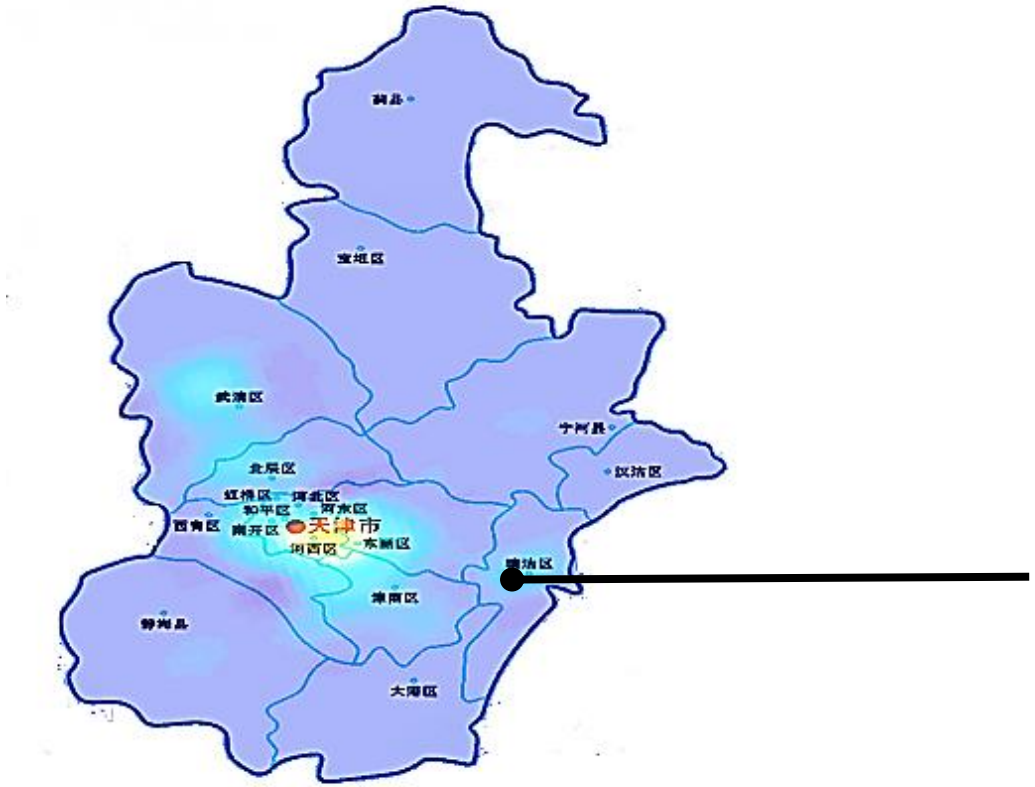
结果分析



武清区中心

群体活动密集度分析

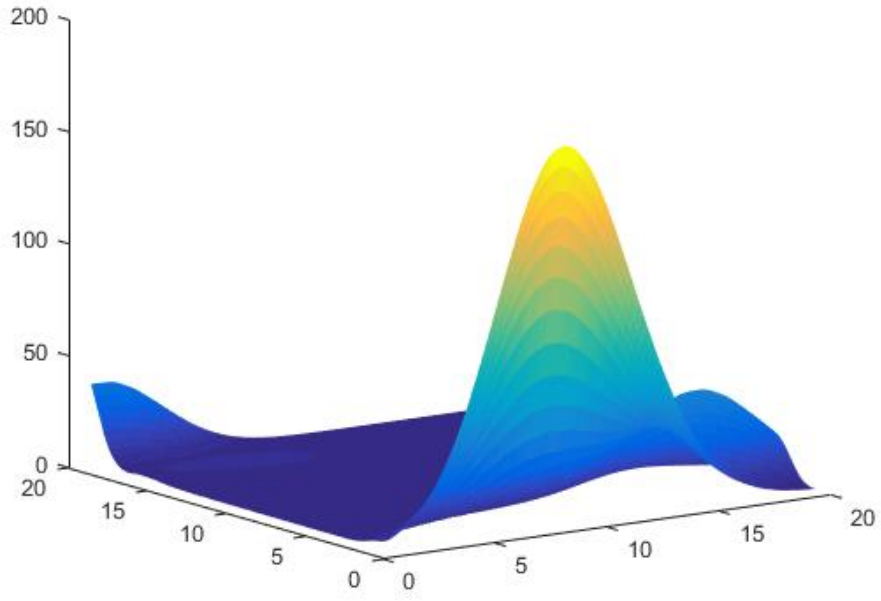
结果分析

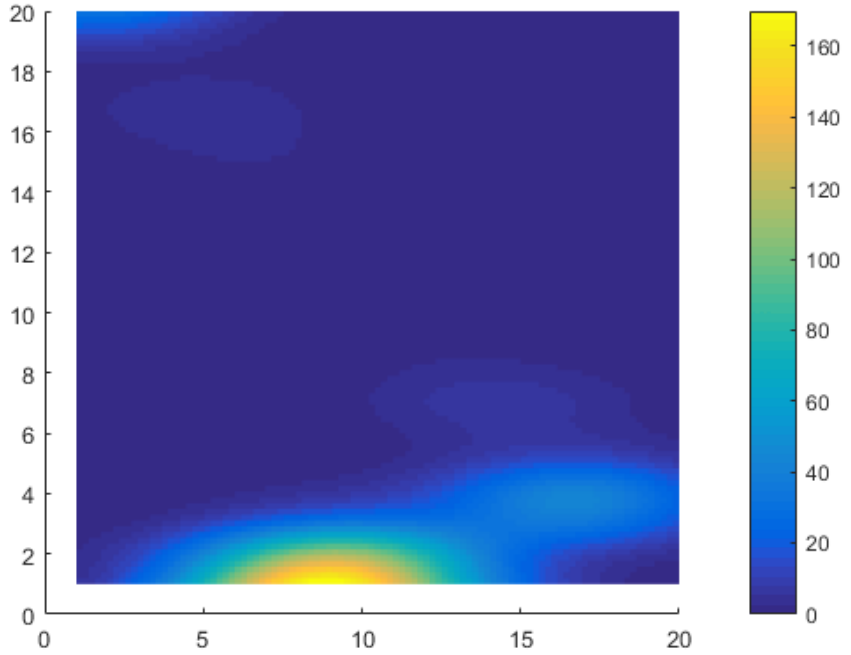


塘沽区中心

个体活动密集度分析

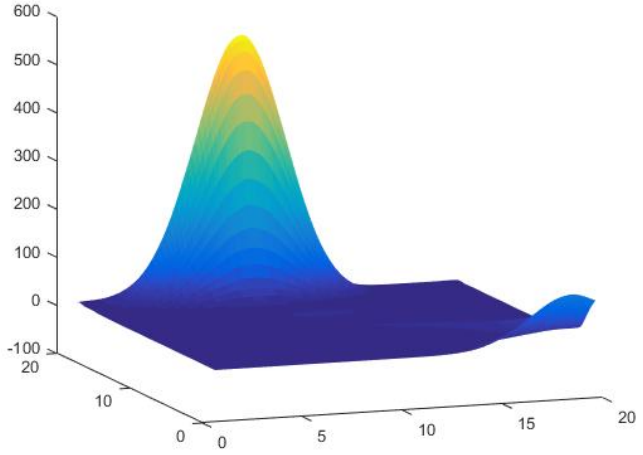
结果分析

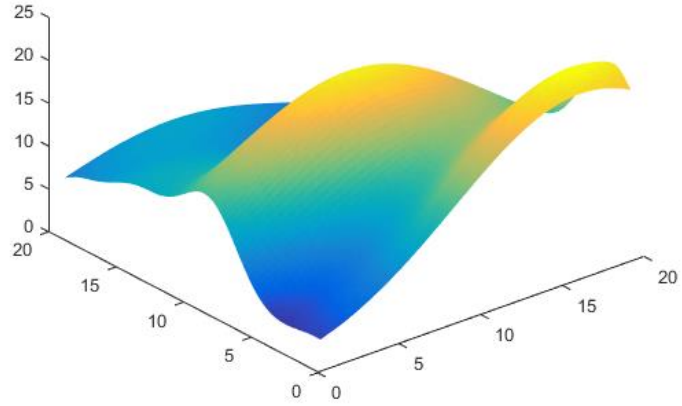


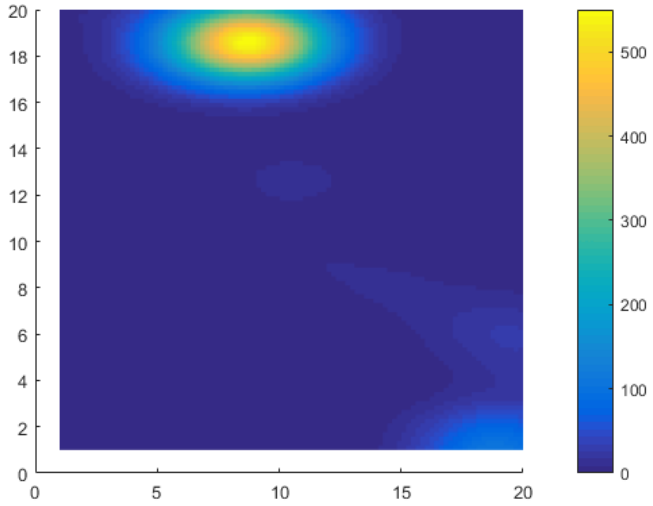


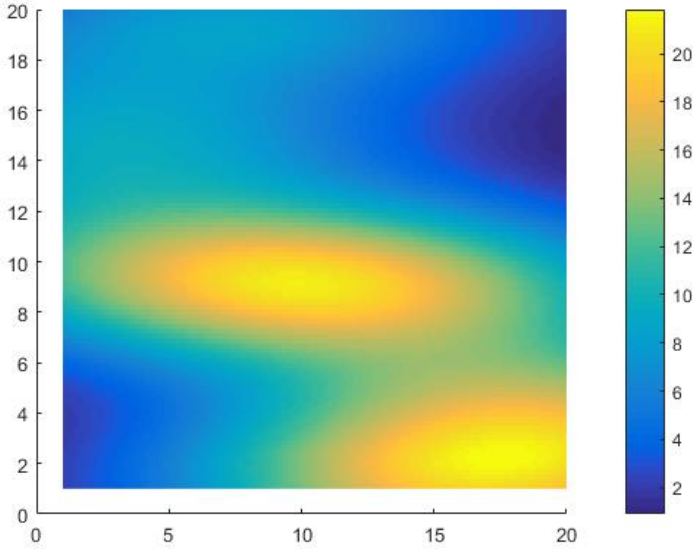
个体活动密集度分析

模式一



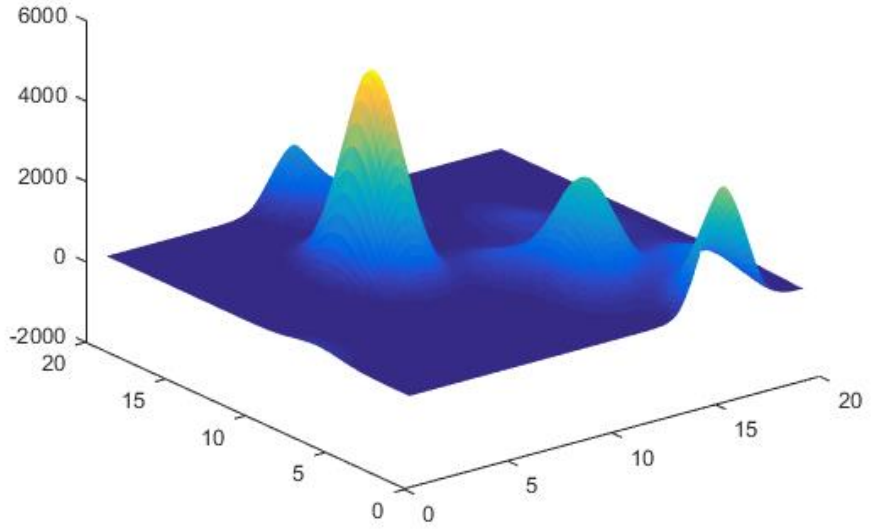


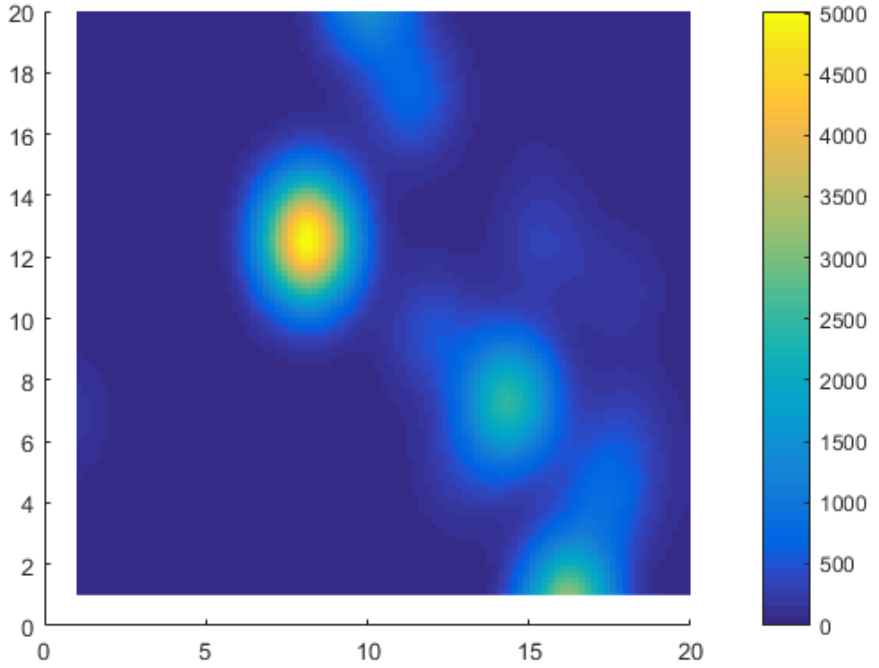




个体活动密集度分析

模式二





个体活动周期挖掘

快速傅里叶变换

对每个重点活动区域：

- 挑选数据，时序建模

  用二进制序列表示到达与否

- 利用 DFT 挖掘时间周期

$B = b_1, b_2, b_3 \ldots b_n$

e.g.

$= 0,0,1,1,1,0,0 \ldots$

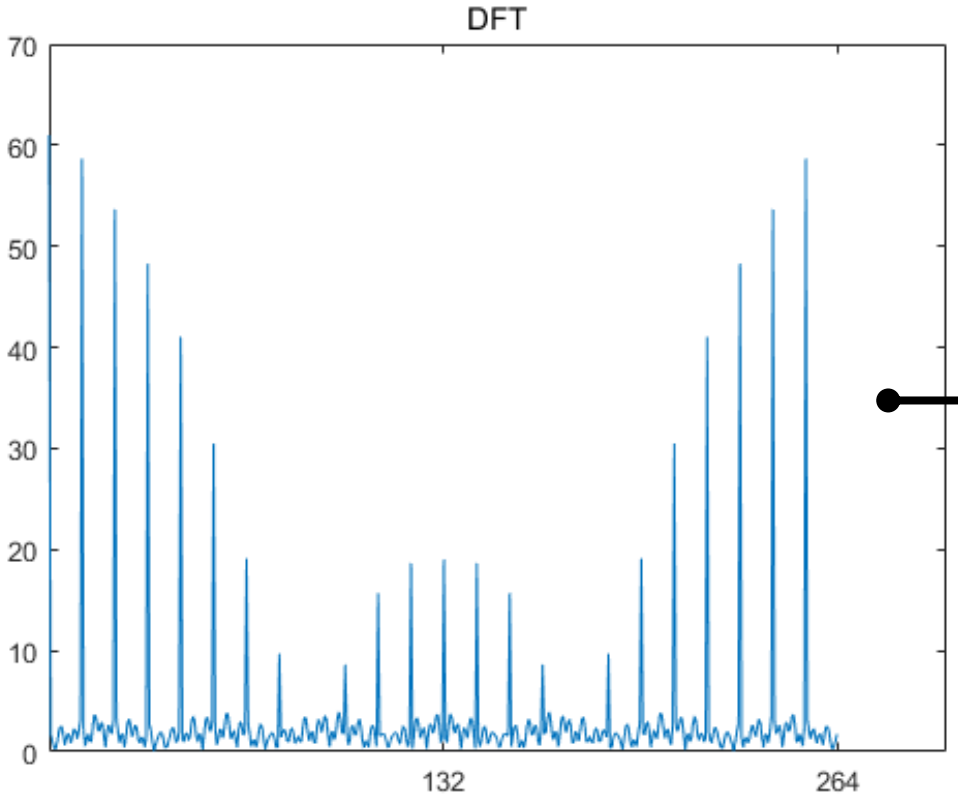
⇩ DFT

$X = X_1, X_2 \ldots X_n$

M. Vlachos, P. S. Yu, and V. Castelli. On periodicity detection and structural periodic similarity. In SDM, 2005.

活动周期挖掘

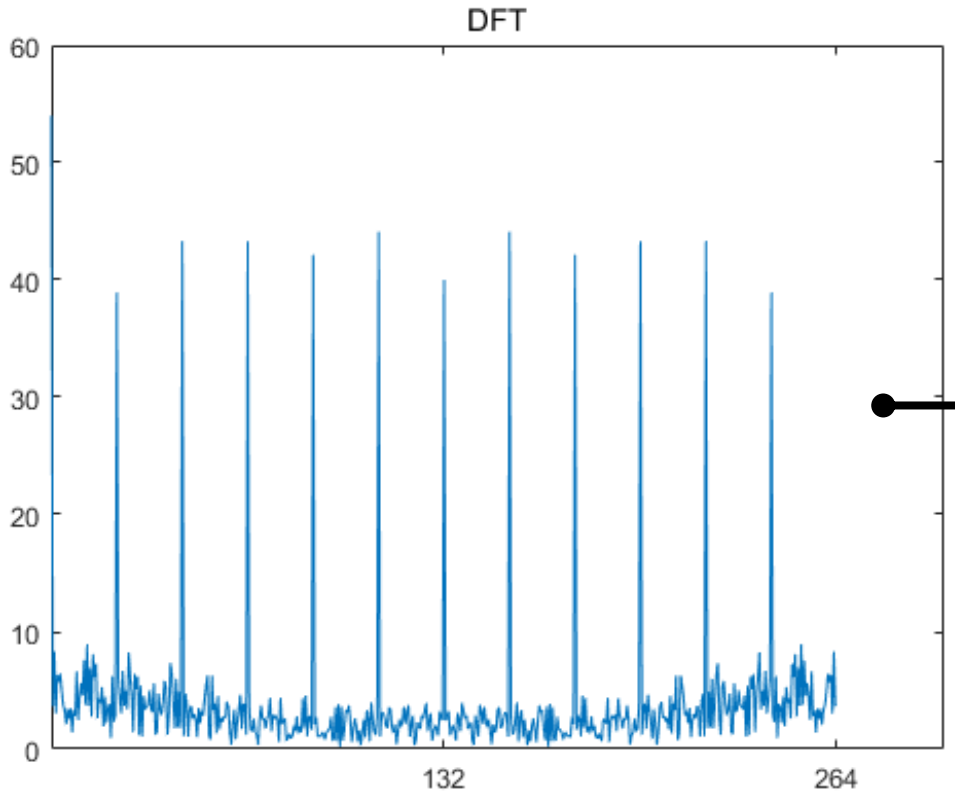
快速傅里叶变换



T = 24h

活动周期挖掘

快速傅里叶变换



T = 12h

活动周期挖掘

瓶颈

时间维度上数据量匮乏

活动模式挖掘

根据周期进行轨迹分段

轨迹分段：$l = l_1\ l_2\ l_3 \ldots$

其中：$l_j = l_j^1\ l_j^2\ l_j^3 \ldots$　　第 j 段

其中：$l_j^k \in [1, d]$　　第 j 段中第 k 个时间单位

活动模式挖掘

求解最优化问题

最优化问题 $$\max_{\mathbf{P}} \left\{ L(\mathbf{P}|\mathcal{I}) = \log P(\mathcal{I}|\mathbf{P}) = \sum_{I^j \in \mathcal{I}} \sum_{k=1}^{T} p(x_k = I_k^j) \right\}$$

ML（最大似然概率）

求解得到 $$p(x_k = i) = \frac{\sum_{I^j \in \mathcal{I}} \mathbf{1}_{I_k^j = i}}{|\mathcal{I}|}$$

最优转移概率

Li Z, Ding B, Han J, et al. Mining periodic behaviors for moving objects[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010:1099-1108

活动模式挖掘

得到最优转移概率

家？

0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7272	0.1818	0.0910	0.0910	0.0910
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5455
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2727	0.8182	0.9090	0.9090	0.3635

0.0000	0.0000	0.0000	0.0910	1.0000	1.0000	0.3636	0.0000	0.0000	0.0000	0.0000	0.0000
0.0910	0.0910	0.0910	0.0910	0.0000	0.0000	0.6364	1.0000	1.0000	1.0000	1.0000	1.0000
1.0000	0.9090	0.9090	0.8182	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.9090	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

活动模式挖掘

得到最优转移概率

工作地点一？

0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7272	0.1818	0.0910	0.0910	0.0910
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5455
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2727	0.8182	0.9090	0.9090	0.3635

0.0000	0.0000	0.0000	0.0910	1.0000	1.0000	0.3636	0.0000	0.0000	0.0000	0.0000	0.0000
0.0910	0.0910	0.0910	0.0910	0.0000	0.0000	0.6364	1.0000	1.0000	1.0000	1.0000	1.0000
1.0000	0.9090	0.9090	0.8182	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.9090	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

活动模式挖掘

得到最优转移概率

0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7272	0.1818	0.0910	0.0910	0.0910
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5455
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2727	0.8182	0.9090	0.9090	0.3635

0.0000	0.0000	0.0000	0.0910	1.0000	1.0000	0.3636	0.0000	0.0000	0.0000	0.0000	0.0000
0.0910	0.0910	0.0910	0.0910	0.0000	0.0000	0.6364	1.0000	1.0000	1.0000	1.0000	1.0000
1.0000	0.9090	0.9090	0.8182	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.9090	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

工作地点二？

活动模式挖掘

得到最优转移概率

0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7272	0.1818	0.0910	0.0910	0.0910
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.5455
0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.2727	0.8182	0.9090	0.9090	0.3635

0.0000	0.0000	0.0000	0.0910	1.0000	1.0000	0.3636	0.0000	0.0000	0.0000	0.0000	0.0000
0.0910	0.0910	0.0910	0.0910	0.0000	0.0000	0.6364	1.0000	1.0000	1.0000	1.0000	1.0000
1.0000	0.9090	0.9090	0.8182	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
0.9090	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

接人 / 晚饭？

基于活动密集度的轨迹语义化

回顾

群体活动密集度分析

城市人口分布

个体活动度分析

活动区域挖掘　活动周期挖掘　活动模式挖掘

基于活动密集度的轨迹语义化

前瞻

时间维度更长的数据集

更多活动周期 & 活动模式

拥有同一周期的 活动模式分类（KL聚类）

基于活动密集度的轨迹语义化

前瞻——活动模式分类（KL聚类）

$$
\begin{aligned}
KL(\mathbf{P}\|\mathbf{Q}) &= \sum_{k=1}^{T}\sum_{i=0}^{d} p(x_k=i)\log p(x_k=i) \\
&\quad -\sum_{k=1}^{T}\sum_{i=0}^{d} p(x_k=i)\log q(x_k=i) \\
&= -H(\mathbf{P}) - \sum_{k=1}^{T}\sum_{i=0}^{d} \frac{\sum_{I^j\in\mathcal{I}} \mathbf{1}_{I_k^j=i}}{|\mathcal{I}|}\log q(x_k=i) \\
&= -H(\mathbf{P}) - \frac{1}{|\mathcal{I}|}\sum_{I^j\in\mathcal{I}}\sum_{k=1}^{T}\log q(x_k=I_k^j) \\
&= -H(\mathbf{P}) - \frac{1}{|\mathcal{I}|}\log P(\mathcal{I}|\mathbf{Q}),
\end{aligned}
$$

衡量相同事件空间内两个概率分布的差异

Li Z, Ding B, Han J, et al. Mining periodic behaviors for moving objects[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010:1099-1108

基于活动密集度的轨迹语义化

意义

城市人口密度分析和区域划分

行为模式 → 轨迹预测

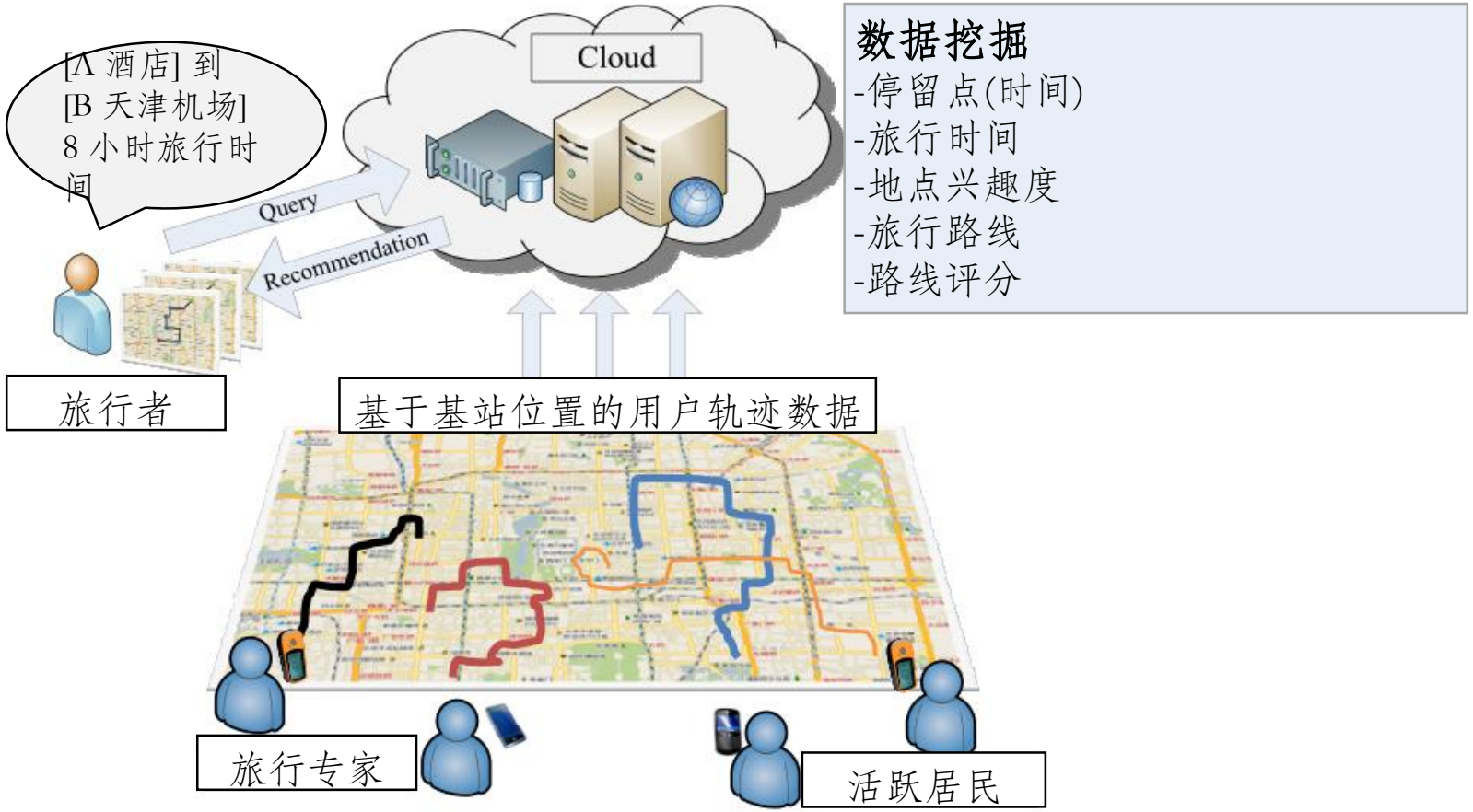
基站GPS数据分析与可视化

概览

1. 基于关键点提取和时序分析的轨迹语义化
2. 基于活动密集度的轨迹语义化
3. 用户行程推荐与可视化

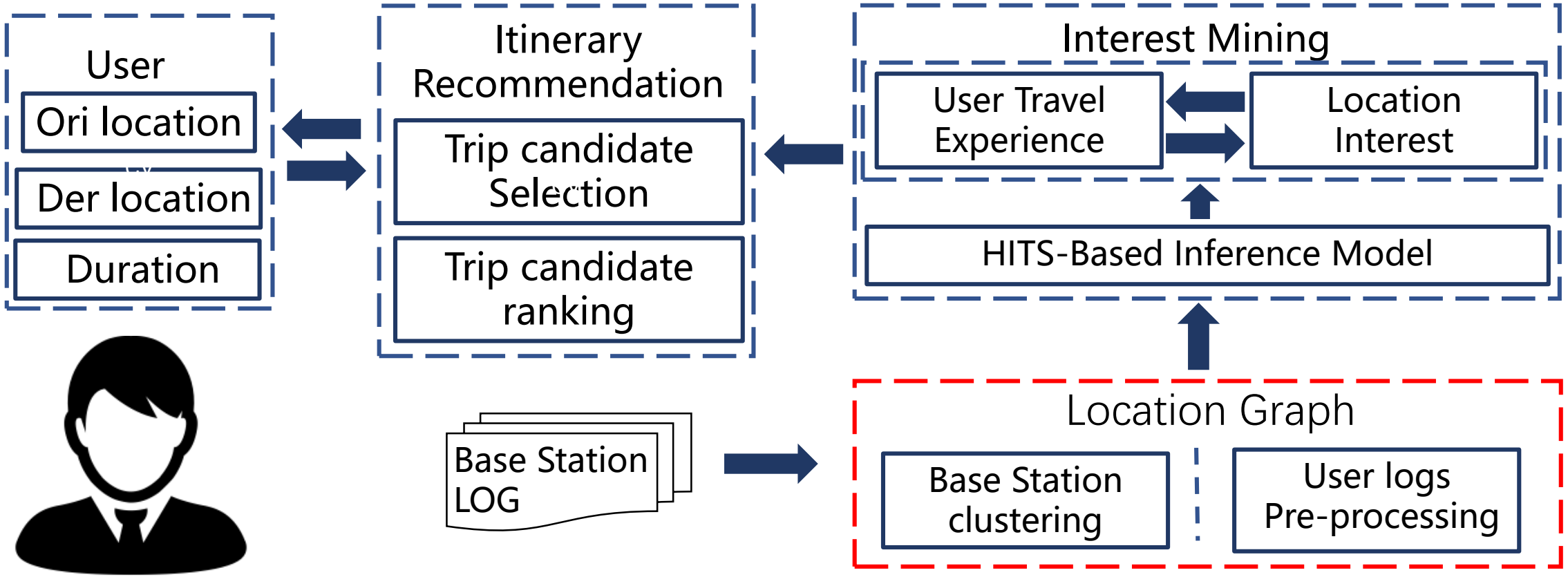
用户行程推荐

概念图


[A 酒店] 到
[B 天津机场]
8 小时旅行时间
Query
Recommendation
Cloud
数据挖掘
-停留点(时间)
-旅行时间
-地点兴趣度
-旅行路线
-路线评分
旅行者
基于基站位置的用户轨迹数据
旅行专家
活跃居民

用户行程推荐

系统结构


User
Ori location
Der location
Duration
Itinerary Recommendation
Trip candidate Selection
Trip candidate ranking
Interest Mining
User Travel Experience
Location Interest
HITS-Based Inference Model
Location Graph
Base Station clustering
User logs Pre-processing
Base Station LOG

用户行程推荐

数据清洗

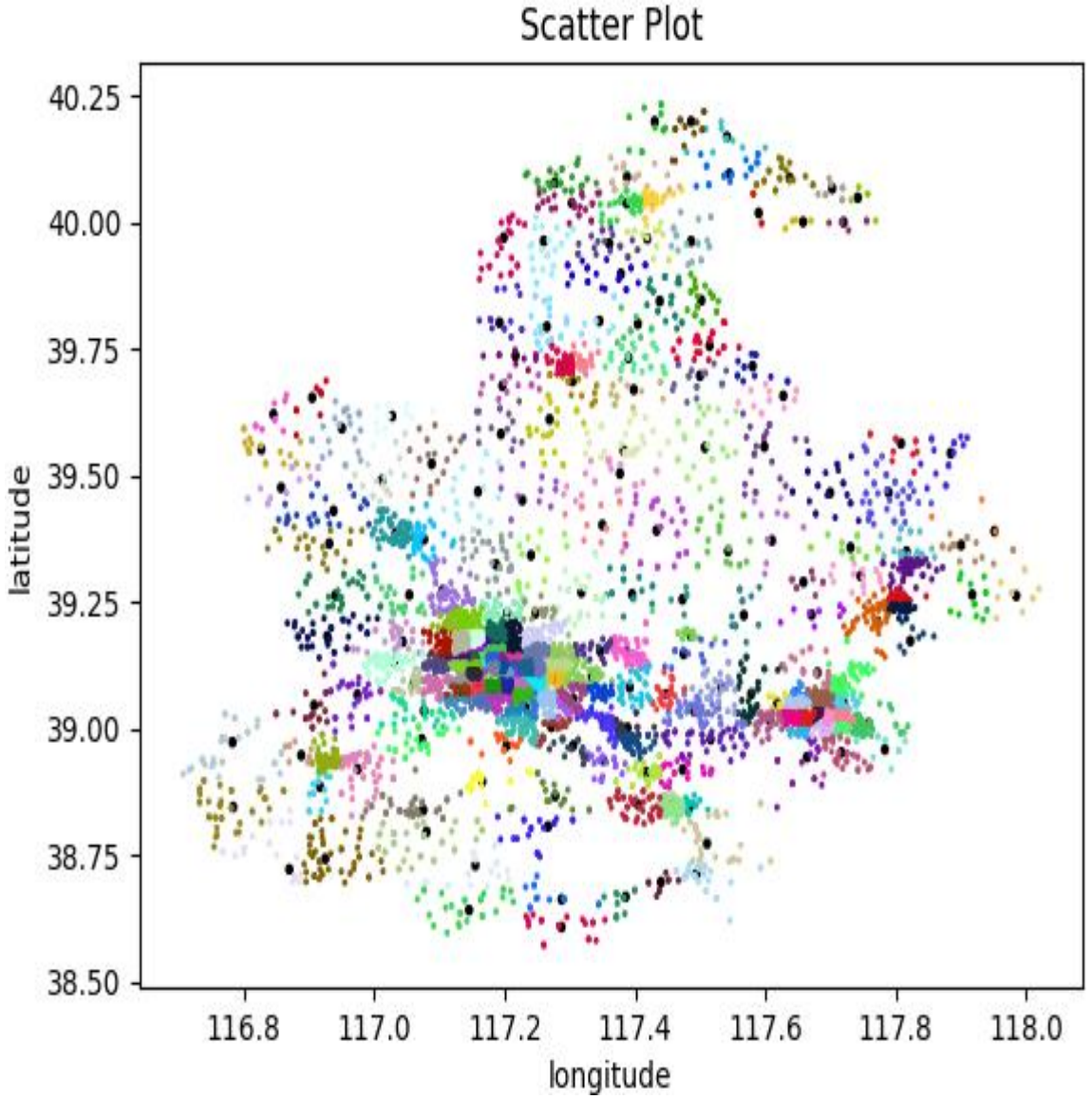
- 原始数据如图所示
- 每位用户都有大量基站GPS
- 从中筛选出连接持续时间大于5分钟、连接开始时间为每天6：30以后的记录
- 得到每个用户在每个基站访问的总时间、每个基站被每个用户访问的总时间

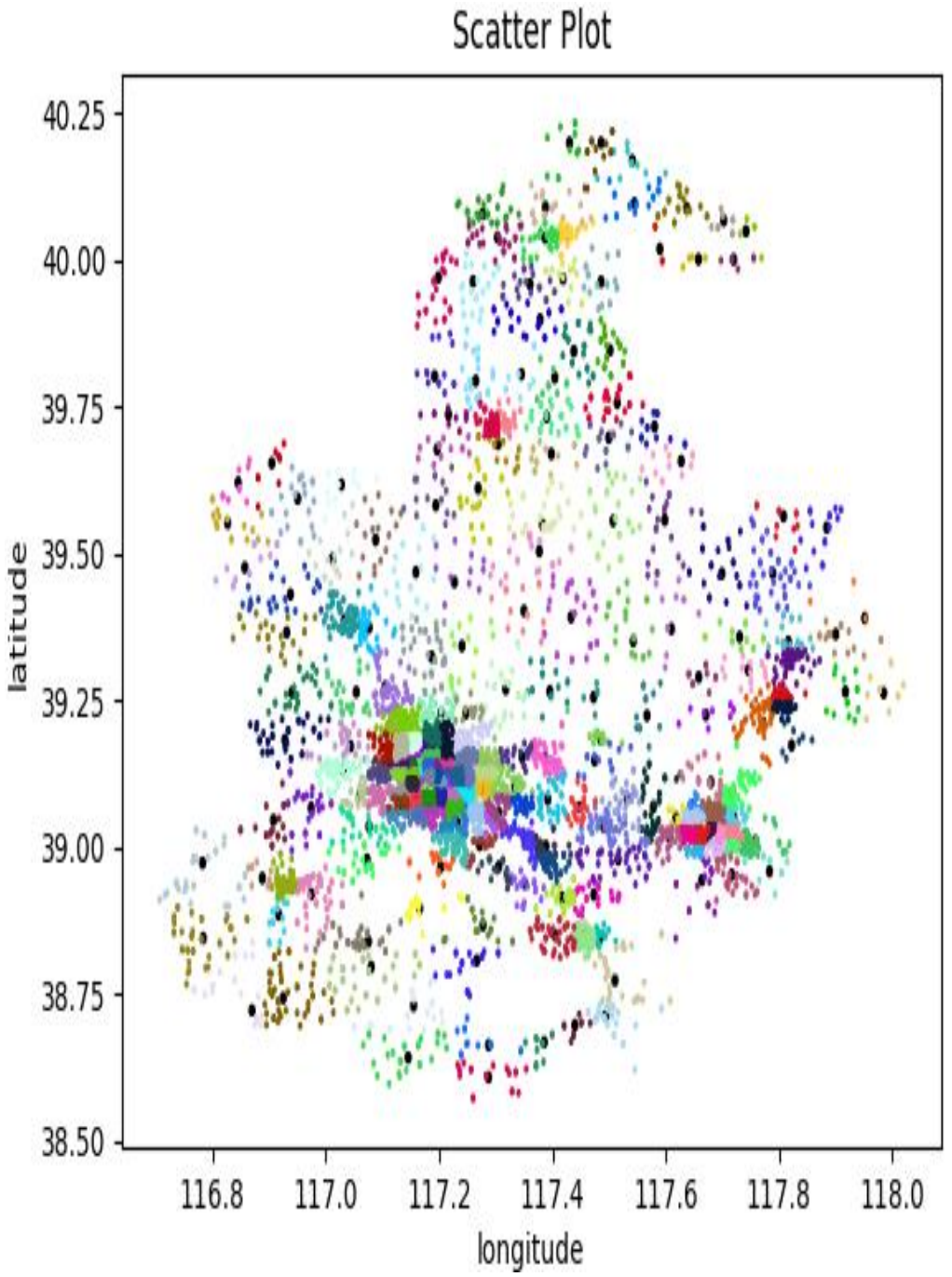
44054	11353	117.248450	39.080030	20140506000028	20140506002106	1238
44081	24363	117.361440	39.015360	20140506002546	20140506003652	666
1317	58623	117.375681	38.986116	20140506003809	20140506003852	43
44081	34201	117.370760	38.981610	20140506010626	20140506073433	23287
44081	24242	117.363340	38.992200	20140506075817	20140506110823	11406
44081	24201	117.370760	38.981610	20140506112542	20140506120039	2097
44081	34242	117.363340	38.992200	20140506120501	20140506122703	1322
44081	24242	117.363340	38.992200	20140506123215	20140506153705	11090
44081	24193	117.444470	38.986560	20140506160440	20140506162648	1328
44081	14193	117.444470	38.986560	20140506164906	20140506170843	1177
44054	20351	117.349130	39.027450	20140506171633	20140506172551	558
44057	11211	117.234657	39.091129	20140506172646	20140506172646	0
44054	11083	117.239111	39.092049	20140506172705	20140506172726	21
44057	11372	117.233250	39.096220	20140506172808	20140506172812	4
44057	11373	117.233250	39.096220	20140506172838	20140506211026	13308
44057	21103	117.221992	39.100487	20140506211943	20140506225901	5958
44054	11083	117.239111	39.092049	20140506225932	20140506230138	126
44054	11402	117.240300	39.083750	20140506230206	20140506231026	500
44081	24421	117.364920	39.022460	20140506231202	20140506231215	13
44081	34183	117.404200	39.002840	20140506231442	20140506232716	754
44081	24242	117.363340	38.992200	20140506233422	20140506235255	1113
1317	50331	117.365193	38.983627	20140506235332	20140506235357	25

用户行程推荐

基站聚类

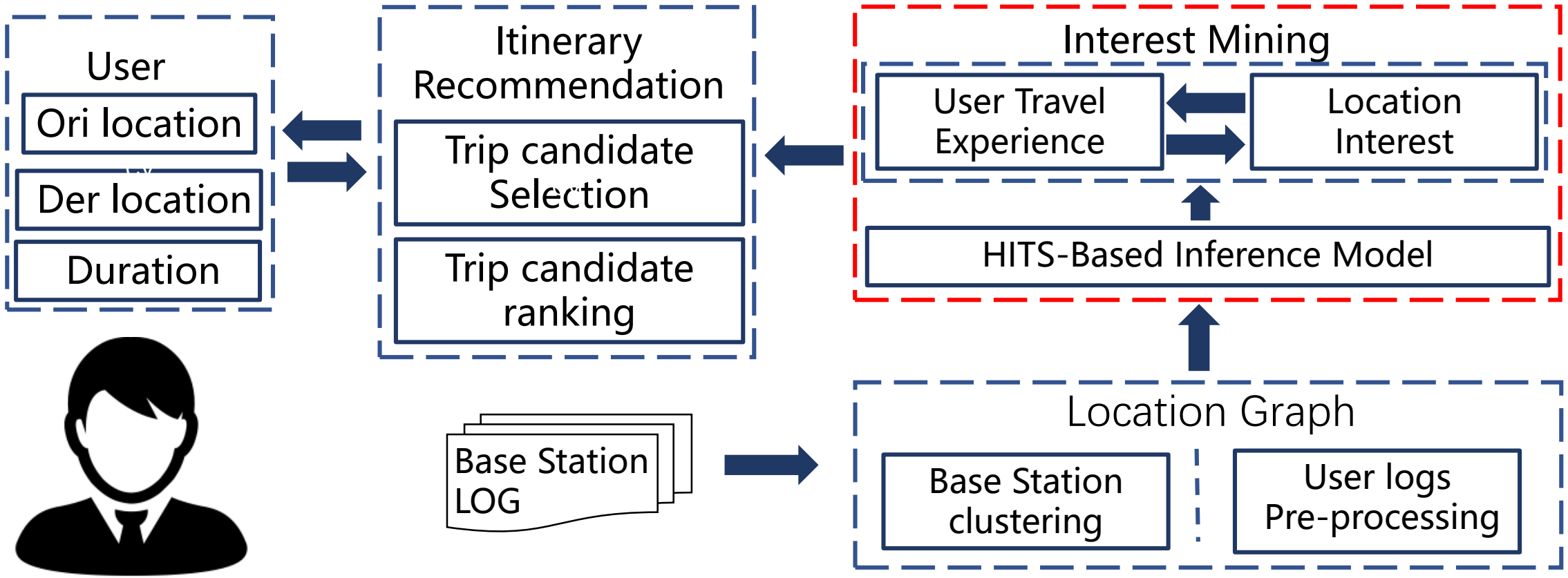
- 运用K-means算法
- 首先聚成100个location
- 将所有包含基站数超过阈值的location进行分割，直到其包含基站数不超过阈值（保证流量密集区域的划分粒度足够细致）
- 最后聚类的效果如右图所示




Scatter Plot
latitude
longitude
40.25
40.00
39.75
39.50
39.25
39.00
38.75
38.50
116.8
117.0
117.2
117.4
117.6
117.8
118.0

用户行程推荐

系统结构


User
Ori location
Der location
Duration
Itinerary Recommendation
Trip candidate Selection
Trip candidate ranking
Interest Mining
User Travel Experience
Location Interest
HITS-Based Inference Model
Base Station LOG
Location Graph
Base Station clustering
User logs Pre-processing

用户行程推荐

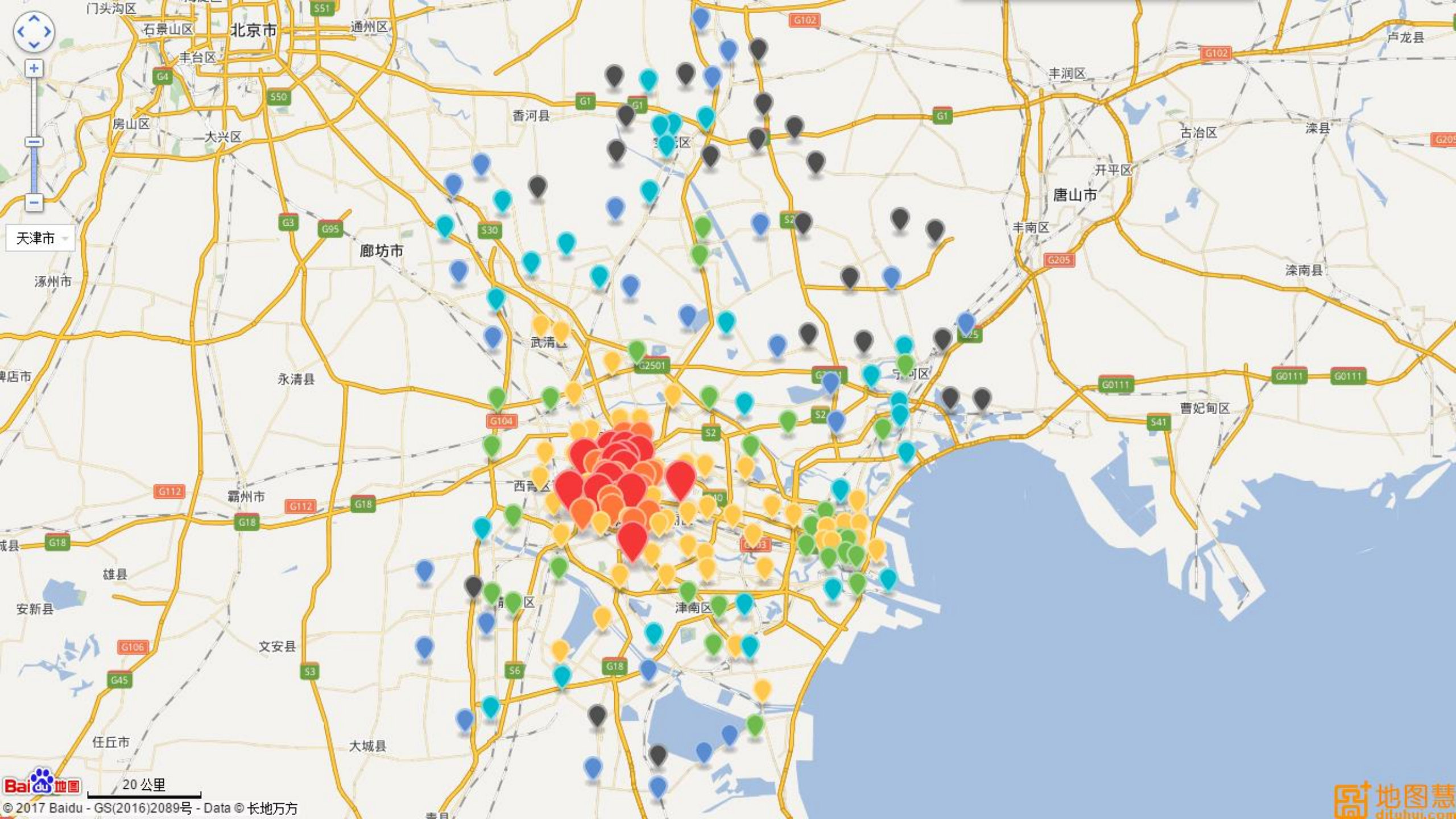
兴趣度计算

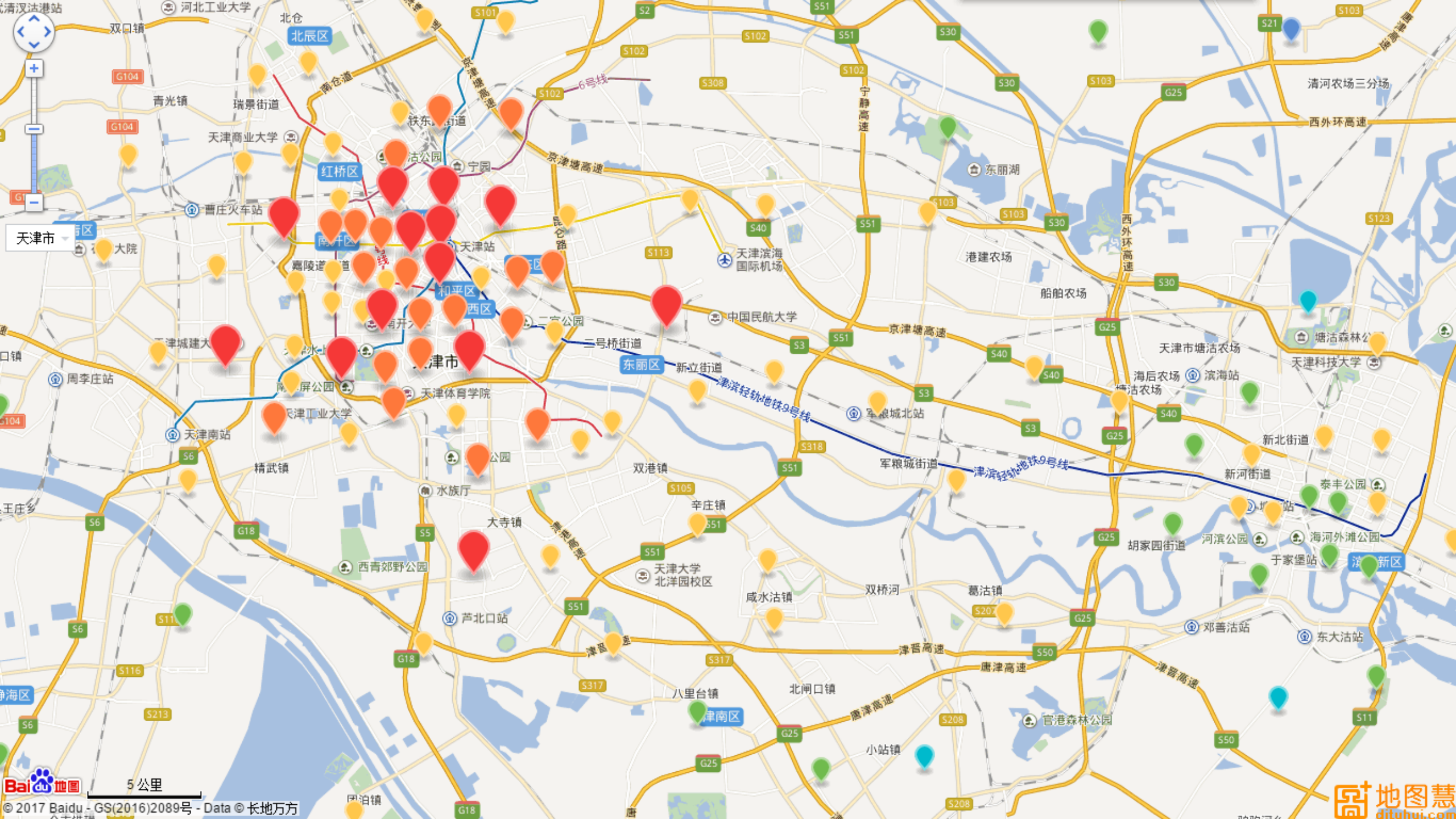
- 首先得到user-location访问时间矩阵

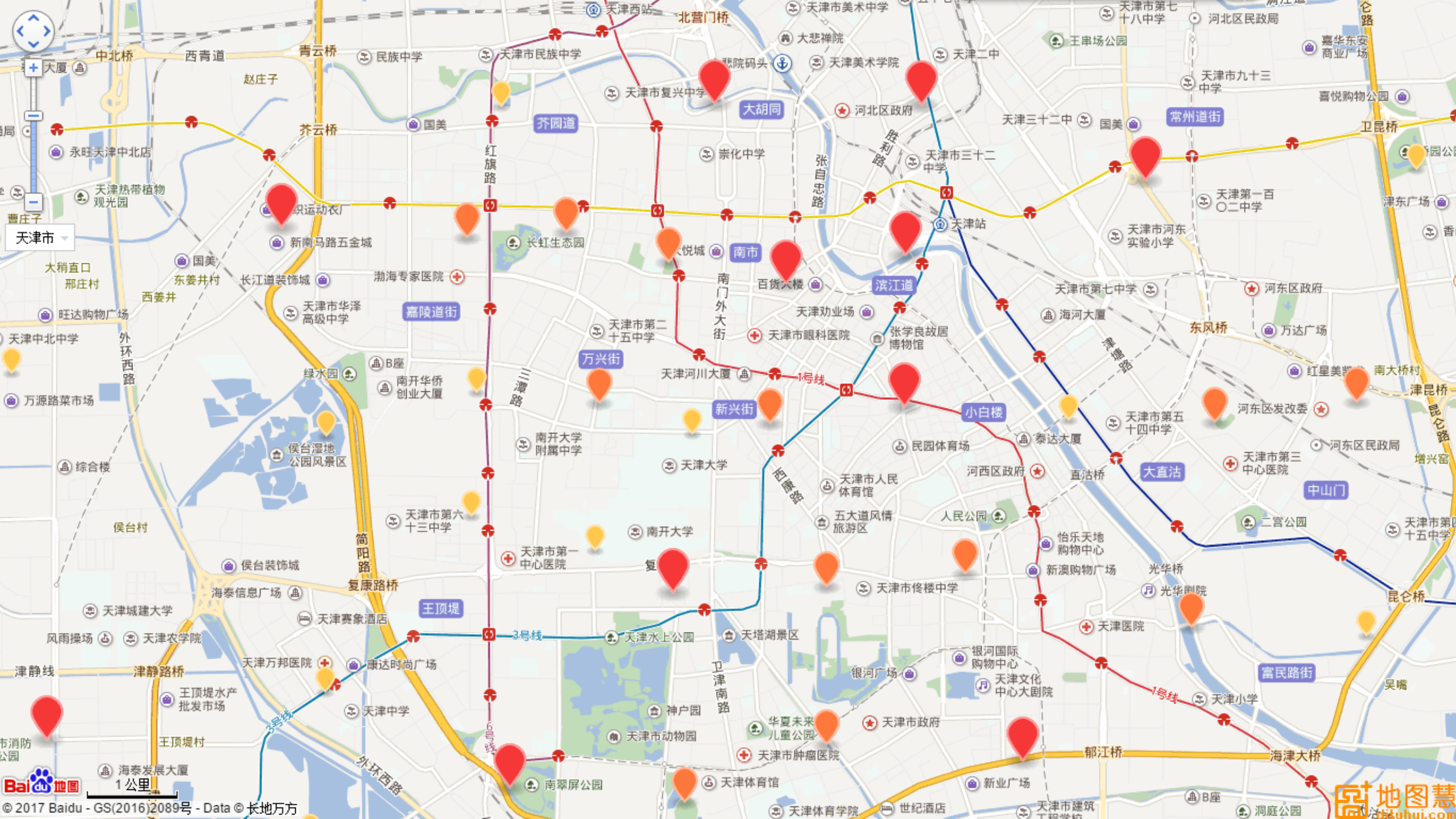
$$M = \begin{array}{c} \\ u_1 \\ u_2 \\ u_3 \\ u_4 \end{array} \begin{array}{c} \begin{array}{ccccc} c_{31} & c_{32} & c_{33} & c_{34} & c_{35} \end{array} \\ \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 2 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \end{array}$$

- 每个location

$$\begin{cases} I_j = \sum_{ui \in U} r_{ji} \times e_i \\ e_i = \sum_{lj \in L} r_{ij} \times I_j \end{cases} \Rightarrow \begin{cases} \mathrm{I}_t = \mathrm{M} \times \mathrm{E}_{t-1} \\ \mathrm{E}_t = \mathrm{M}^T \times \mathrm{I}_{t-1} \end{cases} \Rightarrow \mathrm{I}_t = \mathrm{M} \times \mathrm{M}^T \times \mathrm{E}_{t-1}$$

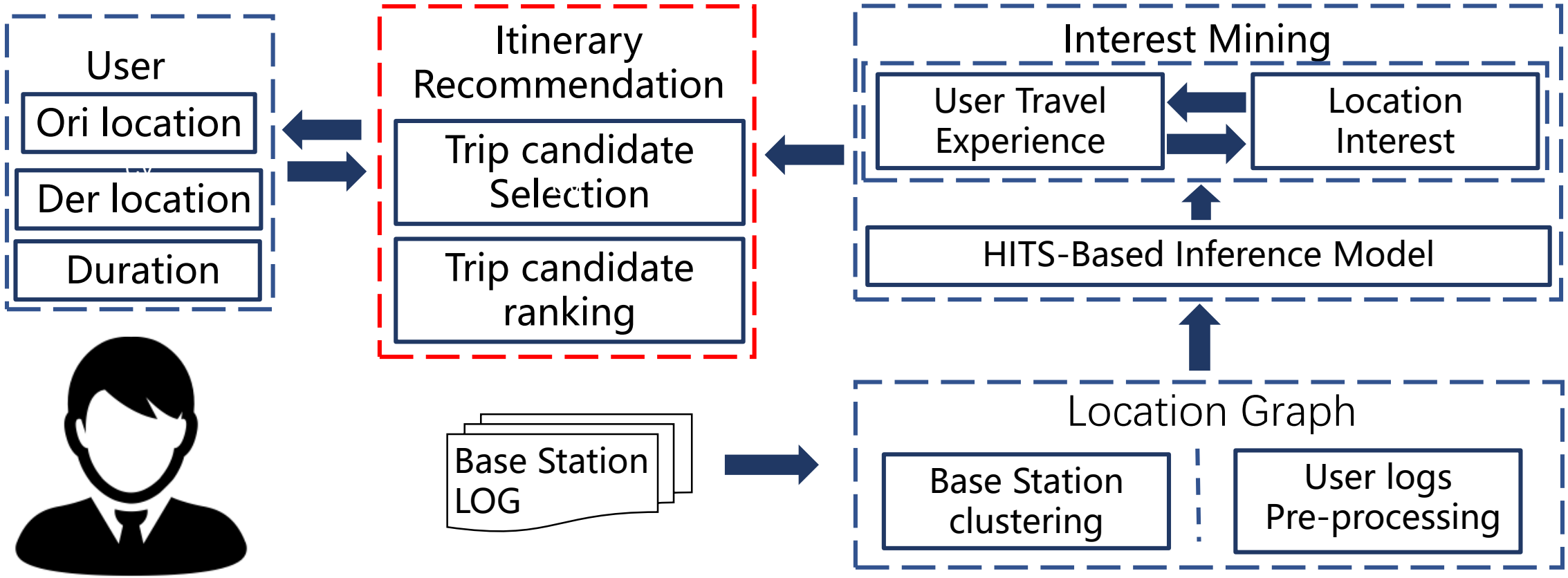

北京市
天津市
廊坊市
唐山市
通州区
丰台区
房山区
大兴区
石景山区
门头沟区
香河县
丰润区
古治区
开平区
丰南区
卢龙县
滦县
滦南县
曹妃甸区
涿州市
永清县
霸州市
雄县
安新县
文安县
大城县
任丘市
津南区
武清区
西青区
20 公里
© 2017 Baidu - GS(2016)2089号 - Data © 长地万方


天津市
北辰区
红桥区
和平区
西区
东丽区
南开区
津南区
河北工业大学
双口镇
北仓
青光镇
瑞景街道
天津商业大学
铁东街道
曹庄火车站
大院
嘉陵道街道
天津站
周李庄站
天津城建大学
天津南站
天津工业大学
天津体育学院
南开大学
二宁公园
二号桥街道
新立街道
中国民航大学
天津滨海国际机场
东丽湖
港建农场
船舶农场
天津市塘沽农场
海后农场
滨海站
塘沽森林公园
天津科技大学
新北街道
新河街道
泰丰公园
胡家园街道
河滨公园
海河外滩公园
于家堡站
滨海新区
邓善沽站
东大沽站
官港森林公园
军粮城北站
军粮城街道
双港镇
辛庄镇
大寺镇
水族厅
精武镇
西青郊野公园
芦北口站
天津大学北洋园校区
咸水沽镇
双桥河
葛沽镇
八里台镇
北闸口镇
小站镇
团泊镇
清河农场三分场
西外环高速
京津塘高速
津滨轻轨地铁9号线
津晋高速
唐津高速
津保高速
京津高速
宁静高速
昆仑路
6号线
5公里
© 2017 Baidu - GS(2016)2089号 - Data © 长地万方
地图慧
dituhui.com


天津市
北营门桥
天津西站
天津市美术中学
天津市第七十八中学
河北区民政局
王串场公园
嘉华东安商业广场
大悲禅院
悲院码头
天津美术学院
天津二中
天津市九十三中学
喜悦购物公园
天津市复兴中学
大胡同
河北区政府
天津三十二中
国美
常州道街
卫昆桥
中北桥
西青道
青云桥
民族中学
天津市民族中学
赵庄子
芥园道
芥云桥
永旺天津中北店
崇化中学
红旗路
张自忠路
胜利路
天津市三十二中学
天津热带植物观光园
天津第一百〇二中学
津东广场
曹庄子
织运动衣厂
新南马路五金城
长虹生态园
南市
天津站
天津市河东实验小学
大稍直口
邢庄村
东姜井村
西姜井
国美
长江道装饰城
渤海专家医院
百货大楼
滨江道
天津市第七中学
河东区政府
旺达购物广场
嘉陵道街
天津市华泽高级中学
天津劝业场
海河大厦
东风桥
万达广场
天津中北中学
外环西路
天津市第二十五中学
天津市眼科医院
张学良故居博物馆
津塘路
万兴街
天津河川大厦
1号线
B座
绿水园
南开华侨创业大厦
南大桥村
津昆桥
昆仑路
万源路菜市场
三潭路
新兴街
小白楼
天津市第五十四中学
泰达大厦
河东区发改委
侯台湿地公园风景区
南开大学附属中学
民园体育场
综合楼
天津大学
西康路
天津市人民体育馆
河西区政府
直沽桥
大直沽
天津市第三中心医院
河东区民政局
中山门
增兴窑
侯台村
天津市第六十三中学
五大道风情旅游区
人民公园
二宫公园
天津市第十五中学
南开大学
天津市第一中心医院
怡乐天地购物中心
新澳购物广场
光华桥
简阳路
复康路桥
侯台装饰城
天津市佟楼中学
光华剧院
海泰信息广场
昆仑桥
王顶堤
天津城建大学
天津赛象酒店
天津医院
风雨操场
天津农学院
天津水上公园
天塔湖景区
3号线
银河国际购物中心
津静线
津静路桥
天津万邦医院
康达时尚广场
银河广场
天津文化中心大剧院
富民路街
王顶堤水产批发市场
天津中学
天津小学
吴嘴
神户园
卫津南路
天津市动物园
华夏未来儿童公园
天津市政府
1号线
郁江桥
海津大桥
海泰发展大厦
王顶堤村
外环西路
南翠屏公园
天津体育馆
新业广场
天津体育学院
世纪酒店
B座
洞庭公园
1公里
© 2017 Baidu - GS(2016)2089号 - Data © 长地万方
地图慧

用户行程推荐

系统结构


User
Ori location
Der location
Duration
Itinerary Recommendation
Trip candidate Selection
Trip candidate ranking
Interest Mining
User Travel Experience
Location Interest
HITS-Based Inference Model
Location Graph
Base Station clustering
User logs Pre-processing
Base Station LOG

用户行程推荐

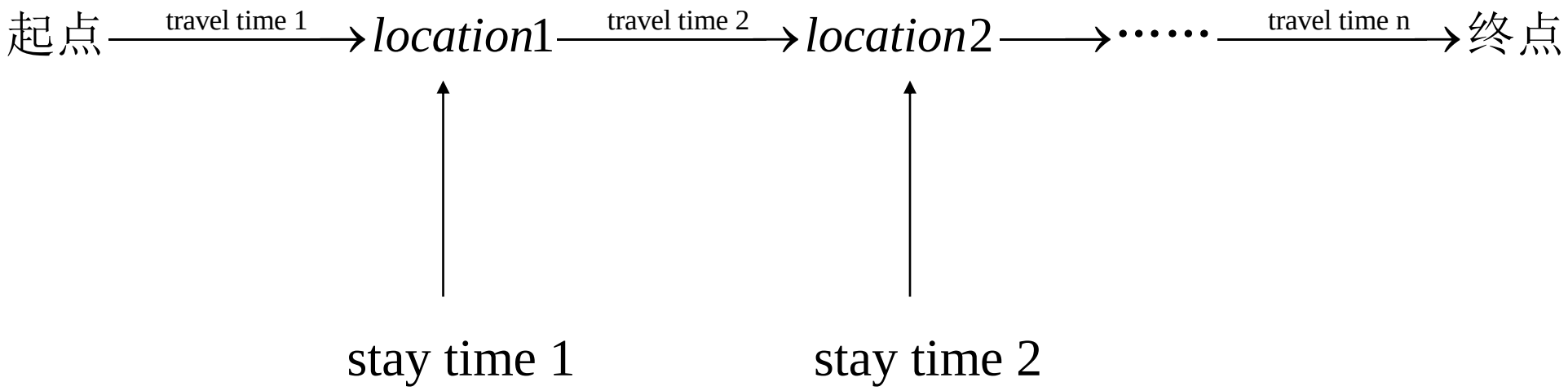
行程推荐

起点
终点
最大旅行时间 } ⇒ 最佳行程路线

- 如何表示一条行程路线？
- 如何评价一条行程路线？
- 如何获取一条最好的行程路线？

用户行程推荐

如何表示一条行程路线？



用户行程推荐

如何评价一条行程路线？

$$score = \sqrt{(\text{staytime ratio})^2 + (\text{interest})^2}$$

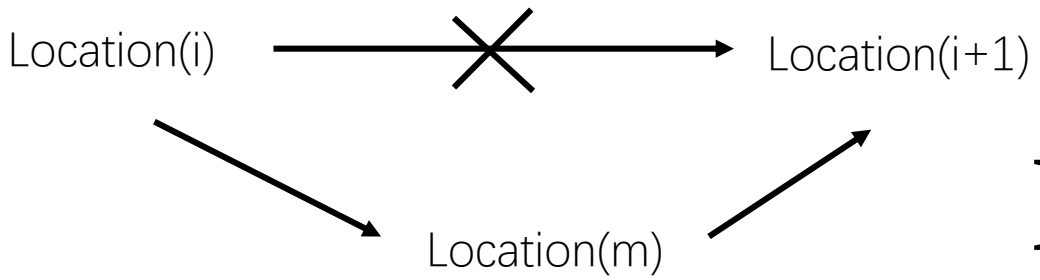
$$\text{其中，} \ \text{staytime ratio} = \frac{\text{所有}location\text{停留总时间}}{\text{给定最大旅行时间}}$$

$$\text{interest} = \text{所有路线中location兴趣度的总和}$$

用户行程推荐

如何获取一条最好的行程路线？

- 初始化：Itinerary = 起点→终点
- 每一次更新：向现有路径中加入一个location，使更新后路线的评分最高



$$\max_{l_m \in L, e_i \in E} Score$$

- 终止条件：对任何更新选择，更新后路线评分低于更新前评分；或总时间高于最大旅行时间，更新结束，此时路线为最优路线。

用户行程推荐

效果展示

起点：徐官屯街道

终点：张家窝村

旅行时间：10小时

基站GPS数据分析与可视化

回顾

1. 基于关键点提取和时序分析的轨迹语义化
2. 基于活动密集度的轨迹语义化
3. 用户行程推荐与可视化

谢谢！