

利用人类活动轨迹确定城市的功能区

每周学习总结

利用人类活动轨迹确定城市的功能区

1 Apriori Alg--- focus on 频繁项集 ,

2 Discovering Urban Functional Zones Using Latent Activity Trajectories

注：

1. Location Semantics
2. Mobility Semantics
3. Map Semantics
4. Discovery activities

3 主要参考文献

1 Apriori Alg--- focus on 频繁项集 ,

- 缺点：大数据集实现 慢，
- 适用场景：购物，投票，网站流量分析
- 适用数据类型：数值型，或者标称型数据

- 评估标准：**支持度(Support)**--**频繁** or **置信度(Confidence)**--**关联**
- 迭代的思想
 - 找出所有的包含1项的集合，小于支持度的去除
 - 将剩余的 1 项集 进行 连接（按照顺序），构成2 项集
 - 直到出现空，或者剩1
- 支持度计算
 - Hash 树进行支持度计算
- 项集连接（合并）

- $F_{k-1} * F_{k-1}$
 - 按照字典序
 - 遍历方式
 - 深度优先--- a-> ab-> abc -> 非频繁节点：便于确定 **极大频繁项集**
 - 广度优先
 - 扩展
 - 处理分类属性
 - 处理连续属性
 - 基于离散化
 - 基于统计学
 - 非离散化的---动态支持度(min-Apriori)
-

应用：发现毒蘑菇的相似特征

- 挖掘包含某特定元素的项集
 - 将蘑菇的每一个特征对应一个标称数据值，标称值转化为一个集合
 - 第一个特征值对应有毒没毒
 - 挖掘频繁项集，找出包含 对应有毒特征值 的项集

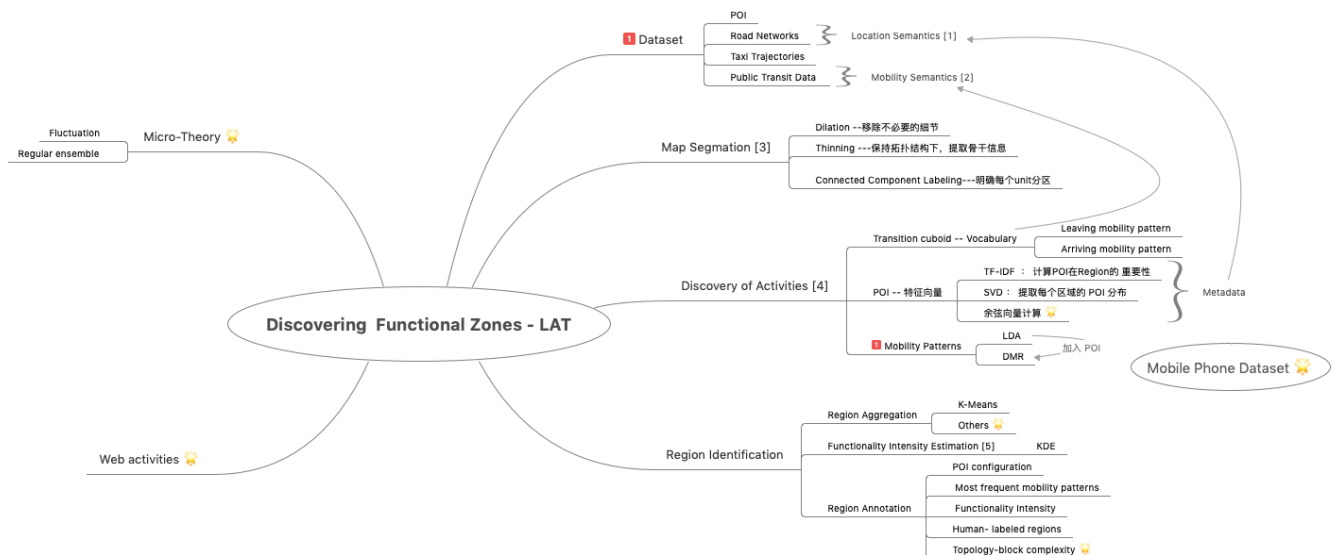
类比于 刘婉甜学姐 介绍的 “交通场景和犯罪发生的对应”

具有相同，相似场景特征的容易催发某类犯罪

- 挖掘包含犯罪特征值的频繁项集
 - 问题1：犯罪行为太多种，对应多个值，而且要表示犯罪数量
 - 问题2：场景除了有无，还有程度的表示
 - 问题3：倾斜支持度分布---支持度阈值选取
 - solve 1：犯罪类型合并归类
 - solve 2：使用一个复合项()来表示某一个特征 [1, 2, 3, 9, [10, 1]]
 - [10,1]-- 1 对应场景包含特征程度，或者犯罪数量分为 少，中，多的数值对应
 - 使用 Apriori Alg 的时候，如果有复合项，先只关注第一项，挖掘
 - 频繁项集挖掘第二项

2 Discovering Urban Functional Zones Using Latent Activity Trajectories

Key : 利用位置序列 反应 经济活动 , 从而确定分区



- 红色---代表 重点需要解决的问题
- □----代表有可能改进的 点
- [大图link](#)

注 :

1. Location Semantics

- 城市的交通道路将城市分割成不同的区
- POI 某些程度上决定一个区的function (大学城--教育区)
- 但是不能区分不同的功能区 (餐厅)

2. Mobility Semantics

- 人类在区域内的模式---人类的时空行为模式

谁，从哪来，到哪去

工作日 早上离开住宅区，晚上返回

- 不同区域之间 由人类的移动连接

工作日从 工作区 来到 娱乐场所
休息日 从住宅区 来到娱乐场所

- 如果人们（在时间限制下）从相似的区域来，到相似的区域区，则两个区域功能相似

3. Map Semantics

- 划分区域，将每一个区域看作承载功能的基本单元
- 选取 raster-based model 便于区域分析

4. Discovery activities

- 基于 概率主题模型
- 将区域活动 类比到文本挖掘
- LPA：将文档集合中每篇文档主题以概率分布的形式给出，结合主题分布，进行主题聚类分析
- 文档包含多个主题，文档的每个词包含于主题

3 主要参考论文

2016年 KDD 论文 --微软亚洲研究院

2018-11-06