更多内容,请关注





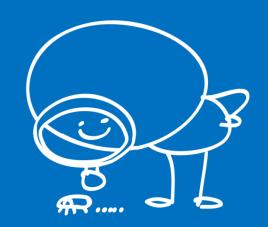
事件分析技术基础与 应用(二) LBS数据探索与挖掘





目录

- 1 LBS中的空间思维
- 2 LBS数据分析模式



3 案例:基于Spark的LBS数据处理



LBS





北大保安哥的传奇









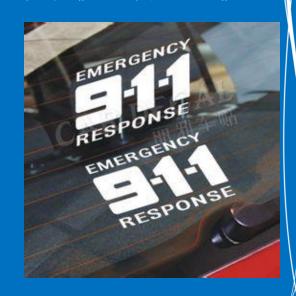
定位是生物本能





LBS:基于位置的服务(Location base services)

- 又称为定位服务,指通过移动终端和移动网络的配合,确定移动用户的实际地理位置,从而提供用户所需要的与位置相关的服务信息的一种移动通信与导航融合的服务形式。
- 最早可以追溯到美国70年代颁布的《911服务规范》





LBS内涵

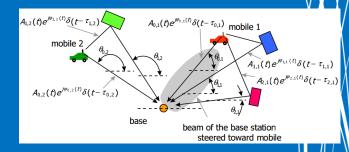
- 位置服务的两个内涵:
 - 让我知道自己所在的位置
 - 让我知道我需要知道的对象(人、物、事件)所在的位置
- 基于以上,提供相应的指示或者建议服务。

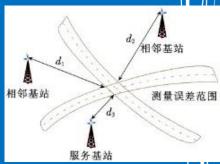


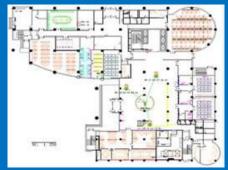


LBS的定位方式

- A)AOA(angle of arrival)指通过 两个基站的交集来获取移动台 (Mobile station)的位置;
- B)TDOA(time difference of arrival)工作原来类似与GPS。通过一个移动台和多个基站交互的时间差来定位;
- C)location signature位置标记。
 对每个位置区进行标识来获取位置;
- D)卫星定位。



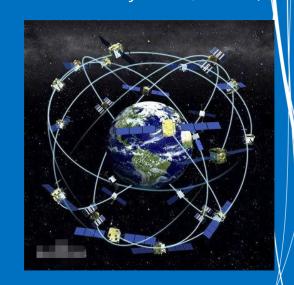






全球四大卫星定位系统

- 美国: GPS (Global Positioning System)
- 俄罗斯: GLONASS (Глобальная навигационная спутниковая система)
- 欧洲: GALILEO (Galileo satellite navigation system)
- 中国: 北斗 (BeiDou Navigation Satellite System, BDS)





LBS数据

- > 空间数据
- > 时间数据
- ▶ 唯一ID
- 附加的其他信息 (可选)





GPS信息记录

```
20,116.15400-
           150000,116.4858704,35.
        1104150000,116.4551315,39.90
       121104150000,116.3755188,39.94826
     40121104150000,116.3961487,40.004398
    20121104150000,116.3363495,40.0271797,
  0,20121104150000,116.3597641,39.9424210,0
  .0.20121104150000.116.4993362.40.0363541.0.
4,0,20121104150000,116.3597641,39.9424210,0,3
4,0,20121104150000,116.3243866,40.0275116,0,1
,4,0,20121104150000,116,2189102,40,0380287,0,0,
,4,0,20121104150000,116.3703995,39.9312477,0,80
,4,0,20121104150000,116.3714676,39.9672852,0,19
,4,0,20121104150000,116.1761169,40.0732231,0,26
,4,0,20121104150001,116.3197632,39.9759026,0,96
4,0,20121104150001,116.4296799,39.9059715,34,3
4,0,20121104150001,116.4901733,39.9707565,0,0
  ,0,20121104150001,116.1754684,40.0729752,0,
   0,20121104150002,116.5566177,40.1080360,0
    20121104150002,116.3280029,40.0266418,
     20121104150002,116.2178345,40.0741158
       121104150002,116.3408508,39.98559
         1104150002,116.4883194,39.912
            1150002,116.4994965,39
                202 116.289184
```











公共交通记录







2012-3-30 7:47	储值消费	八方达客运公司820主线	上车站: 0023 -> 下车站:0037
2012-3-28 17:38	储值消费	北京巴士股份有限公司11特	上车站: 0028 -> 下车站:0030
2012-3-14 15:18	储值消费	北京巴士股份有限公司11特	上车站: 0031 -> 下车站:0033
2012-3-14 15:04	储值消费	第二客运分公司652主线	上车站: 0024 -> 下车站:0021
2012-3-14 14:22	储值消费	北京巴士股份有限公司741主线	上车站: 0013 -> 下车站:0010
2012-3-14 8:56	储值消费	第四客运分公司40主线	上车站: 0000 -> 下车站:0001
2012-3-14 8:30	储值消费	第二客运分公司71主线	上车站: 0000 -> 下车站:0030
2012-3-9 18:24	储值消费	八方达客运公司820主线	上车站: 0023 -> 下车站:0026
2012-3-7 17:31	储值消费	第二客运分公司69主线	上车站: 0000 -> 下车站:0027
2012-3-7 17:05	储值消费	地铁2号线	上车站: 0003 -> 下车站:0008
2012-3-7 16:41	储值消费	北京巴士殷份有限公司714主线	上车站: 0020 -> 下车站:0022
2012-3-7 13:55	储值消费	北京巴士殷份有限公司714主线	上车站: 0022 -> 下车站:0021
2012-3-7 13:04	储值消费	地铁2号线	上车站: 0061 -> 下车站:0003
2012-3-7 12:27	储值消费	北京巴士股份有限公司11特	上车站: 0033 -> 下车站:0029
2012-3-7 9:49	储值消费	北京巴士殷份有限公司11特	上车站: 0025 -> 下车站:0026
	2012-3-28 17:38 2012-3-14 15:18 2012-3-14 15:04 2012-3-14 14:22 2012-3-14 8:56 2012-3-14 8:30 2012-3-9 18:24 2012-3-7 17:31 2012-3-7 17:05 2012-3-7 16:41 2012-3-7 13:55 2012-3-7 13:04 2012-3-7 12:27	2012-3-28 17:38 储值消费 2012-3-14 15:18 储值消费 2012-3-14 15:04 储值消费 2012-3-14 14:22 储值消费 2012-3-14 8:56 储值消费 2012-3-14 8:30 储值消费 2012-3-9 18:24 储值消费 2012-3-7 17:31 储值消费 2012-3-7 17:05 储值消费 2012-3-7 16:41 储值消费 2012-3-7 13:04 储值消费 2012-3-7 13:04 储值消费 2012-3-7 13:04 储值消费 2012-3-7 13:04 储值消费	2012-3-28 17:38 储值消费 北京巴士服份有限公司11特 2012-3-14 15:18 储值消费 北京巴士服份有限公司11特 2012-3-14 15:04 储值消费 第二客运分公司652主线 2012-3-14 14:22 储值消费 北京巴士服份有限公司741主线 2012-3-14 8:56 储值消费 第四客运分公司40主线 2012-3-14 8:30 储值消费 第二客运分公司71主线 2012-3-9 18:24 储值消费 从方达客运公司820主线 2012-3-7 17:31 储值消费 第二客运分公司69主线 2012-3-7 17:05 储值消费 地铁2号线 2012-3-7 16:41 储值消费 北京巴士服份有限公司714主线 2012-3-7 13:04 储值消费 北京巴士服份有限公司714主线 2012-3-7 13:04 储值消费 地铁2号线 2012-3-7 13:04 储值消费 地铁2号线



电商物流信息

23:36 【广州番禺洛溪】的收件员【陈镜油】已收件 23:52 由【广州番禺洛溪】发往【番禺中转仓】 快件已到达【番禺中转仓】 扫描员是【蔡建锋】上一站是【广州番禺洛溪】 23:31 25:14 【番禺中转仓】正在进行【装袋】扫描 22:48 【广州番禺洛溪】已进行【运单图片】扫描。【】 37:17 【北京分拨中心】已进行【拆袋】扫描、【总部】 57:24 由【北京分拨中心】发往【北京东城】 52:07 快件已到达【北京】 扫描员是【常变宁】上一站是【北京分拨中心】 17:19 快件已到达【北京东城】 扫描员是【北京东城代派】上一站是【北京东城】 86:25 【北京东城】已收入 36:25 【北京东城】的派件员【划)方仁】正在派件 32:20 已签收,签收人是草签 【北京东城】已收入 37:28 37:28 【北京东城】的派件员【刘振彪】正在派件 卖家已发货 6:08

信息来源: 淘宝 运单号: 2380294835











更多……





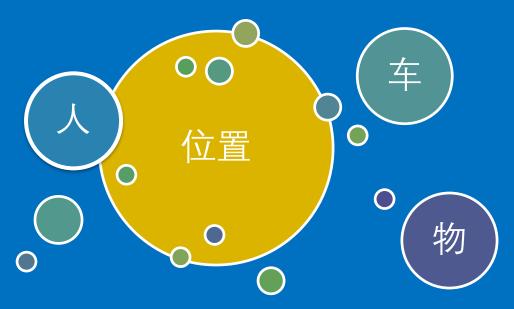






LBS分析的意义

• 城市是由人组成的。





LBS分析的意义





LBS分析的意义





LBS



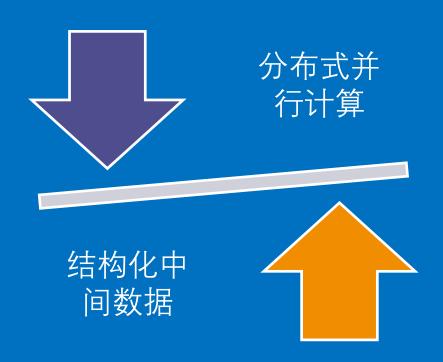


LBS分析的特点



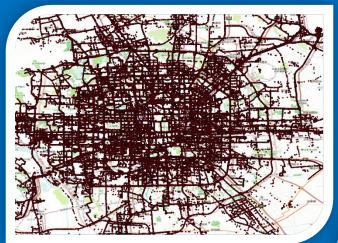


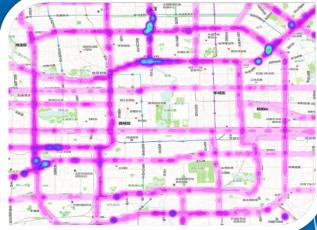
海量数据处理的最佳方法





定性分析







定量分析





LBS

案例:基于 Spark的LBS数 据处理 ——以车辆轨迹数据为 例



轨迹研究

轨迹的是一个非常庞大的研究领域

轨迹计算

- •轨迹面积
- •轨迹廊道
- •轨迹地图
- •轨迹压缩
- •轨迹分段

轨迹预测

- •轨迹函数拟合
- •轨迹插值
- •轨迹修正
- •轨迹预测
- •预测评估
- •不确定性模拟

轨迹通用分析

- 轨迹伴随关系
- •异常轨迹识别
- •轨迹停留区域
- •轨迹离群识别
- •轨迹关键识别
- •目的地预测
- •轨迹规划
- 轨迹隐私

轨迹检索

- •轨迹指数
- •轨迹相似度
- •轨迹查询

轨迹模式分析

- •相对运动模式
- •基干距离的模式
- •基干位置的模式
- •轨迹度量模式
- •轨迹复杂度模式
- •轨迹空间自相关性
- •轨迹迁移模式
- •轨迹周期模式
- •轨迹语义模式
- •轨迹聚类
- •轨迹分类

行业专题分析

- •社交网络轨迹应用
- •公共安全轨迹应用
- •交通轨迹应用 •文化经济
- •生物生态
- •军事
- •自然气象

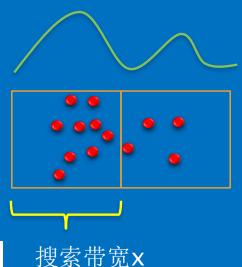
轨迹预处理

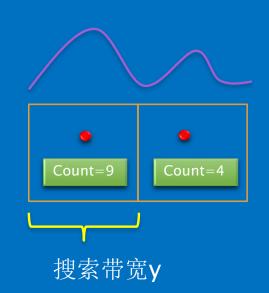
- •轨迹过滤
- •滤波/去燥
- •轨迹采样



海量数据的LBS密度分析流程

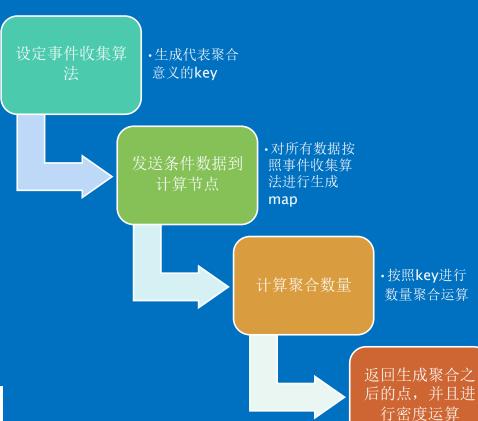
- 因为核函数的特征,所以使之点事件的聚合成为核密度 函数的重要前置条件
- 示例: 当x = y的时候,以下两种情况生成的核密度曲面 完全一样。







点聚合的流程



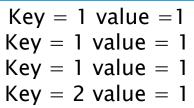


算法示例:map部分

• 利用空间包含算法进行事件收集















算法示例: redcue部分

Key = 1 value = 1

Key = 1 value = 1Key = 1 value = 1

Key = 2 value = 1

Key = 2 value = 1

....

reduceByKey



Key = 1 value = 1 Key = 1 value = 1 Key = 1 value = 1

.....



Key = 1 value = 3Key = 2 value = 2

Key = 2 value = 1 Key = 2 value = 1



挂接属性计算核密度

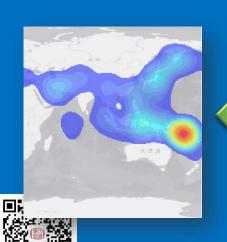
UID = 1

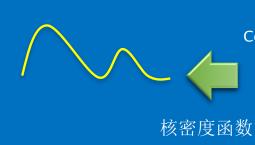
UID = 2

JOIN



Key = 1 value = 3Key = 2 value = 2







设定合理的Key进行统计信息提取

总数量信息

每个研究区的信息 总量

每个时段的信息总 量

研究区的最高、最 低、平均时速

研究区的停泊、行 进车辆数据

不同区域的车辆流 动情况

单个车辆的信息统计

单个车辆的轨迹信息 提取

单个车辆的驾驶行为 特征

分量统计

总量统计



细节统计

单个统计量的设计(一)

• 每个区域的信息总量统计

Мар

- key: 研究区域的id和车辆的id组合
- value:每个区域内每个记录记为1

reduceByKey

• 按key进行进行过滤,保证每个区域内每辆车 只留下一条记录

Мар

- key: 每个研究区的id
- value: 每条记录计数1



reduceByKey

• 按key进行累加

'单个统计量的设计(二)

• 时间段的信息总量统计

Мар

- key: 研究时段和车辆的id组合
- value:每个区域内每个记录记为1

reduceByKey

• 按key进行进行过滤,保证每个时段内每辆车 只留下一条记录

Мар

- key: 每个时段的id
- value: 每条记录计数1



reduceByKey

• 按key进行累加

单个统计量的设计(三)

• 每个区域内的车辆最高、最低时速信息统计

Мар

- key: 研究区域id
- value:每个区域内每个记录 的时速记录

reduceByKey

按key进行进行过滤,保证 每个区域内每辆车只留下速 度最高(最低)的一条记录



单个统计量的设计(四)

区域流动情况(OD分析)

Мар

- key: 车辆的 ID
- value:每个记录的时间戳和出现的区域

groupByKey

• 按车辆的ID 进行分组。

reduceByKey

• 对每个key里 面的分组数 据进行处理

Map

• 将每个车辆 按照区域变 换进行分类 组合

reduceByKey

累加value, 完成OD分析



• 数据情况:

车辆ID	时间	х	У
A01	2019-7-20 13:15:19	116.39	40.22
A01	2019-7-20 14:15:10	116.44	40.13
A01	2019-7-20 14:17:33	117.04	41.44
A01	2019-7-20 14:25:19	116.44	40.13
A01	2019-7-20 15:44:19	118.42	41.11
A01	2019-7-20 16:12:11	118.45	41.23
A01	2019-7-20 18:18:19	116.55	40.22
A02	2019-7-20 14:15:55		
A02	2019-7-20 18:33:44		
A03	2019-7-20 13:15:19		



- 步骤1: Map
 - 组合成以车辆ID为KEY,时间戳和区域ID为VALUE的组合

key	value
A01	(2019-7-20 13:15:19,区域1)
A01	(2019-7-20 14:15:10,区域1)
A01	(2019-7-20 14:17:33,区域2)
A01	(2019-7-20 14:25:19,区域1)
A01	(2019-7-20 15:44:19,区域3)
A01	(2019-7-20 16:12:11,区域3)
A01	(2019-7-20 18:18:19,区域1)
A02	(2019-7-20 14:15:55,区域1)
A02	(2019-7-20 18:33:44,区域1)
A03	



• 步骤2: groupbyKey

分组后, 顺序



第三步: reduceByKey(1)

```
A01
              (2019-7-20 16:12:11. 区域3).
                                          (2019-7-20 14:15:10. 区域1).
              (2019-7-20 14:25:19, 区域1).
                                          (2019-7-2014:17:33、区域2)、
                                          (2019-7-2013:15:19, 区域1),
              (2019-7-20 15:44:19, 区域3),
              (2019-7-20 18:18:19. 区域1)
```



按时间排序

```
A01
              (2019-7-20 13:15:19, 区域1),
                                           (2019-7-20 14:15:10, 区域1)
              (2019-7-20 14:17:33. 区域2).
                                          (2019-7-20 14:25:19. 区域1)
              (2019-7-20 15:44:19, 区域3).
                                           (2019-7-20 16:12:11, 区域3)
              (2019-7-20 18:18:19, 区域1)
```











• 第三步: reduceByKey (2)

```
A01 (2019-7-20 13:15:19, 区域1), (2019-7-20 14:15:10, 区域1) (2019-7-20 14:17:33, 区域2), (2019-7-20 14:25:19, 区域1) (2019-7-20 15:44:19, 区域3), (2019-7-20 16:12:11, 区域3) (2019-7-20 18:18:19, 区域1)
```







• 第三步: reduceByKey (3)

```
A01 (2019-7-20 13:15:19, 区域1), (2019-7-20 14:15:10, 区域1) (2019-7-20 14:17:33, 区域2), (2019-7-20 14:25:19, 区域1) (2019-7-20 15:44:19, 区域3), (2019-7-20 16:12:11, 区域3) (2019-7-20 18:18:19, 区域1)
```



















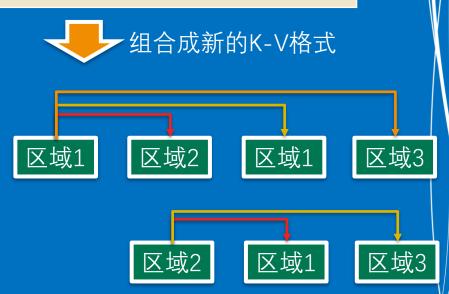


区域3

• 第四步: Map

A01 (区域1,区域2,区域1,区域3,区域1······)

key	value
区域1-区域2	1
区域1-区域1	1
区域1-区域3	1
区域1-区域1	1
区域2-区域1	1
区域2-区域3	1





• 第五步: ReduceByKey

Key	value
区域1-区域2	1
区域2-区域1	1
区域1-区域3	1
区域3-区域1	1
区域1-区域3	1
区域3-区域1	1



Key	value
区域1-区域2	34
区域2-区域1	76
区域1-区域3	55
区域3-区域1	21





车辆轨迹特征的提取

急刹车/百 速度 角度 加速度 公里 急加速/百 停车次数 驾驶时间 停车时长 公里 轨迹线性 停车位置 变速频率 驾歇间隔 规则度

行驶路型

路况应对



插播广告

• 所有的PPT、数据、文章、代码·····均可以通过此公众号 获取。

公众号:

虾神daxialu





谢谢!