



VGI大数据、关键技术 及在城市科学领域的应用

姚尧 博士, 副教授

中国地质大学 (武汉) 地理与信息工程学院

阿里巴巴集团

Email: yaoy@cug.edu.cn

UrbanComp团队主页: <https://www.urbancomp.net>





主要内容



- 1 VGI大数据简介
- 2 VGI大数据关键技术
- 3 VGI在城市科学领域的应用



主要内容

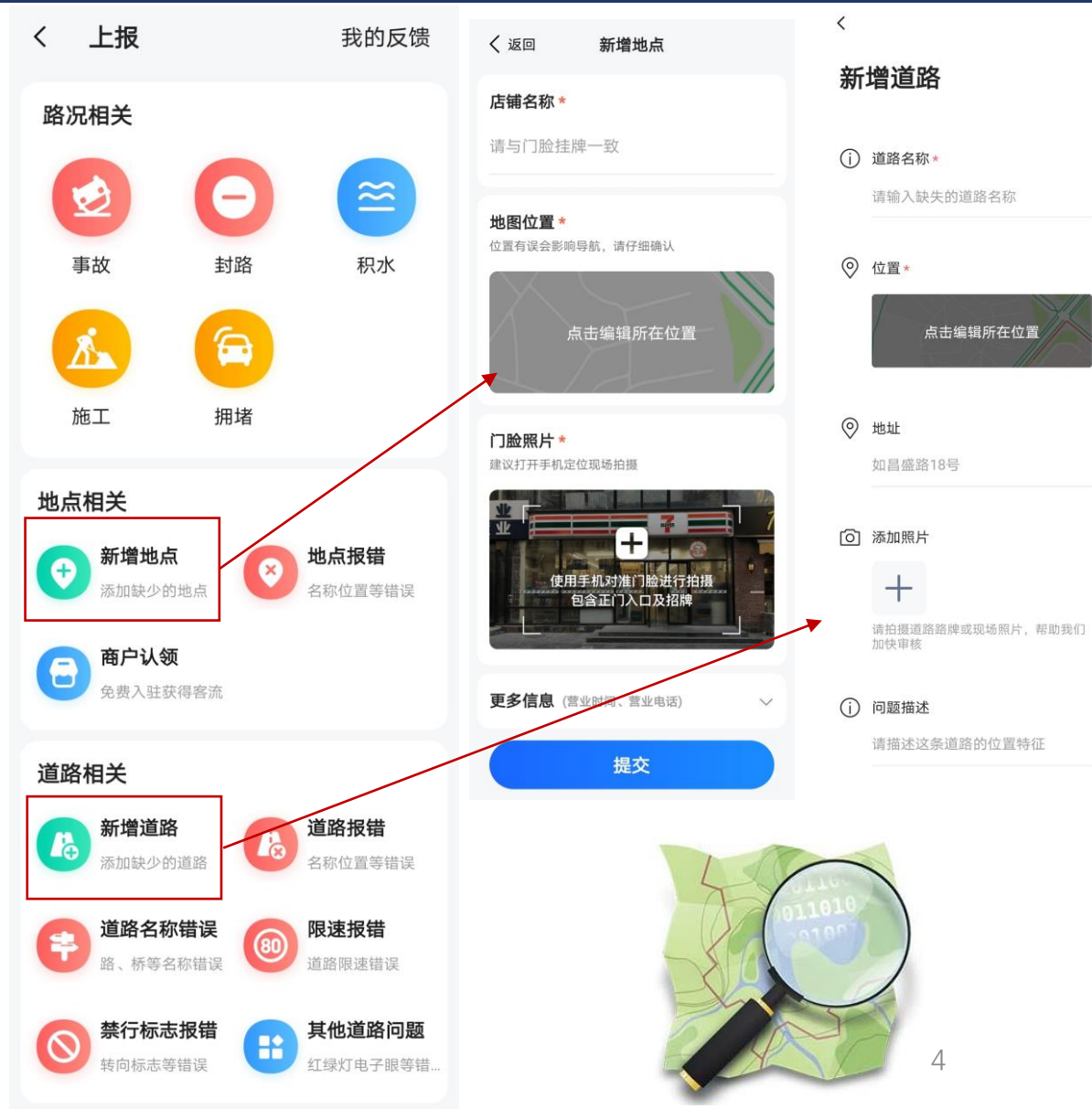


- 1 VGI大数据简介
- 2 VGI大数据关键技术
- 3 VGI在城市科学领域的应用

1.1 | VGI大数据概念



- 自发式地理信息 (Volunteered Geographic Information, 简称VGI):
- 随着地球信息科学在新地理信息时代中发展出现的新概念, 由M.F. Goodchild 于2007年提出;
 - 指任何人都可以通过移动互联设备自发地贡献地理空间信息;
 - 认为地理信息的创建、维护、应用可由大众完成, 反映了互联网时代地理信息新的获取与应用方式;



1.2 | VGI大数据来源



➤ **VGI信息收集平台:** OpenStreetMap、Wikimapia、Google Map Maker、微信、美团...

➤ **OpenStreetMap(OSM):**

- https://wiki.openstreetmap.org/wiki/Main_Page
- 由大众共同打造的免费开发和可编辑的地图服务;
- 随着开源意识的普及, 以及电子地图应用的普及, OSM数据的质量和体量不断增加, 在一些领域的精确度已经不逊于google地图。

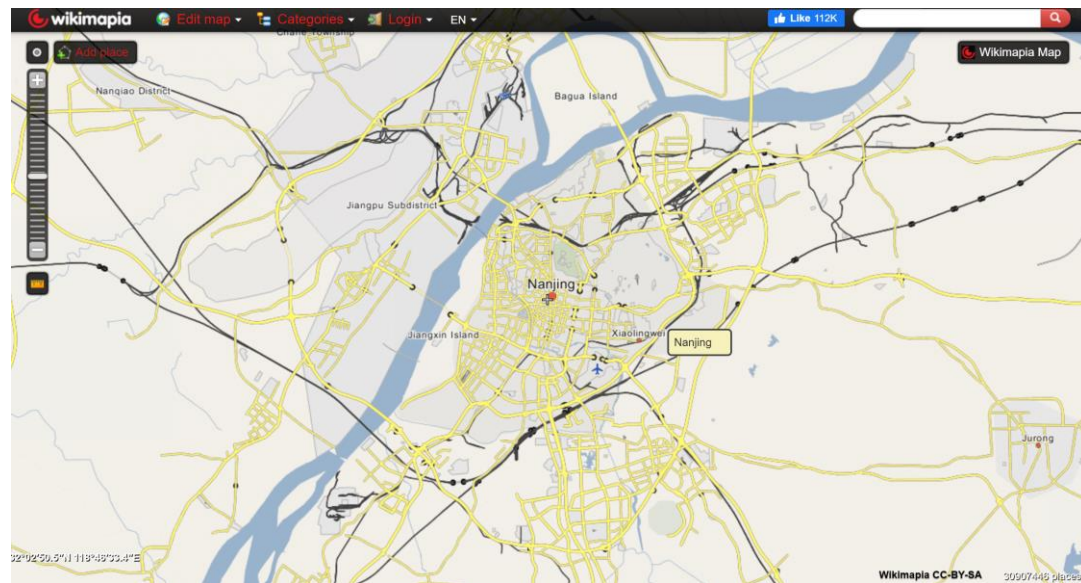
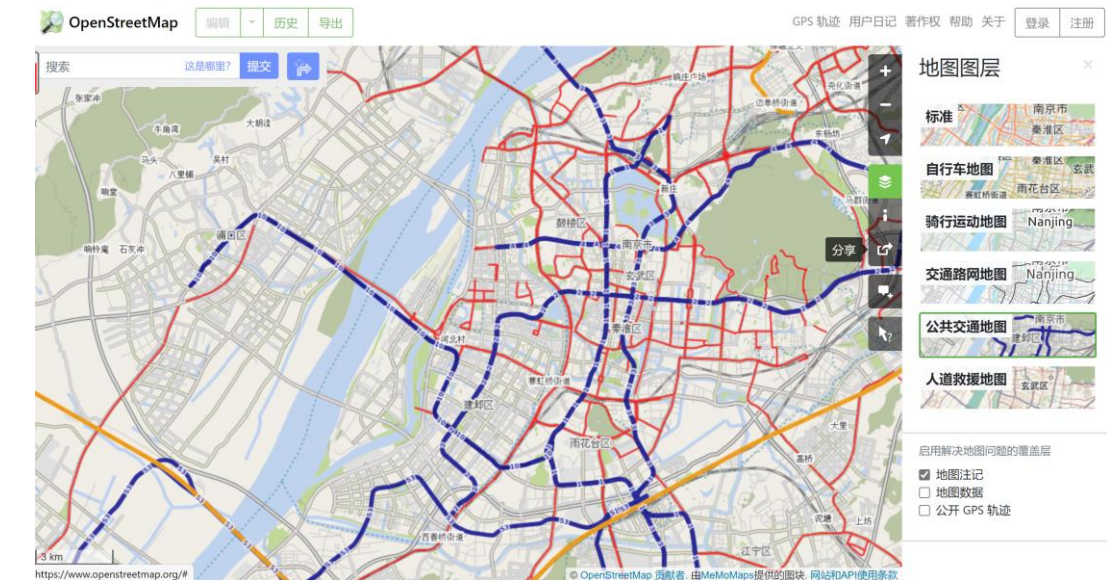
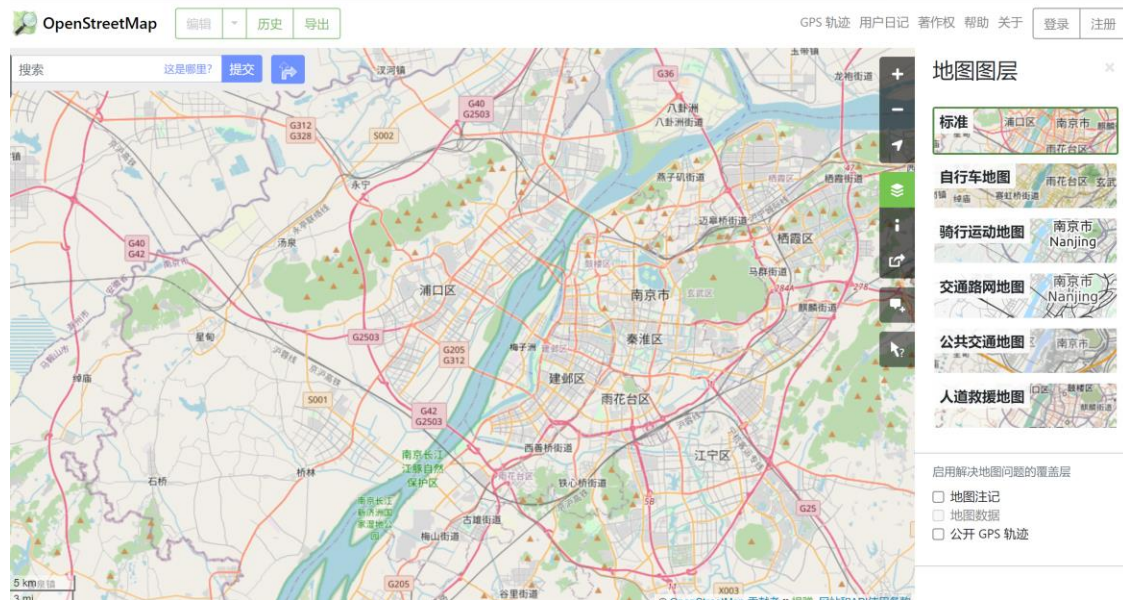
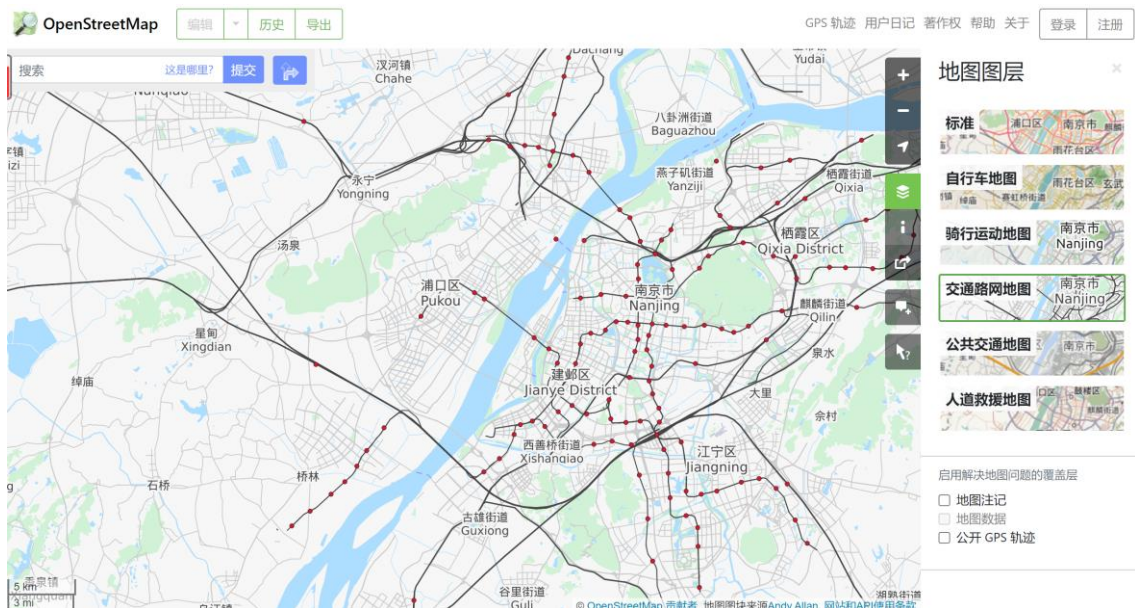
➤ **WikiMapia:**

- <http://wikimapia.org/>
- 其目标是“创建和维护一个自由, 完整, 多语言和最新的全世界地图。将包含有关地球上每一个地方的详细信息”。

➤ **Google Map Maker:**

- Google为了扩展Google地图提供的服务范围而开发的一项地图在线编辑产品 (2017年已关闭)

1.2 VGI大数据来源



1.2 | VGI大数据来源



➤ 带有地理坐标的社交媒体数据也被认为是VGI数据

- 易获取，数据量大，数据源多样，实时更新等。

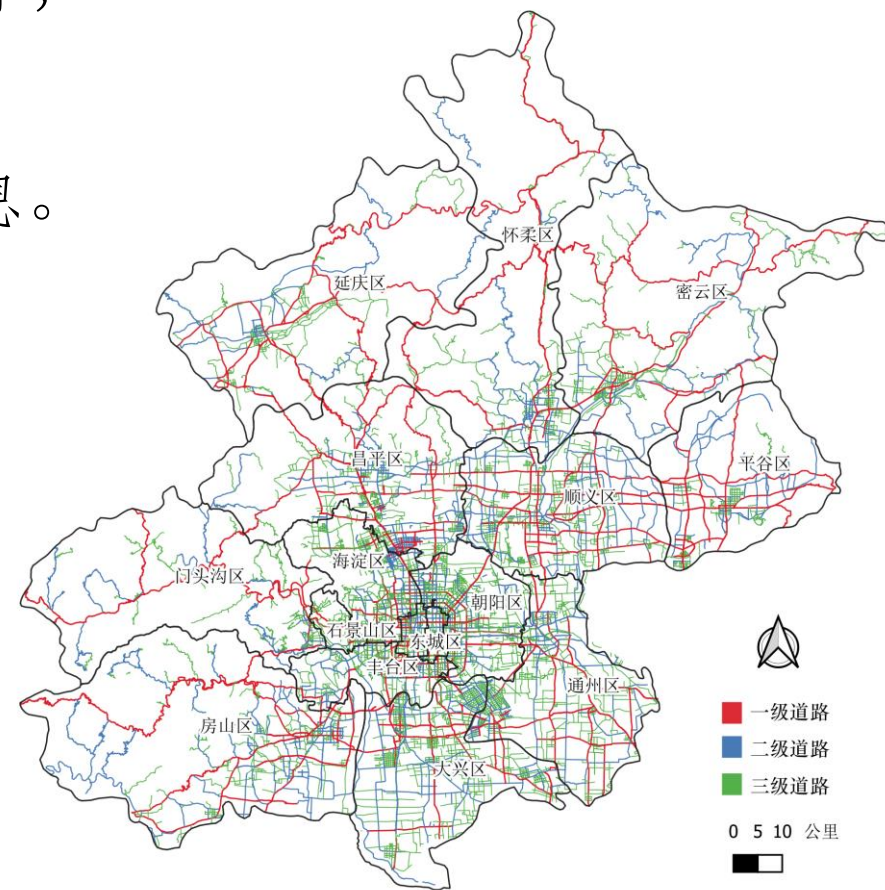
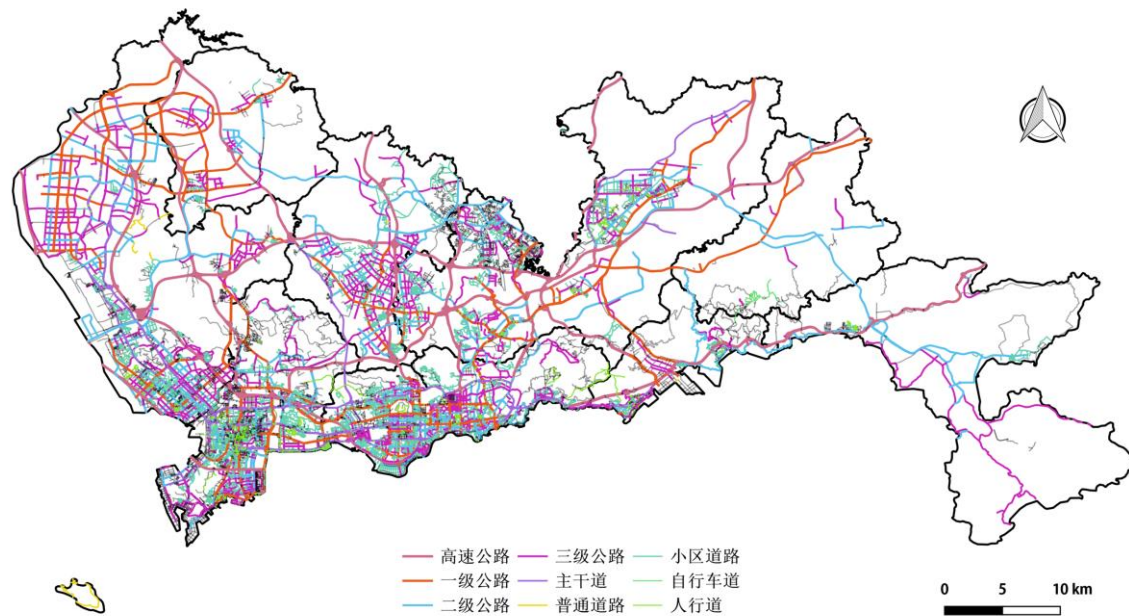


1.3 | VGI大数据示例—OSM路网



➤ OSM路网特点:

- 数据丰富、覆盖度广、免费获取、更新速度快等;
- 数据包含经纬度等基本空间信息以及道路名称、道路类型、最大行驶速度和单向街道等属性信息。



1.3 | VGI大数据示例—OSM路网获取



➤ 1.网站直接下载:

- <http://download.geofabrik.de/>;
- Asia→China→China-latest-free.shp

Click on the region name to see the overview page for that region, or select one of the file extension links for quick access.

Sub Region ①	Quick Links		
	.osm.pbf	.shp.zip	.osm.bz2
Africa	[.osm.pbf] (4.6 GB)	×	[.osm.bz2]
Antarctica	[.osm.pbf] (29.1 MB)	[.shp.zip]	[.osm.bz2]
Asia	[.osm.pbf] (9.9 GB)	×	[.osm.bz2]
Australia and Oceania	[.osm.pbf] (899 MB)	×	[.osm.bz2]
Central America	[.osm.pbf] (453 MB)	×	[.osm.bz2]
Europe	[.osm.pbf] (23.8 GB)	×	[.osm.bz2]
North America	[.osm.pbf] (10.7 GB)	×	[.osm.bz2]
South America	[.osm.pbf] (2.5 GB)	×	[.osm.bz2]

Sub Regions

Click on the region name to see the overview page for that region, or select one of the file extension links for quick access.

Sub Region	Quick Links		
	.osm.pbf	.shp.zip	.osm.bz2
Afghanistan	[.osm.pbf] (76 MB)	[.shp.zip]	[.osm.bz2]
Armenia	[.osm.pbf] (33.6 MB)	[.shp.zip]	[.osm.bz2]
Azerbaijan	[.osm.pbf] (29.7 MB)	[.shp.zip]	[.osm.bz2]
Bangladesh	[.osm.pbf] (249 MB)	[.shp.zip]	[.osm.bz2]
Bhutan	[.osm.pbf] (11.2 MB)	[.shp.zip]	[.osm.bz2]
Cambodia	[.osm.pbf] (26.9 MB)	[.shp.zip]	[.osm.bz2]
China	[.osm.pbf] (769 MB)	[.shp.zip]	[.osm.bz2]
GCC States	[.osm.pbf] (120 MB)	[.shp.zip]	[.osm.bz2]

China

[\[one level up\]](#)

The OpenStreetMap data files provided on this server do **not** contain the user names, user IDs and changeset IDs of the OSM objects because these fields are assumed to contain personal information about the OpenStreetMap contributors and are therefore subject to data protection regulations in the European Union. [Extracts with full metadata](#) are available to OpenStreetMap contributors only.

Commonly Used Formats

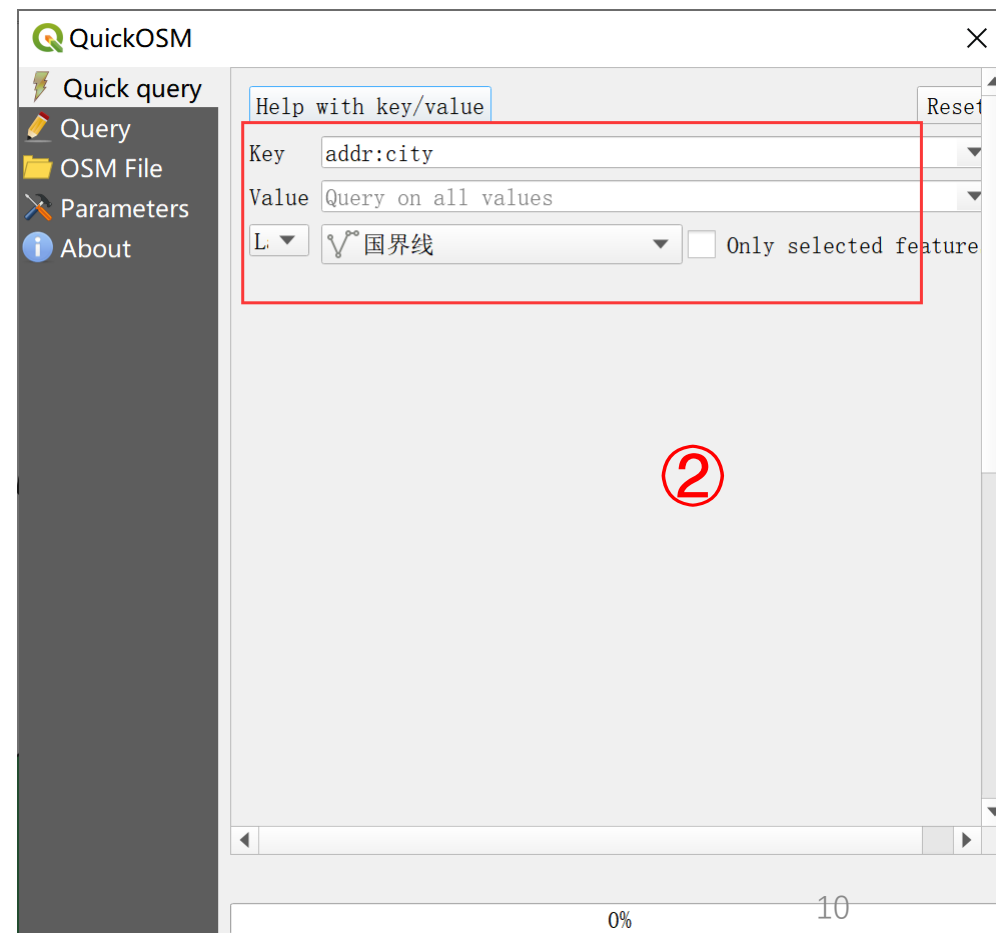
- [china-latest.osm.pbf](#), suitable for Osmium, Osmosis, imposm, osm2pgsql, mkgmap, and others. This file was last modified 12 hours ago and contains all OSM data up to 2021-08-04T20:21:26Z. File size: 769 MB; MD5 sum: [b806ce86b69dde1104a6918180604890](#).
- [china-latest-free.shp.zip](#), yields a number of ESRI compatible shape files when unzipped. ([Format description PDF](#)) This file was last modified 10 hours ago. File size: 1.5 GB; MD5 sum: [d205c599102d4b73fc456a8e22a609a4](#).

1.3 | VGI大数据示例—OSM路网获取



➤ 2.使用QGIS软件获取:

- 安装QuickOSM插件→ 菜单Vector→ QuickOSM →设置;



1.3 | VGI大数据示例—OSM路网处理



➤ OSM路网数据预处理

1

检查数据的一致性，排除相互重叠的道路

2

将双线路转化为单线路，筛选有效道路数据

3

进行拓扑检查，删除悬挂道路和独立路段，保证路网连通性

1.3 | VGI大数据示例—美团店铺及评分



➤ 美团：

- 中国领先的生活服务电子商务平台，拥有美团大众点评、美团外卖等消费者熟知App，服务涵盖餐饮、外卖、生鲜零售、打车、共享单车、酒店旅游、电影、休闲娱乐等200多个品类，业务覆盖全国2800个县区市。

店铺类型

评论日期

用户名称

蓓蓓Eva的家



用户打分

2020年11月6日

评论文本

时间一长就想去吃一下，尤其是冬天，对于不怎么能吃辣的人来说绝对是福音，牛肉太赞了，入口即化，而且牛肉丸很劲道，两个人吃个138的套餐刚刚好，如果胃口好的再加个菜也合适，排队时间有点长，一个小时都属于正常，只是天冷了，户外排队有点辛苦



评论图片



1.3 VGI大数据示例——美团店铺及评分



店铺数据

省	城市	店铺地址	店铺名称	店铺店铺类型代码	类目1	类目2	店铺星级	店铺人均	店铺经度	店铺纬度	活动类型	经纬度省	经纬度市	经纬度	
江苏	B	52, 无锡, 梁溪区崇宁路东方巴黎二楼 (三凤酒家斜对面, 锦江大酒店东行30米)	無界沙龙 (锡城旗舰店),	122	22,	丽人	丽人	5,	177,	120.30172,	31.573751,	个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 梁溪区崇安寺地下商场老街35号 (近崇安寺),	完美指艺美甲美睫韩式定妆,	145	22,	丽人	丽人	3.8,	56,	120.30099,	31.57721,	个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 滨湖区达汇街36号 (万达广场C区门旁),	豪莱坞护肤造型,	248	22,	丽人	丽人	5,	64,	120.2652,	31.55995,	个人护理	江苏省	无锡市	滨湖区
江苏	D	454, 宜兴, 宜兴市东虹东路571号万达星巴克对面,	LLSG美学馆,	657	22,	丽人	丽人	4.2,	148,	119.848702,	31.363169,	个人护理	江苏省	无锡市	宜兴市
江苏	B	52, 无锡, 滨湖区梁溪路万达公寓29号1701室 (星巴克东, 靠近高架桥),	佑米造型 (滨湖万达店),	436	22,	丽人	丽人	4.8,	158,	120.26735,	31.562615,	个人护理	江苏省	无锡市	滨湖区
江苏	D	454, 宜兴, 宜兴市万达百货对面东氪一号东虹路613号售楼外旁 (万达百货一楼)	芭图造型 (万达店),	258	22,	丽人	丽人	5,	85,	119.84988,	31.362839,	个人护理	江苏省	无锡市	宜兴市
江苏	B	52, 无锡, 梁溪区人民中路123号银辉摩天360大厦715室 (B座电梯),	启明星·时尚 (摩天店),	417	22,	丽人	丽人	5,	147,	120.301182,	31.575609,	个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 梁溪区红豆万花城4楼417 (地铁1号线胜利门站4号出口上四楼, 中山路)	GD KING轻奢沙龙 (人气品质),	161	22,	丽人	丽人	5,	238,	120.29811,	31.581315,	个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 梁溪区金太湖广场1071-1072室 (肯德基旁边时尚生活馆入口第一家)	杨静美甲美睫韩式半永久定妆,	250	22,	丽人	丽人	4,	285,	120.287505,	31.593082,	个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 梁溪区人民中路123号银辉摩天360大厦B座电梯720室 (崇安寺),	名品美发 (摩天360店),	686	22,	丽人	丽人	5,	203,	120.301623,	31.575911,	个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 滨湖区吴都路融绿星光广场二楼南侧 (天鹅湖花园B区172-2009)	丝艺造型 (星光广场店),	494	22,	丽人	丽人	4,	29,	120.279887,	31.483441,	个人护理	江苏省	无锡市	滨湖区
江苏	B	52, 无锡, 新吴区旺庄路188号宝龙城市广场3楼3031-3033蜜桃餐厅斜对面 (家乐福)	GD KING轻奢品质潮牌沙龙 (新吴区),	918	22,	丽人	丽人	5,	327,	120.350966,	31.539771,	个人护理	江苏省	无锡市	新吴区
江苏	B	52, 无锡, 滨湖区金石路88号万象城润街110-101 (近大剧院路),	塞巴斯汀SEBASTIAN专业美发,	402	22,	丽人	丽人	4.6,	104,	120.284387,	31.515965,	个人护理	江苏省	无锡市	滨湖区
江苏	B	52, 无锡, 梁溪区中山路T12时尚购物中心五楼,	I-SPACE艾贝斯美发沙龙,	512	22,	丽人	丽人	4.5,	152,	120.29969,	31.57948,	个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 梁溪区人民中路127号三阳百盛五楼,	廊艺造型 (百盛店),	150	22,	丽人	丽人	4.5,	232,	120.301831,	31.5758,	个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 梁溪区南市桥巷22号 (南禅寺地铁站3号出口对面浦发银行巷里50米)	名品美发 (南禅寺店),	407	22,	丽人	丽人	5,	200,	120.304573,	31.567621,	个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 梁溪区中山路270号天安大厦12楼1206室电梯右转 (近云蝠大厦, 崇安寺)	甲天下轻奢美甲沙龙 (崇安寺店),	503	22,	丽人	丽人	4.3,	152,	120.30002,	31.57701,	个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 梁溪区广益路29号 (近广益路地铁站3号出口对面浦发银行巷里50米)	名品美发 (广益路店),	512	22,	丽人	丽人	5,	200,	120.304573,	31.567621,	个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 梁溪区清扬路12号	店铺ID,	评论ID,	用户名称,	用户ID,	评论内容,					个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 滨湖区观山路玉	3276974,	919213808	null,	64303189,	298.288.299服务好的很。技术好的很。,					个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 滨湖区梁溪路29	3276974,	919348536	null,	237608384	推荐313采耳很好, 手法很好。,					个人护理	江苏省	无锡市	滨湖区
江苏	B	52, 无锡, 梁溪区崇安寺步	3276974,	919254561	匿名用户,	0,	环境不错, 和同事们第二次来的, 技师们也很不错, C009.C023、C003都不错,					个人护理	江苏省	无锡市	梁溪区
江苏	B	52, 无锡, 梁溪区人民东路	3276974,	919335970	匿名用户,	0,	按摩师技术还不错☺️服务也相当不错。。。,					个人护理	江苏省	无锡市	梁溪区
江苏	D	454, 宜兴, 宜兴市太滂西路	3276974,	919014817	null,	417509549	298手法很好, 环境不错, 服务很好, 很满意,					个人护理	江苏省	无锡市	宜兴市
江苏	B	52, 无锡, 滨湖区蠡溪路建	3276974,	918569972	null,	144691710	比较不错, 环境, 服务都好, 推荐给大家,					个人护理	江苏省	无锡市	滨湖区
			3276974,	918896280	null,	36916117,	288跟299按得很舒服, 按完全身都松了, 强烈推荐,					个人护理	江苏省	无锡市	滨湖区

评论数据

➤ VGI数据在城市科学中有什么优势？

- 缩短了地理信息创建和传播的时间，降低了普通大众获取、分享和处理地理信息的门槛，**数据量大，易获取**；
- 补充地理框架数据的不足，为更好的描述现实世界提供丰富的**细节和人类感知**；
- 具有随时编辑性，能尽可能的保证**现势性**。



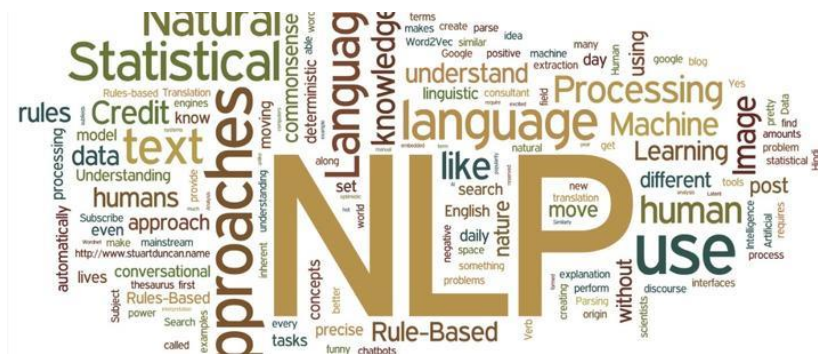
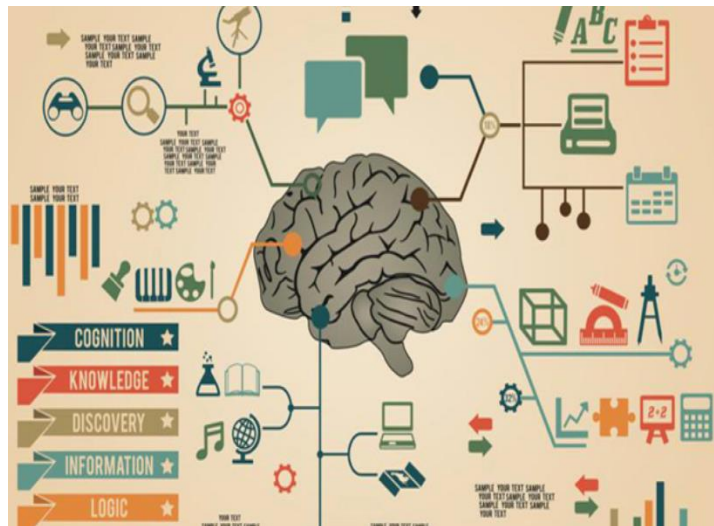
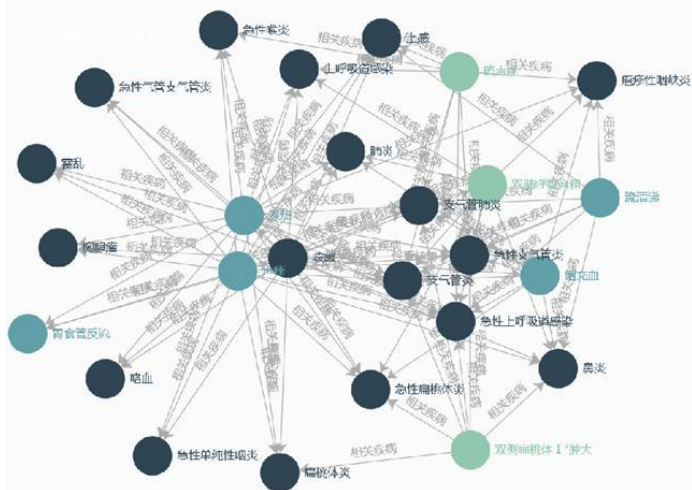
主要内容



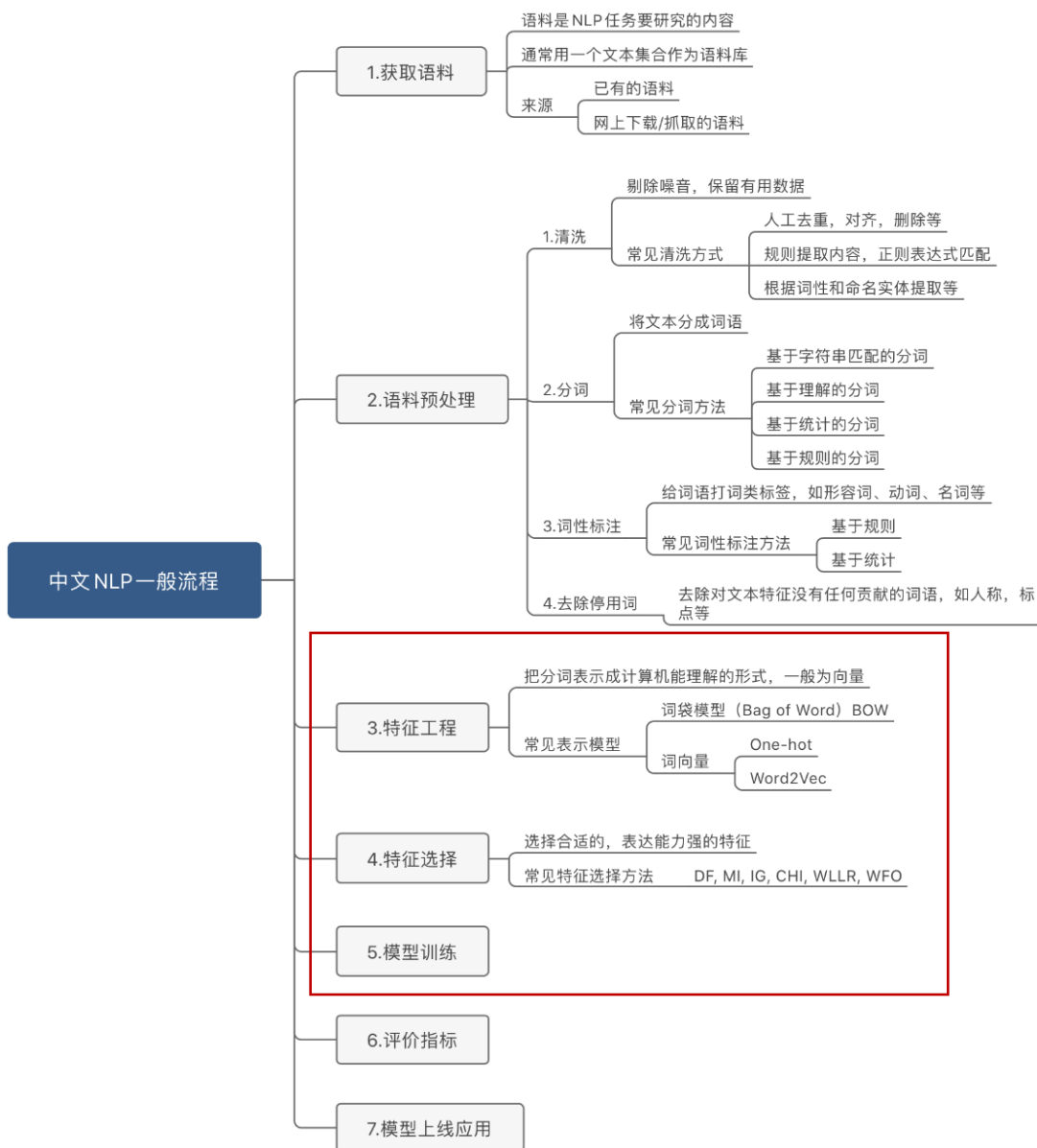
- 1 VGI大数据简介
- 2 VGI大数据关键技术
- 3 VGI在城市科学领域的应用

➤ 自然语言处理(Natural Language Processing, NLP):

- 是计算机科学领域与人工智能领域中的一个重要方向；
- 研究能实现人与计算机之间用自然语言进行有效通信的各种理论和方法；
- 应用于机器翻译、舆情监测、自动摘要、观点提取、文本分类、问题回答、文本语义对比、语音识别、中文OCR等方面。



2.1 VGI大数据关键技术——自然语言处理



词语

↓ 组成

句子

↓ 组成

段落/
篇章/
文档

↑ 包含

主题

如何把这些抽象的形式, 转换为数值 (即嵌入到一个数学空间中)?

如何在大量的文档中自动发现隐含的主题结构信息?

2.1 | VGI大数据关键技术—从简单的开始



1	明天	我们	去	看	电影
2	明天	我们	去	爬山	
3	明天	我们	去	运动	
4	我	喜欢	打	篮球	
5	我	不	喜欢	看	电视

语料



```
[1. 1. 1. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0.]
[1. 1. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]
[1. 1. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.]
[0. 0. 0. 0. 1. 1. 0. 0. 0. 1. 1. 0. 0.]
[0. 0. 0. 1. 1. 1. 0. 0. 0. 0. 0. 1. 1.]
```

One-hot 编码



如果语料库过大会产生什么后果?

TF-IDF模型

"词频" (TF) 和"逆文档频率" (IDF)

词频(TF) = 某个词在文章中的出现次数

词频(TF) = $\frac{\text{某个词在文章中的出现次数}}{\text{文章的总词数}}$

词频(TF) = $\frac{\text{某个词在文章中的出现次数}}{\text{该文出现次数最多的词的出现次数}}$

(语料长短不一)

(语料库过大)



逆文档频率(IDF) = $\log\left(\frac{\text{语料库的文档总数}}{\text{包含该词的文档数} + 1}\right)$



TF - IDF = 词频(TF) × 逆文档频率 (IDF)

➤ 主题模型

- 主题模型是用来在大量文档中发现潜在主题的一种统计模型。
- 主题模型能够自动分析文档 (document) ，不计顺序地统计文档内的单词 (word) ，根据统计的信息判断该文档包含的主题 (topic) 以及各个主题所占比例。

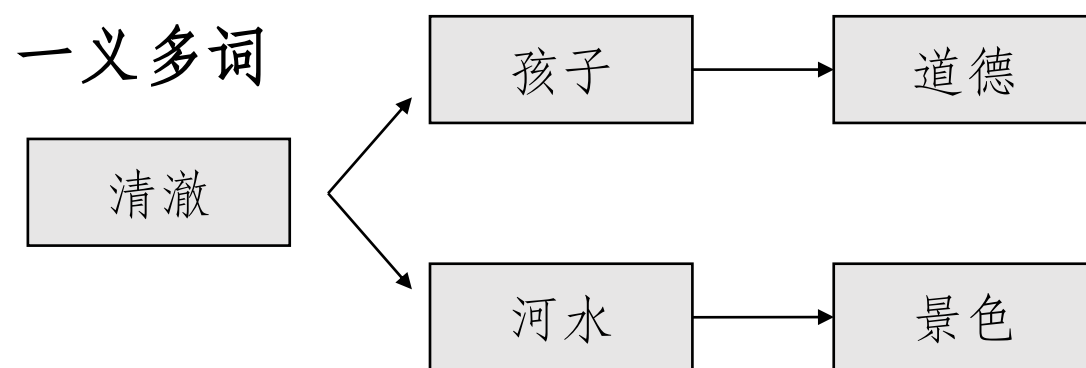
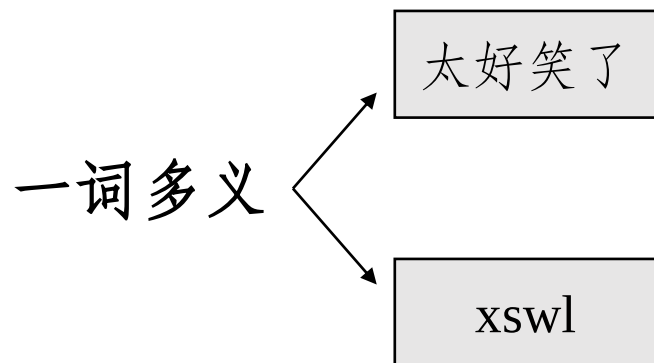
主题模型分类	{	潜在语义分析(LSA)模型 (Christos H et al.1998)
		概率潜在语义分析(PLSA)模型 (Thomas Hofmann H et al.1999)
		潜在狄利克雷分配(LDA)模型 (David M.Blei et al.2003)

2.2 | 自然语言处理—主题模型



➤ 潜在语义分析(LSA)模型

- 一个单词的属性是由它所处的环境刻画的；
- 解决“一词多义”和“一义多词”（术语：指代消解）；



➤ 潜在语义分析(LSA)模型

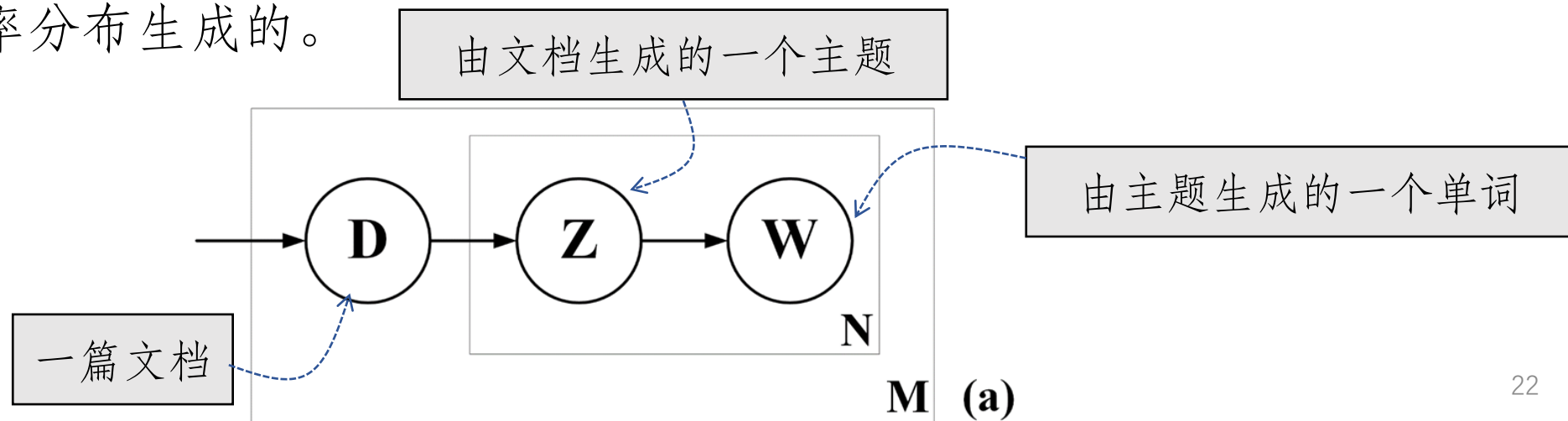
- 核心思想：把拥有的文档-术语矩阵分解成相互独立的文档-主题矩阵和主题-术语矩阵。通过将高维向量映射到潜在语义空间，提取与文档和词项有关的concepts，从而分析文档和词项之间的关系；
- 基于奇异值分解（SVD）的方法得到文档的主题，解决了传统向量空间模型无法处理一词多义或多次同义的问题，LSA方法快速且高效，但它也有一些主要缺点：
 - ✓ 缺乏可解释的嵌入（我们并不知道主题是什么，其成分可能积极或消极，这一点是随机的）
 - ✓ 需要大量的文件和词汇来获得准确的结果
 - ✓ 表征效率低

2.2 | 自然语言处理—主题模型



➤ 概率潜在语义分析(pLSA)模型

- 通过一个生成模型来为LSA赋予了概率意义上的解释；
- 假设每一篇文档都包含一系列可能的潜在主题，文档中的每一个单词都不是凭空产生的，而是在这些潜在的主题的指引下通过一定的概率生成的；
- 主题看作是一种单词上的概率分布，每一个主题都代表着一个不同的单词上的概率分布，而每个文档又可以看成是主题上的概率分布。每篇文档就是通过这样一个两层的概率分布生成的。



➤ 概率潜在语义分析(pLSA)模型

- 基于派生自LCM的混合矩阵分解，属于概率图模型中的生成模型，该类模型还有一元模型、混合一元模型等；
- 其核心思想是找到一个潜在主题的概率模型，该模型可以生成在文档-术语矩阵中观察到的数据。是基于双模式和共现的数据分析方法延伸的经典统计学方法。
- 缺点：
 - ✓ 尽管 PLSA 模型在给定的文档上是一个生成模型，它却无法生成新的未知的文档；
 - ✓ 随着文档数量的增加， $P(Z|D)$ 的参数也会随着线性增加，这就导致无论有多少训练数据，都容易导致模型的过拟合问题。

➤ 潜在狄利克雷分配(LDA)模型

- 主题模型研究领域内应用最为广泛；
- 是在 PLSA 的贝叶斯版本，在 PLSA 模型的基础上引入了参数的先验分布这个概念；
- 它使用狄利克雷先验来处理文档-主题和单词-主题分布，从而有助于更好地泛化。
 - ✓ 每个文档关于主题的概率分布都被赋予了一个先验分布；
 - ✓ 一个主题在所有单词上的概率分布也赋予了先验分布。



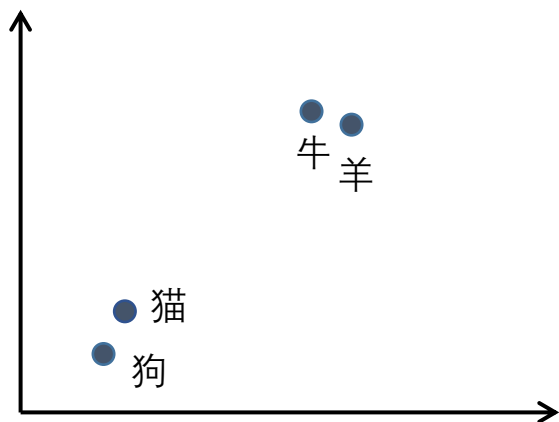
LDA 模型能够比 PLSA 更好地刻画文档-主题-单词这三者的关系

2.3 | 自然语言处理—Word2Vec



➤ Word2Vec

- 将词转化为「可计算」「结构化」的向量的过程



若两个词语词义相近，则二者在向量空间上也相近

如何得到
word2Vec
呢？

对词向量进行计算时，有一个有趣的现象如下所示：

$$\text{vector}(\text{'queen'}) \approx \text{vector}(\text{'king'}) - \text{vector}(\text{'man'}) + \text{vector}(\text{'woman'})$$

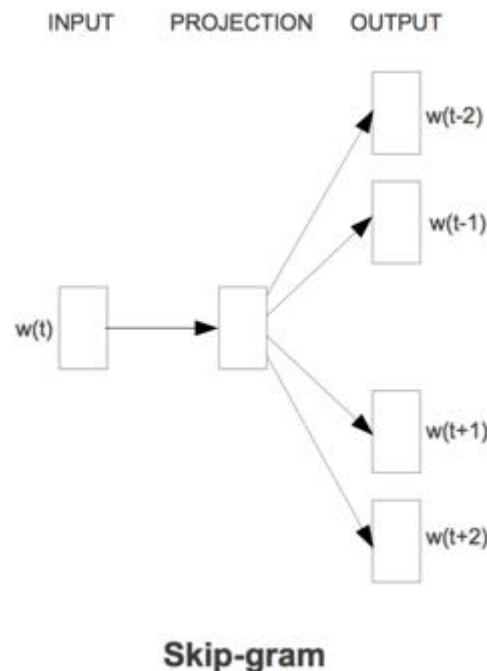
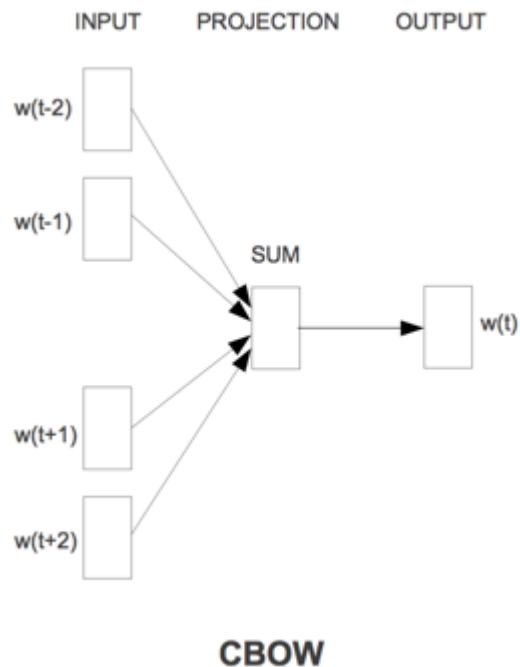
(女王 $\approx$ 国王 - 男人 + 女人)

2.3 | 自然语言处理—Word2Vec



➤ Word2Vec构建:

- 常采用两种模型构建：CBOW 模型& Skip-gram模型；
 - ✓ CBOW模型根据中心词 $w(t)$ 周围的词来预测中心词
 - ✓ Skip-gram模型则根据中心词 $w(t)$ 来预测周围词

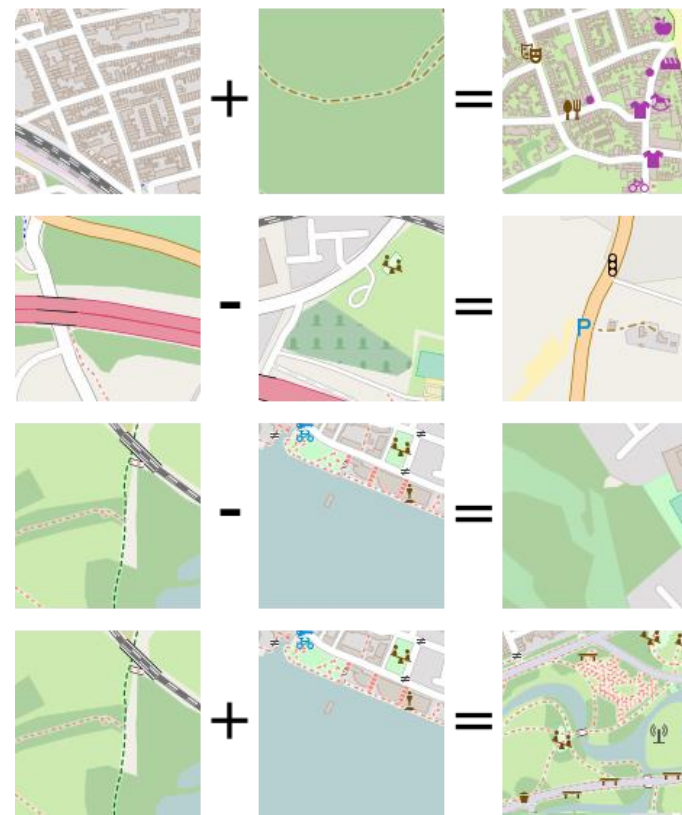


2.3 | 自然语言处理—Word2Vec



➤ Word2Vec衍生：

- Word2vec思想在地理信息研究中得到充分利用，如：POI2vec、Traj2Vec等

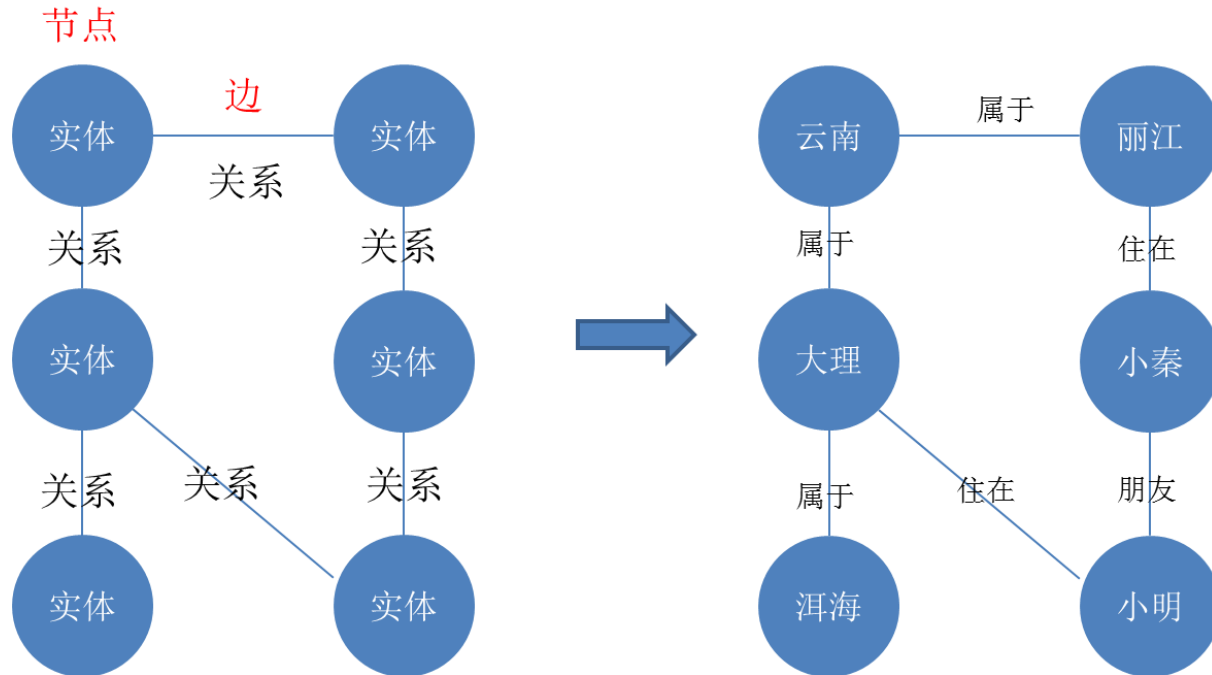


2.4 | 自然语言处理—知识图谱



➤ 知识图谱：

- 一种基于图的数据结构，由节点（Point）和边（Edge）组成；
- 每个节点表示一个“实体”，每条边为实体与实体之间的“关系”；
- 知识图谱本质上是语义网络。



2.4 | 自然语言处理—知识图谱



➤ 知识图谱

多关系数据(multi-relational data)

节点：实体/概念

边：关系/属性

关系事实 = (head, relation, tail)

head：头部实体

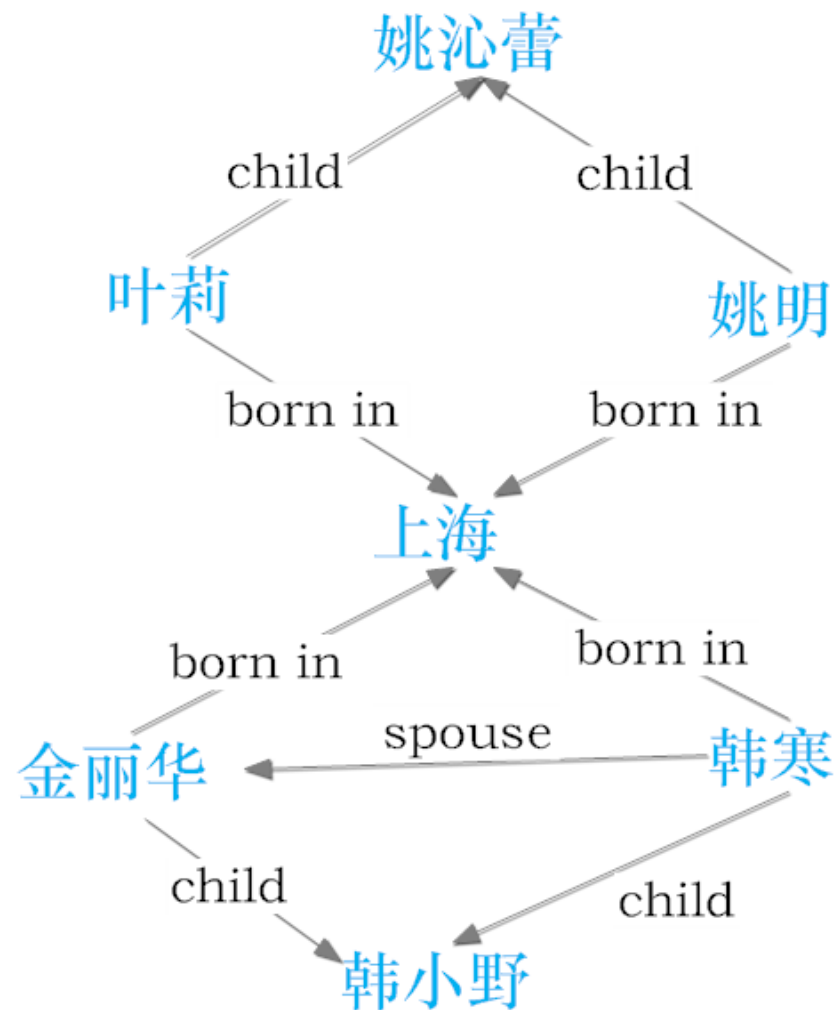
relation：关系/属性

tail：尾部实体

(姚明, born in, 上海)



Head relation tail



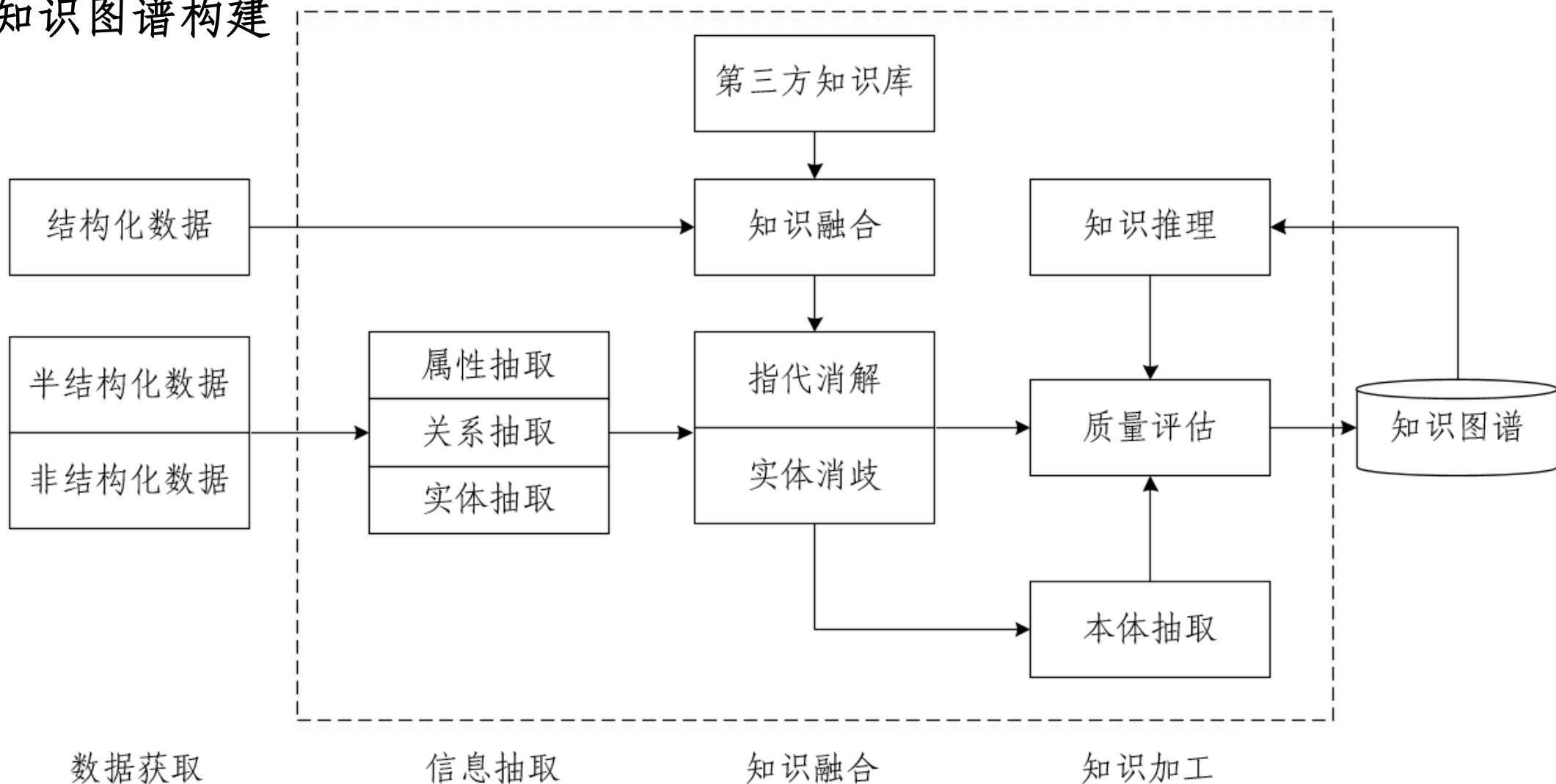
➤ 知识图谱中包含的主要几种节点

- **实体**：指的是具有可区别性且独立存在的某种事物。如某一个人、某一座城市、某一种植物、某一件商品等等。实体是知识图谱中的最基本元素，不同的实体间存在不同的关系。
- **概念**：具有同种特性的实体构成的集合，如国家、民族、书籍、电脑等。
- **属性**：用于区分概念的特征，不同概念具有不同的属性。不同的属性值类型对应于不同类型属性的边。如果属性值对应的是概念或实体，则属性描述两个实体之间的关系，称为对象属性；如果属性值是具体的数值，则称为数据属性。

2.4 | 自然语言处理—知识图谱



➤ 知识图谱构建





主要内容



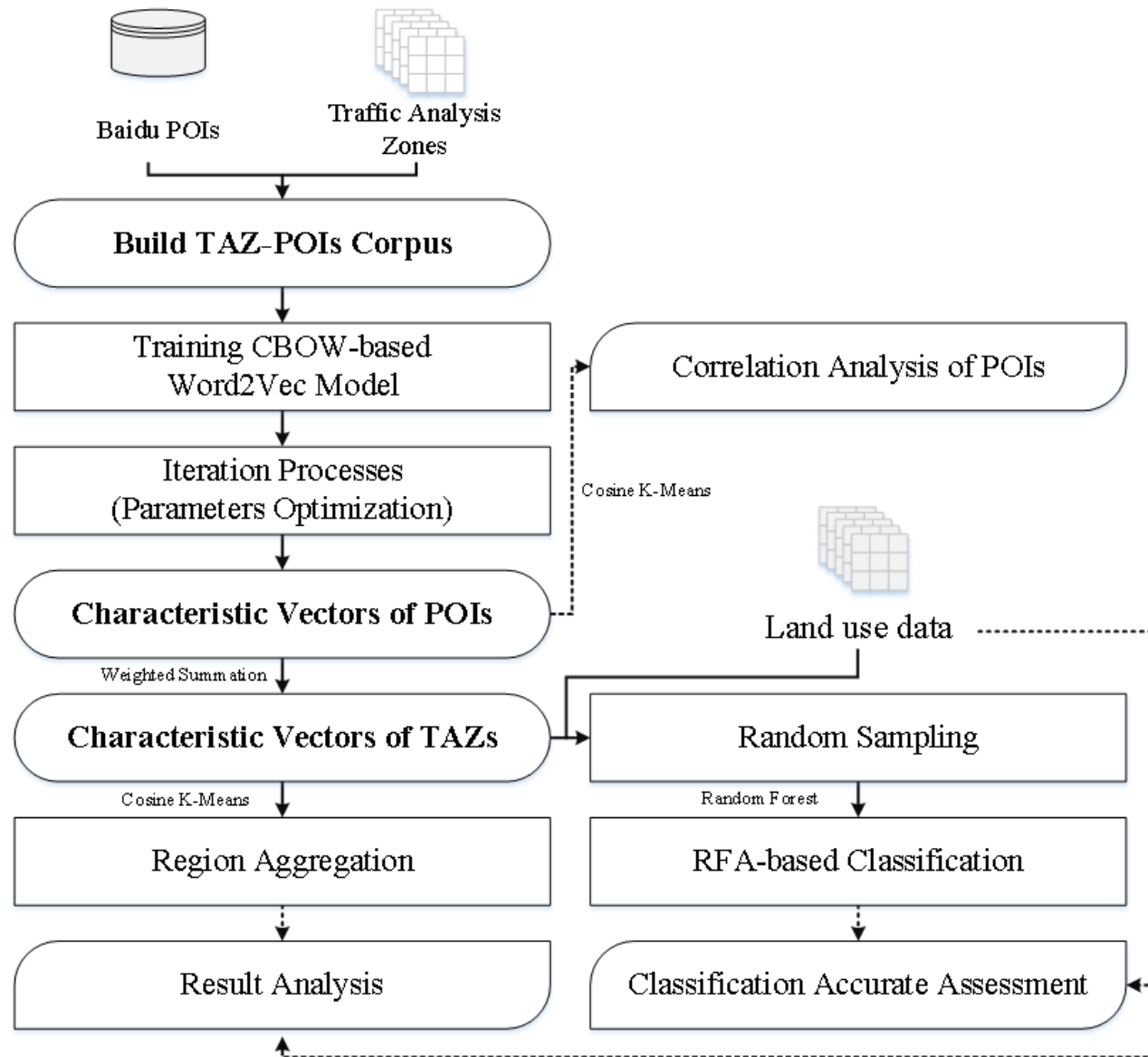
- 1 VGI大数据简介
- 2 VGI大数据关键技术
- 3 VGI在城市科学领域的应用

3.1 | POI向量化—POI2Vec



如何将POI转换为机器可以计算的向量形式，便于对其进行研究分析？

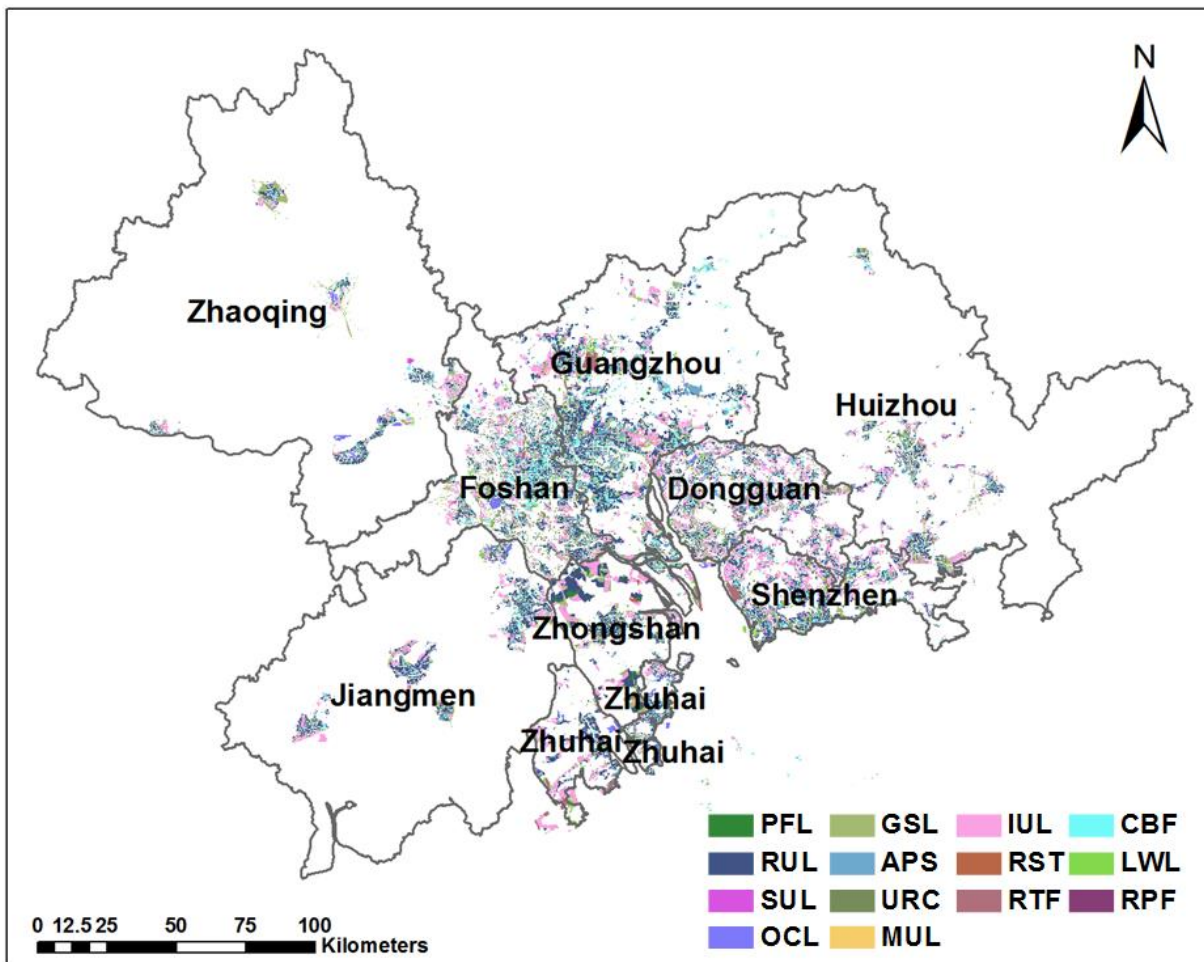
POI2Vec



3.1 | POI向量化—POI2Vec



➤ 研究区域



ID	Short Code	Land-use type	Proportions
1	PFL	Public facilities	3.97%
2	GSL	Green space and square	37.31%
3	IUL	Industrial land	8.19%
4	CBF	Commercial and business facilities	13.23%
5	RUL	Residential land	20.34%
6	APS	Administration and public services	7.52%
7	RST	Road, street and transportation	1.68%
8	LWL	Logistics and warehouse	1.03%
9	SUL	Special use land	0.40%
10	URC	Urban and rural construction land	5.29%
11	RTF	Regional traffic facilities	0.59%
12	RPF	Regional public facilities	0.02%
13	OCL	Other construction land	0.38%
14	MUL	Mining use land	0.05%

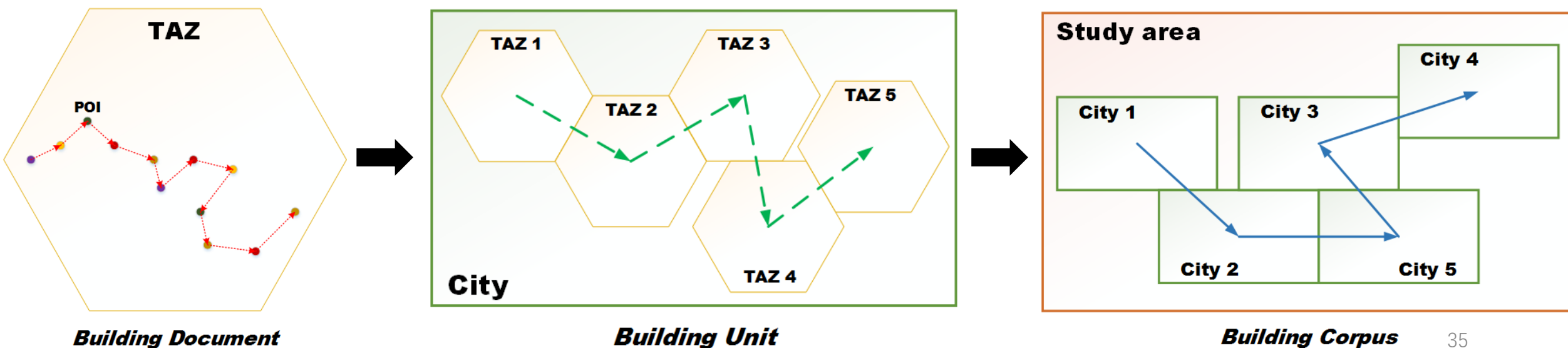
Baidu POIs: **1,403,453 records** in study area

Multi-level categories: **Level1 (20 types) – Level 4 (419 types)**

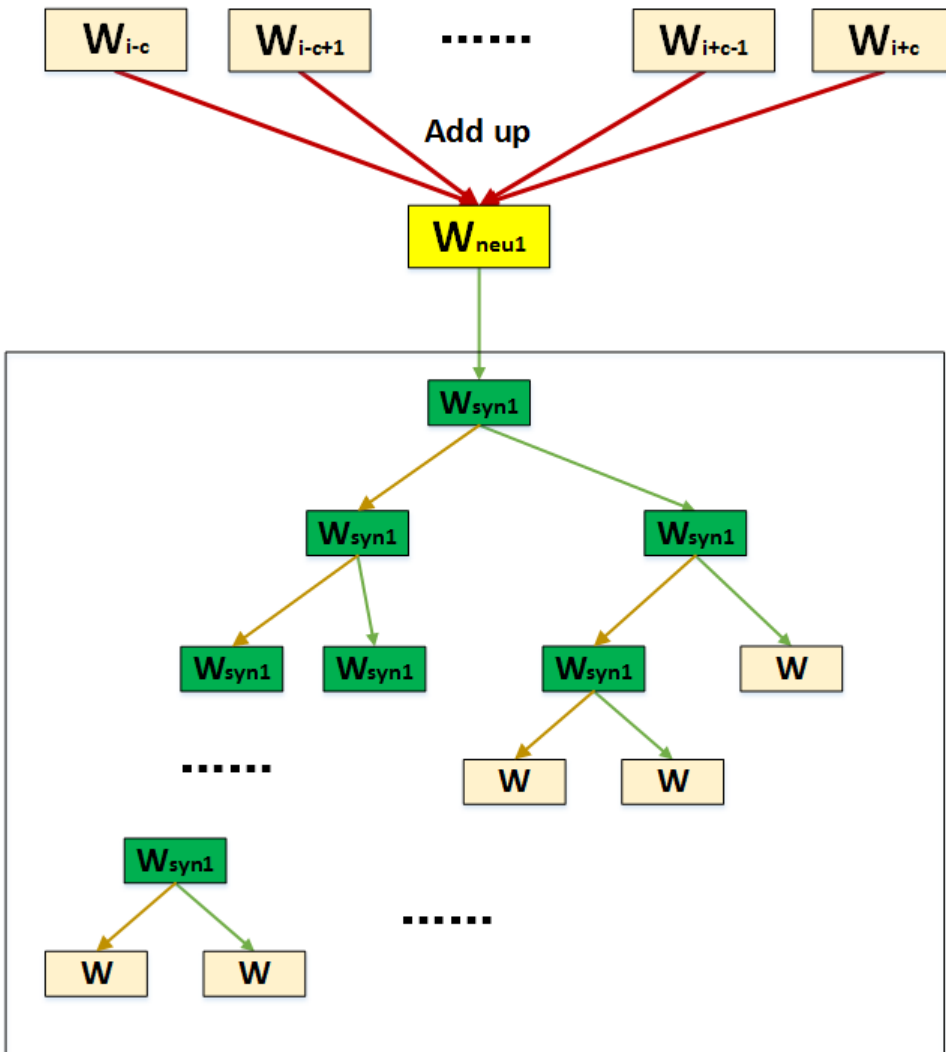
3.1 | POI向量化—POI2Vec



- **Corpus** ↔ Study area
- **Units** ↔ Cities
- **Documents** ↔ TAZs
- **Words** ↔ Categories of POIs (at final-level)
- **Method: *Greedy Algorithm*** (Minimum Distance Optimizing)



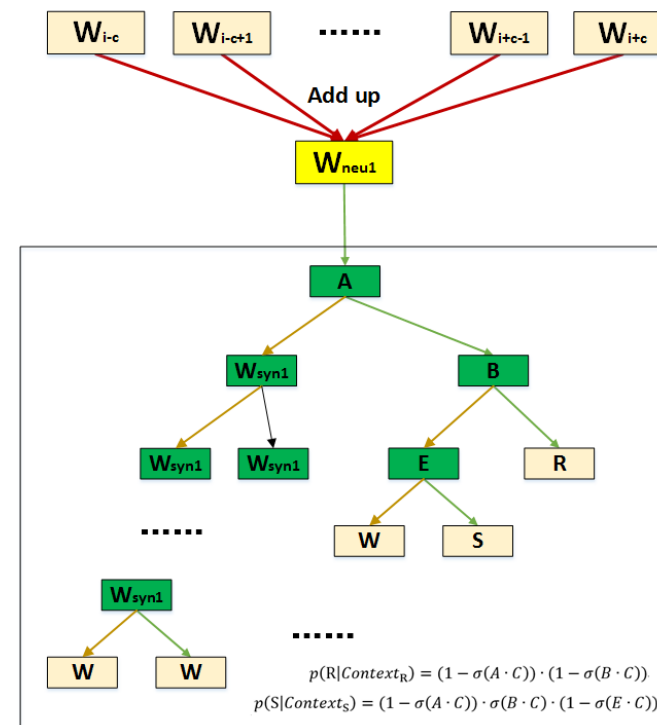
3.1 POI向量化—POI2Vec



Hoffman Trees



* Number of POIs



Maximum likelihood function

$$l(\theta) = \log L(\theta) = \frac{1}{T} \sum_{t=1}^T \log p(w_t | w_{t-c}^{t+c})$$

$$p(w_t | w_{t-c}^{t+c}) = \frac{\exp(-E(w_t, w_{t-c}^{t+c}))}{\sum \exp(-E(w_i, w_{t-c}^{t+c}))}$$

37

3.1 POI向量化—POI2Vec



Enter word or sentence
Word: 加油站
Position in vocabulary: 10

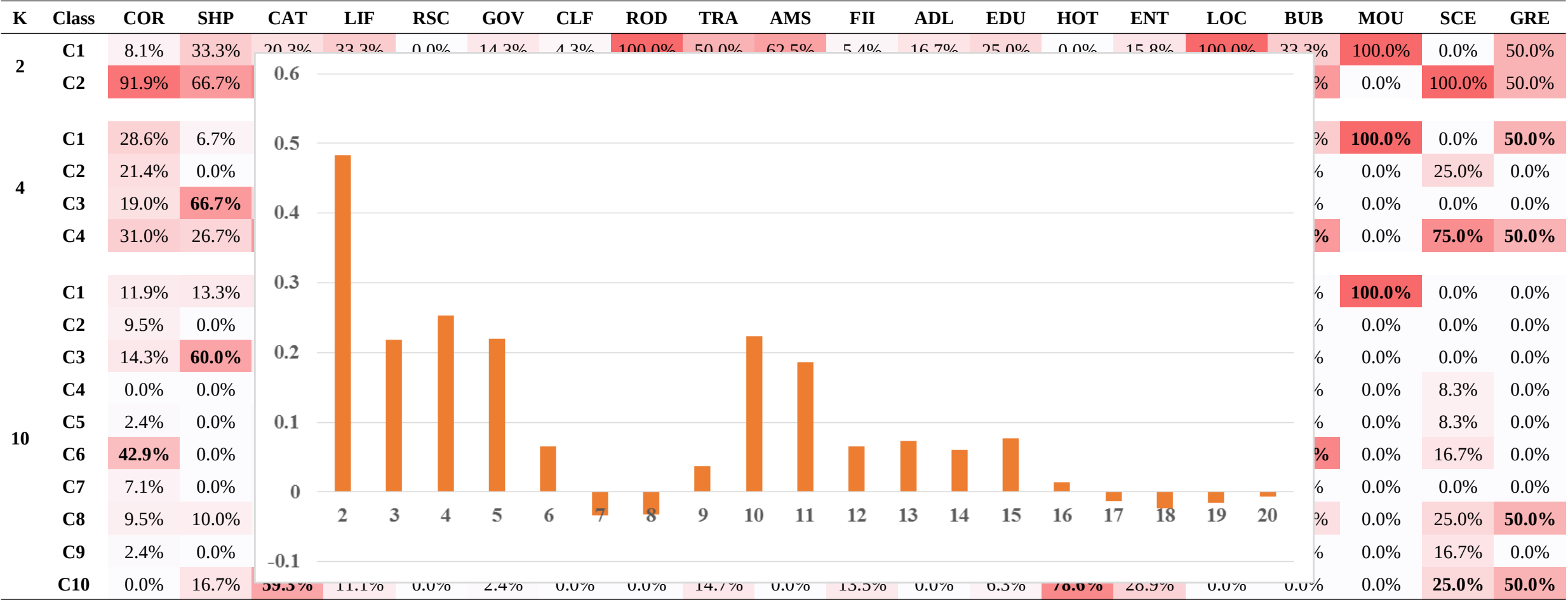
Enter word or sentence
Word: 局
Position in vocabulary: 40

Enter word or sentence
Word: 托儿所
Position in vocabulary: 41

C:\Windows\system32\cmd.exe
Enter word or sentence (EXIT to break): 公安机关
Word: 公安机关 Position in vocabulary: 41

Word	Cosine distance
公检法机构	0.842399
消防局	0.786719
局	0.765451
机关单位	0.713928
居民委员会	0.686625
基金会	0.613507
希望工程	0.586934
司法局	0.573946
交管局	0.556534
社会团体	0.552677
福利机构	0.527460
敬老院	0.526809
村民委员会	0.502741
托儿所	0.492371
街道办事处	0.472069
交通设施	0.469930
活动中心	0.469272
检察院	0.468824
地市级政府	0.460784
药房	0.441578
道路	0.432058
局	0.421293
干洗	0.418817
康复中心	0.410329

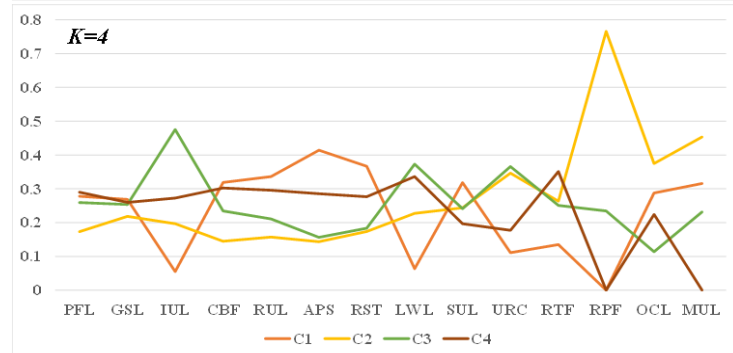
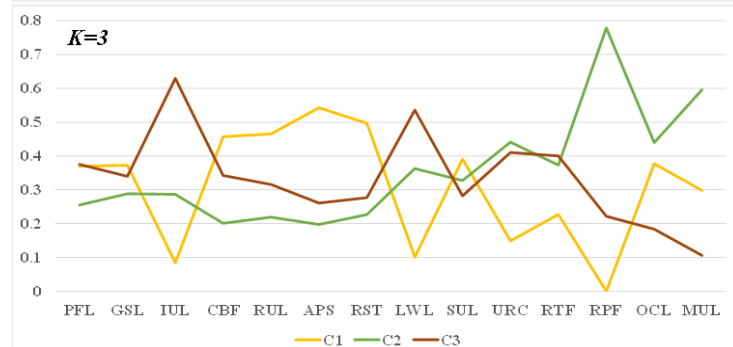
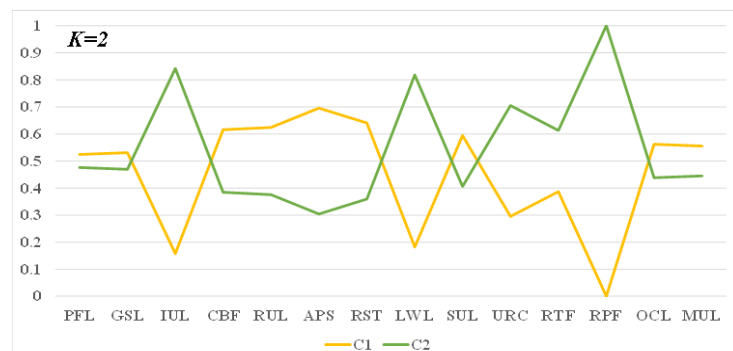
3.1 | POI向量化—POI2Vec



Types of POIs: **COR** = Corporation, **SHP** = Shopping, **CAT** = Catering, **LIF** = Life Service, **RSC** = Residential community, **GOV** = Government, **CLF** = Clinic facility, **ROD** = Road, **TRA** = Traffic facility, **AMS** = Automobile service, **FII** = Financial industry, **ADL** = Administrative landmark, **EDU** = Education, **HOT** = Hotel, **ENT** = Entertainment, **LOC** = Location annotation, **BUB** = Business building, **MOU** = Natural Mountain, **SCE** = Scenic spot, **GRE** = Green space

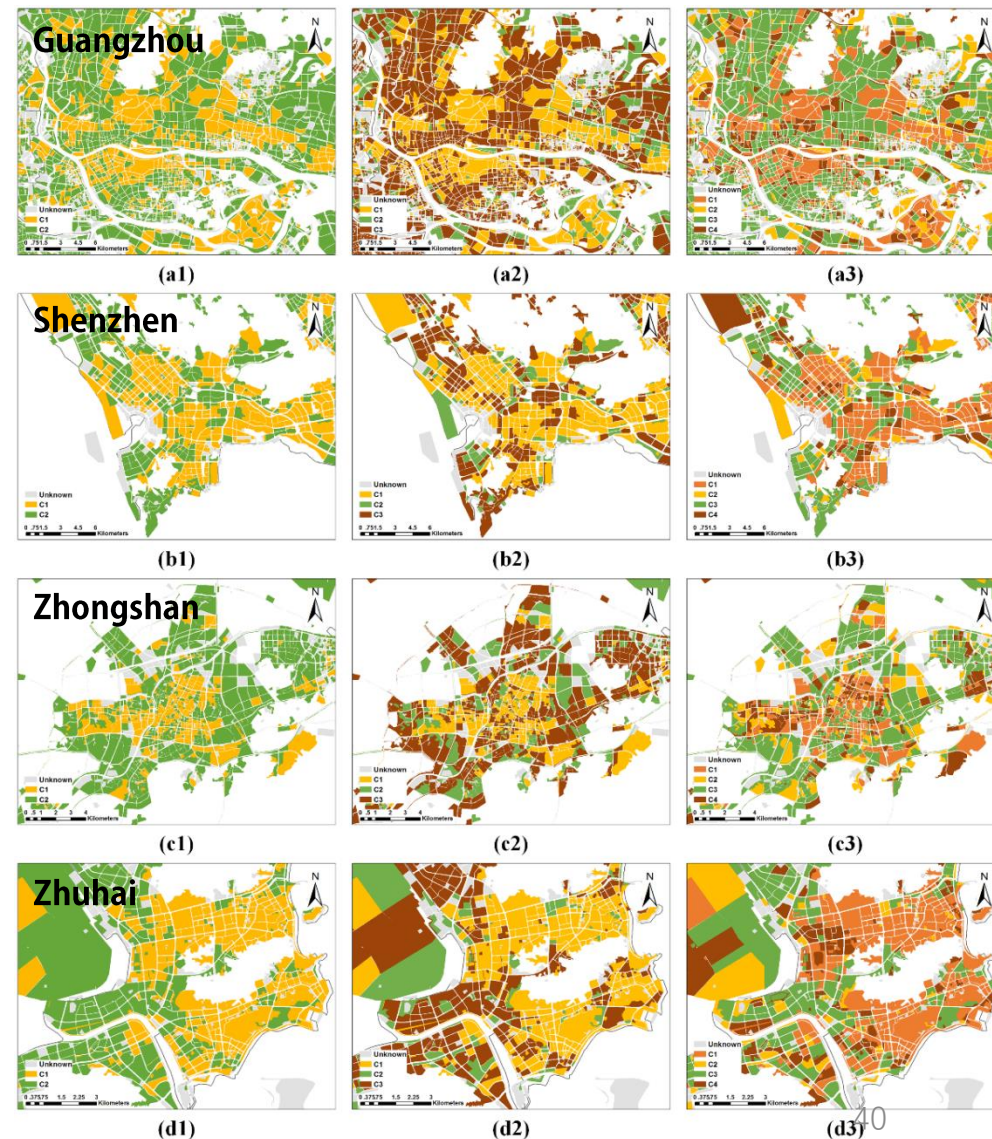
39

3.1 | POI向量化—POI2Vec



Land-use types:

Public facilities (PFL),
Green space and square (GSL),
Industrial land (IUL),
Commercial and business facilities (CBF),
Residential land (RUL),
Administration and public services (APS),
Road street and transportation (RST),
Logistics and warehouse (LWL),
Special use land (SUL),
Urban and rural construction land (URC),
Regional traffic facilities (RTF),
Regional public facilities (RPF),
Other construction land (OCL),
Mining use land (MUL)



3.1 | POI向量化—POI2Vec



Methods	Training Process		Predicting Process		Avg. Comp. Time
	OOB Avg. Error	OOB RMSE	OA	Kappa	
Proposed Method	0.1028 ± 0.0009	0.2301 ± 0.0013	0.8728 ± 0.0012	0.8399 ± 0.0007	161.0040
TF-IDF	0.1527 ± 0.0028	0.3384 ± 0.0043	0.5526 ± 0.0051	0.4162 ± 0.0023	146.3670
pLSA	0.1081 ± 0.0012	0.2361 ± 0.0021	0.7431 ± 0.0018	0.6719 ± 0.0010	3673.2530
LDA	0.1038 ± 0.0008	0.2375 ± 0.0016	0.6763 ± 0.0026	0.5841 ± 0.0019	1221.5500

W2V	PFL	GSL	IUL	CBF	RUL	APS	RST	LWL	SUL	URC	RTF	RPF	OCL	MUL
PFL	0.81	0.09	0.01	0.01	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GSL	0.00	0.89	0.01	0.01	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
IUL	0.00	0.06	0.86	0.01	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CBF	0.00	0.06	0.01	0.82	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RUL	0.00	0.05	0.01	0.01	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
APS	0.00	0.06	0.01	0.01	0.09	0.83	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RST	0.00	0.10	0.01	0.02	0.07	0.01	0.80	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LWL	0.00	0.07	0.05	0.01	0.07	0.00	0.00	0.80	0.00	0.00	0.00	0.00	0.00	0.00
SUL	0.00	0.04	0.01	0.01	0.10	0.01	0.00	0.00	0.84	0.00	0.00	0.00	0.00	0.00
URC	0.00	0.08	0.02	0.01	0.09	0.00	0.00	0.00	0.00	0.80	0.00	0.00	0.00	0.00
RTF	0.00	0.06	0.02	0.01	0.09	0.01	0.00	0.00	0.00	0.00	0.82	0.00	0.00	0.00
RPF	0.00	0.29	0.00	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.57	0.00	0.00
OCL	0.00	0.07	0.01	0.01	0.11	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.79	0.00
MUL	0.00	0.08	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.83

(a) Word2Vec

pLSA	PFL	GSL	IUL	CBF	RUL	APS	RST	LWL	SUL	URC	RTF	RPF	OCL	MUL
PFL	0.59	0.16	0.03	0.03	0.19	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GSL	0.00	0.79	0.03	0.02	0.14	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
IUL	0.00	0.13	0.70	0.01	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CBF	0.00	0.13	0.02	0.64	0.20	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RUL	0.00	0.10	0.02	0.02	0.86	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
APS	0.00	0.12	0.02	0.03	0.20	0.63	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RST	0.00	0.17	0.01	0.04	0.12	0.01	0.66	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LWL	0.01	0.14	0.07	0.01	0.18	0.01	0.00	0.59	0.00	0.00	0.01	0.00	0.00	0.00
SUL	0.00	0.13	0.03	0.03	0.24	0.01	0.00	0.00	0.56	0.00	0.00	0.00	0.00	0.00
URC	0.00	0.16	0.05	0.01	0.17	0.01	0.00	0.00	0.00	0.60	0.00	0.00	0.00	0.00
RTF	0.00	0.13	0.06	0.02	0.20	0.01	0.00	0.00	0.00	0.00	0.58	0.00	0.00	0.00
RPF	0.00	0.57	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.00
OCL	0.00	0.16	0.01	0.03	0.23	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.56	0.00
MUL	0.00	0.17	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.58

(c) pLSA

(b) TF-IDF

LDA	PFL	GSL	IUL	CBF	RUL	APS	RST	LWL	SUL	URC	RTF	RPF	OCL	MUL
PFL	0.43	0.32	0.03	0.03	0.17	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
GSL	0.00	0.78	0.03	0.02	0.16	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
IUL	0.00	0.23	0.58	0.02	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CBF	0.00	0.21	0.03	0.58	0.17	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RUL	0.00	0.15	0.02	0.02	0.80	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
APS	0.00	0.21	0.03	0.02	0.18	0.55	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RST	0.00	0.30	0.03	0.02	0.18	0.03	0.44	0.00	0.00	0.00	0.00	0.00	0.00	0.00
LWL	0.00	0.23	0.08	0.02	0.20	0.00	0.00	0.46	0.00	0.00	0.00	0.00	0.00	0.00
SUL	0.00	0.17	0.04	0.03	0.24	0.01	0.00	0.00	0.52	0.00	0.00	0.00	0.00	0.00
URC	0.00	0.33	0.03	0.02	0.18	0.01	0.00	0.00	0.00	0.44	0.00	0.00	0.00	0.00
RTF	0.00	0.26	0.04	0.04	0.24	0.01	0.00	0.02	0.00	0.01	0.39	0.00	0.00	0.00
RPF	0.00	0.71	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.14	0.00	0.00
OCL	0.00	0.16	0.04	0.02	0.28	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.50	0.00
MUL	0.00	0.67	0.00	0.00	0.17	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.17

(d) LDA

LDA/pLSA Param.:

- Topic Number: 200
- Max Iterations: 100
- Alpha: 0.025

Random Forest (RFA) Param.:

- Number of Trees: 100
- OOB Ratio: 0.5
- Repeat Times: 100

3.1 POI向量化—POI2Vec



Guangzhou



(a1) Actual land use

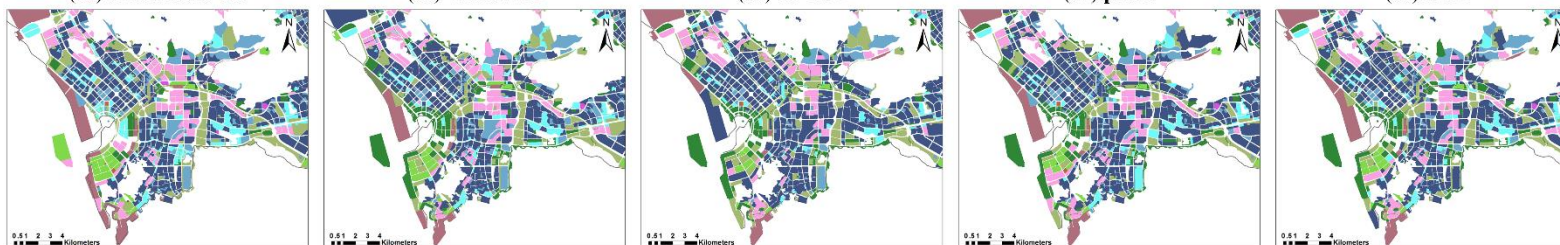
(a2) Word2Vec

(a3) TF-IDF

(a4) pLSA

(a5) LDA

Shenzhen



(b1) Actual land use

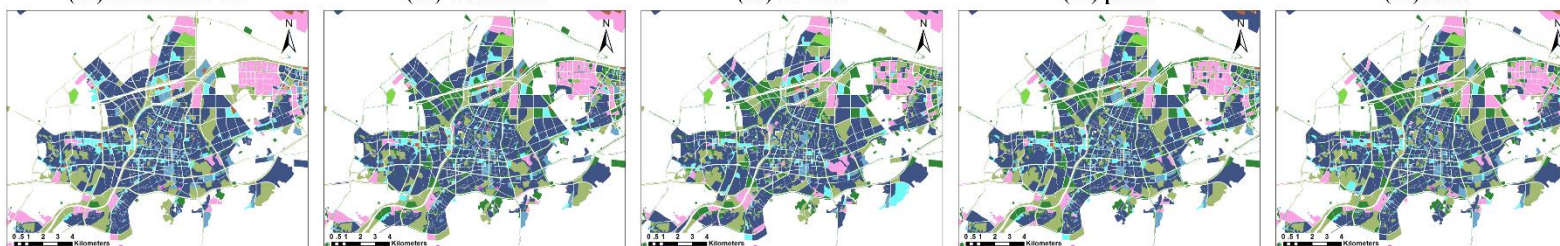
(b2) Word2Vec

(b3) TF-IDF

(b4) pLSA

(b5) LDA

Zhongshan



(c1) Actual land use

(c2) Word2Vec

(c3) TF-IDF

(c4) pLSA

(c5) LDA

Zhuhai



(d1) Actual land use

(d2) Word2Vec

(d3) TF-IDF

(d4) pLSA

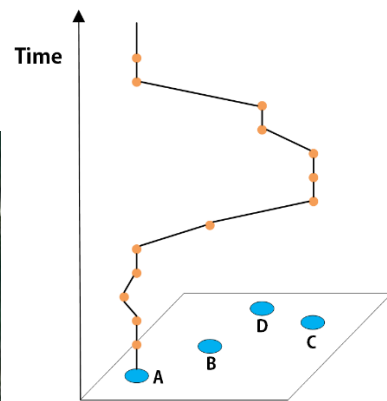
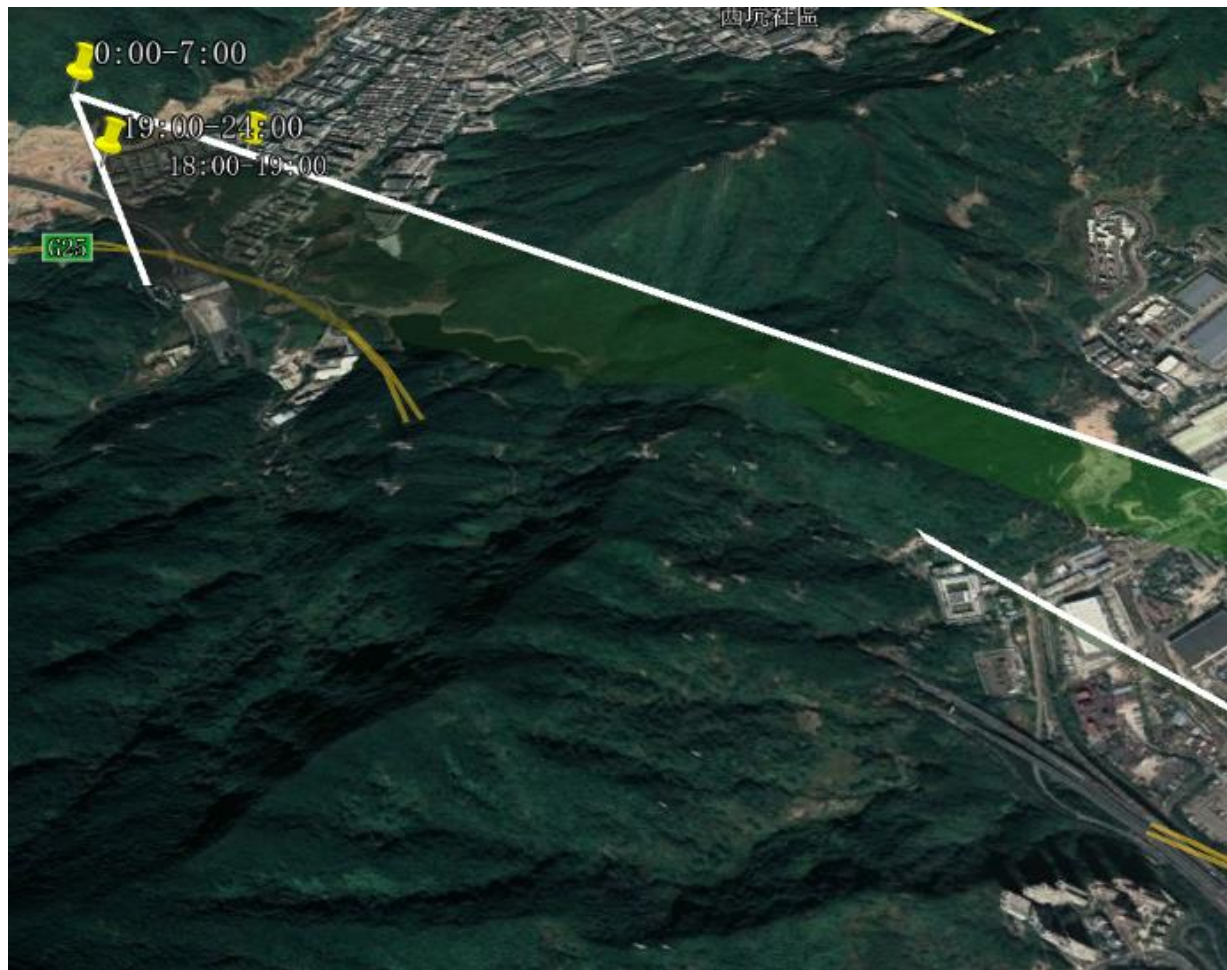
(d5) LDA



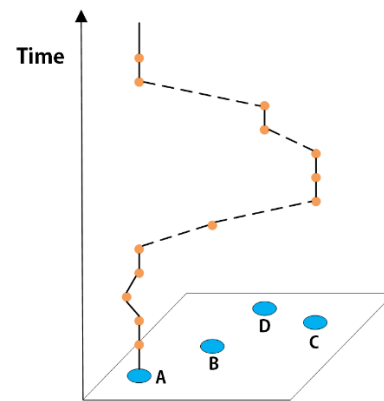
Land-use types:

Land-use types are Public facilities (PFL)
 Green space and square (GSL)
 Industrial land (IUL)
 Commercial and business facilities (CBF)
 Residential land (RUL)
 Administration and public services (APS)
 Road street and transportation (RST)
 Logistics and warehouse (LWL)
 Special use land (SUL)
 Urban and rural construction land (URC)
 Regional traffic facilities (RTF)
 Regional public facilities (RPF)
 Other construction land (OCL)
 Mining use land (MUL)

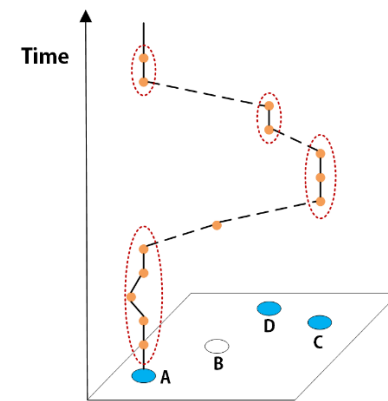
3.2 | 轨迹向量化—Traj2Vec



(a)trajectory

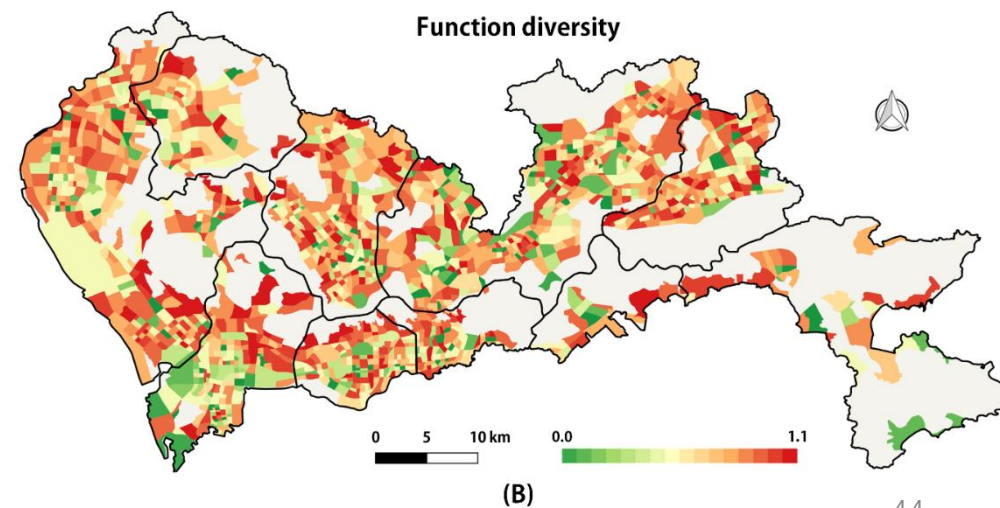
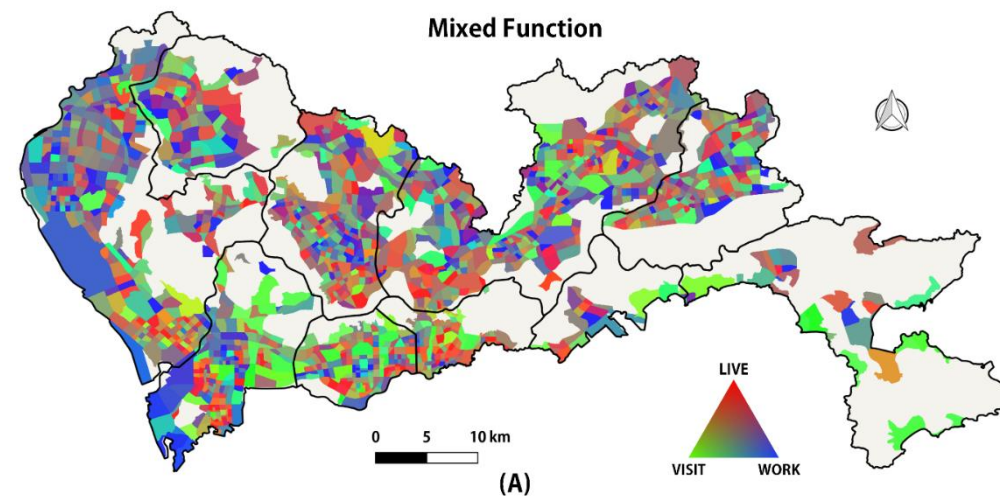
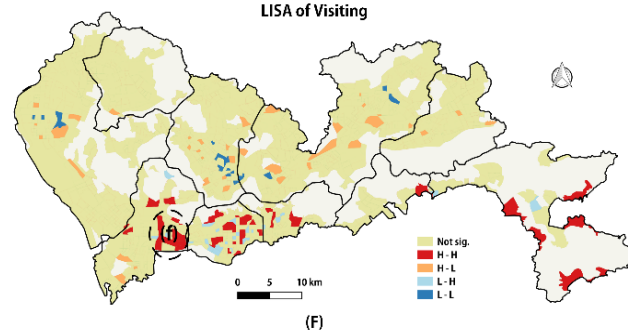
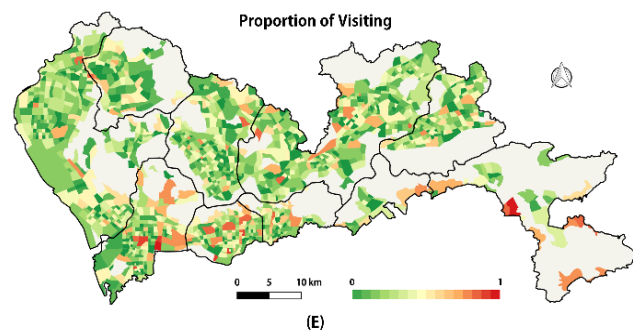
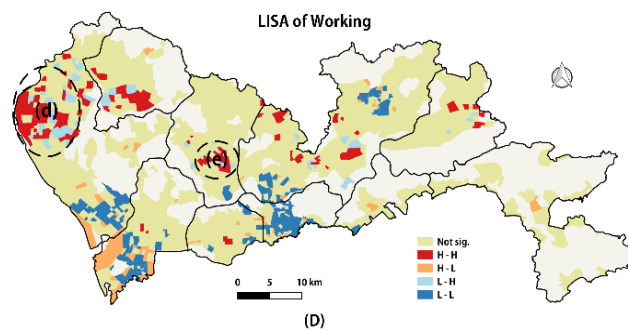
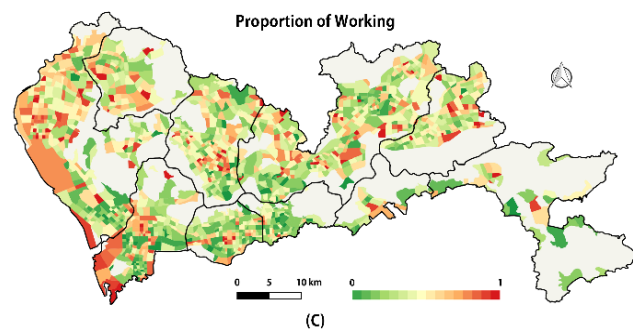
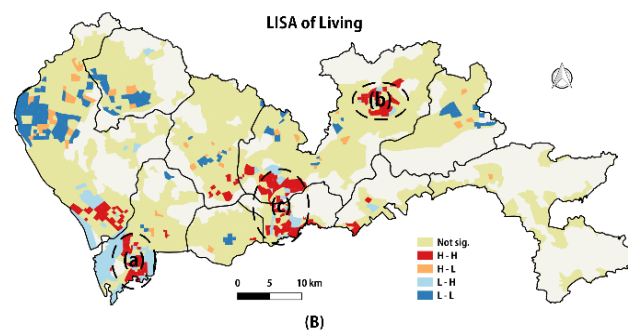
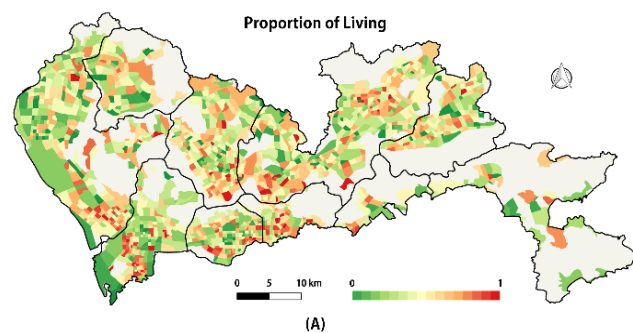


(b)segmentation



(c)anchor points

3.2 | 轨迹向量化—Traj2Vec



3.3 | 基于知识图谱的城市居民活动推断



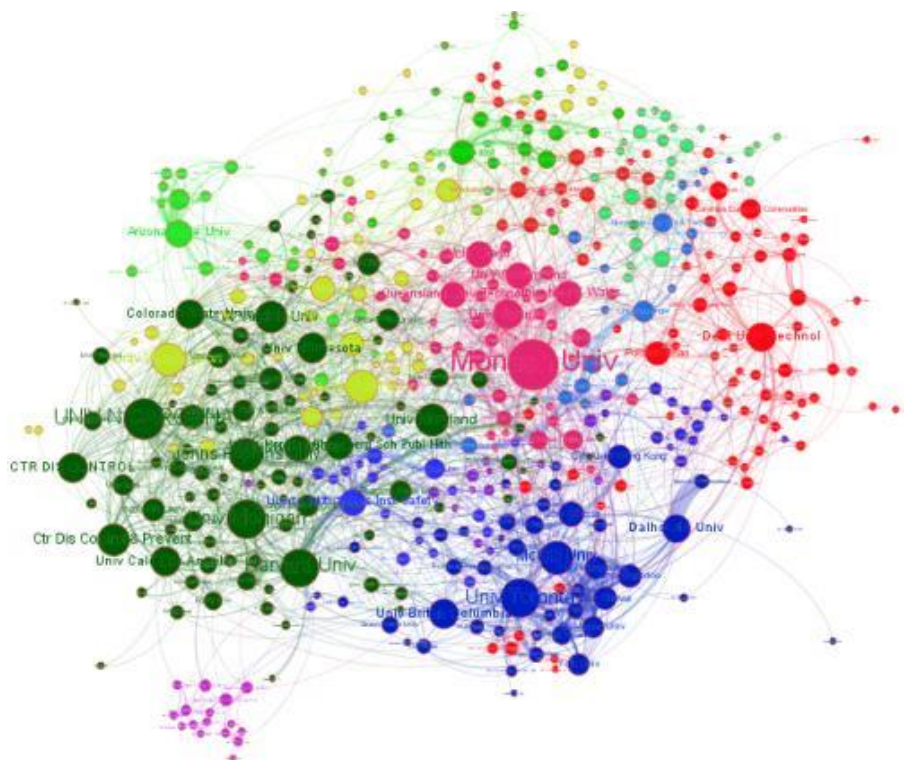
➤ 人们正在做什么？下一步要做什么？

采用数据：美团店铺数据及评论数据

店铺数据主要包含：店铺id、店铺名称、店铺人均消费、店铺星级、
店铺所属类型、店铺经纬度、店铺所在省市、店铺具体地址

评论数据主要包含：店铺名称、店铺类型、用户id、
评论发布时间、评论发布日期、评论文本内容、用户打分

- 如何从点评数据中有效挖掘与人类活动相关的信息?
- **构建知识图谱**为挖掘人类活动信息提供了有力支撑。



3.3 | 基于知识图谱的城市居民活动推断



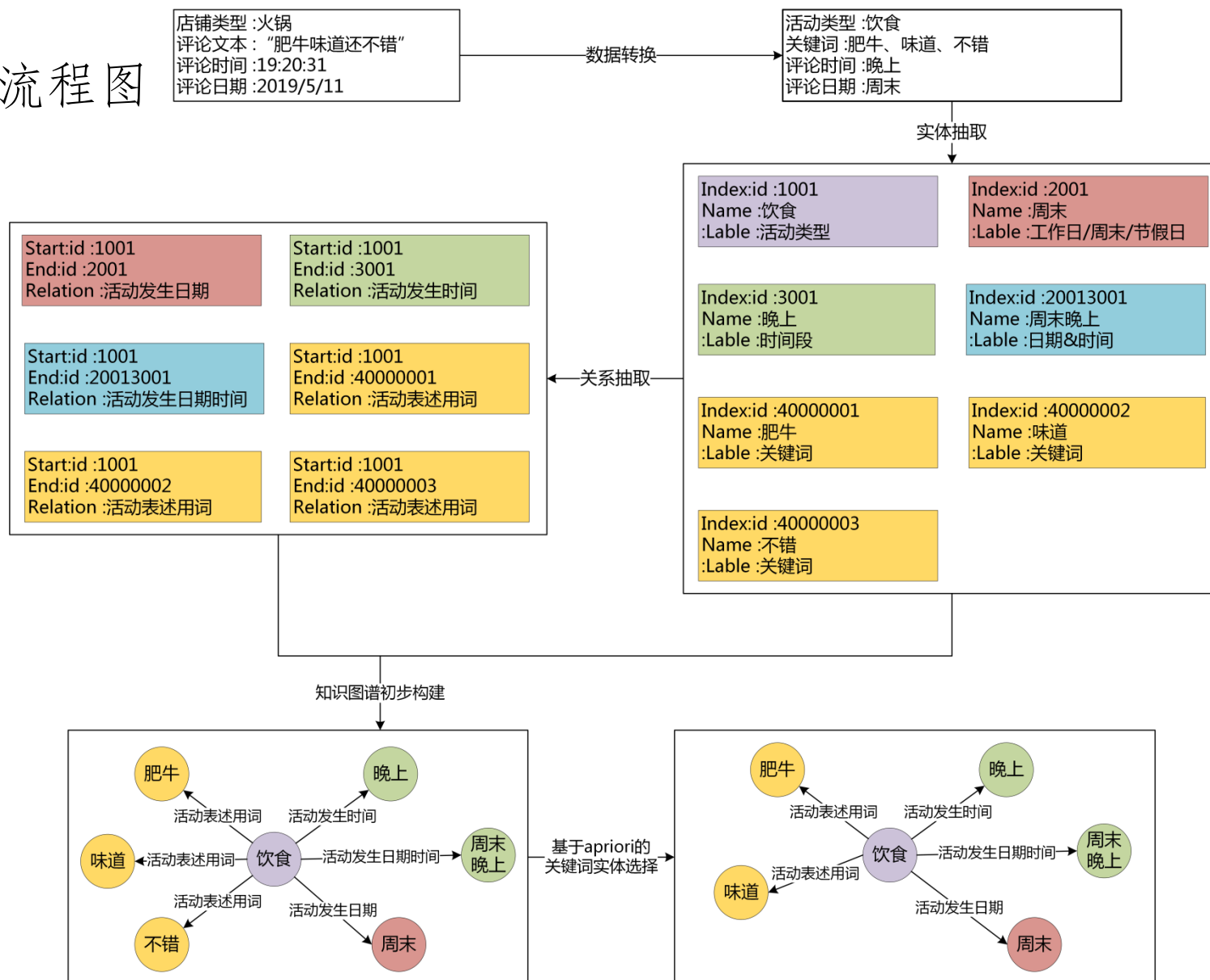
➤ 知识图谱构建——数据组成

活动类型	数据组成	抽取数据量
个人护理	抽取 全国部分 个人护理活动类型数据	75000
学习	抽取 全国所有 学习活动类型数据	67211
饮食	抽取 全国部分 饮食活动类型数据	75000
看病	抽取 全国部分 看病活动类型数据	74719
购物	抽取 全国部分 购物活动类型数据	76796
运动	抽取 全国部分 运动活动类型数据	76662
娱乐	抽取 全国部分 娱乐活动类型数据	70627
住宿	抽取 全国部分 住宿活动类型数据	73103
游玩	抽取 全国各景点 游玩活动类型数据	78186

3.3 基于知识图谱的城市居民活动推断



➤ 活动知识图谱构建流程图



➤ 知识图谱构建——数据转换

- 店铺类型转活动类型
 - 将活动类型划分为9类，分别为：饮食、娱乐、住宿、运动、游玩、个人护理、学习、看病、购物。店铺类型到活动类型的转换规则如下表所示：



店铺类型	活动描述	活动类型
海鲜、汤/粥/炖菜、烧烤烤肉、小吃快餐、自助餐、火锅等	与吃喝有关的所有活动	饮食
桌游、轰趴馆、酒吧、量贩式KTV等	室内场所的娱乐性质活动	娱乐
酒店住宿	在室内休息场所进行的活动	住宿
网球、羽毛球、健身中心、游泳等	关于健身的所有活动	运动
周边游	在户外场所进行的娱乐性质活动	游玩
足疗/按摩、医学美容、美甲美睫、美容美体、美发等	个人放松或美容美发的相关活动	个人护理
美术培训、语言培训、驾校、自习室、音乐培训等	有关学习或技能提升的活动	学习
体检中心、齿科、眼科、中医、妇幼医院等	有关于恢复/保持自身健康的活动	看病
化妆品、家具家居、配镜、鲜花等	除饮食外的实体消费	购物

3.3 | 基于知识图谱的城市居民活动推断



➤ 知识图谱构建——数据转换

- 文本转关键词

通过jieba库抽取文本关键词

店铺类型 :火锅
评论文本 :“肥牛味道还不错”
评论时间 :19:20:31
评论日期 :2019/5/11

数据转换

活动类型 :饮食
关键词 :肥牛、味道、不错
评论时间 :晚上
评论日期 :周末

- 时间转时间段

时间划分为时间段，如：凌晨，清晨，上午，中午，……

00:00:00-04:59:59 凌晨

05:00:00-06:59:59 清晨

07:00:00-08:59:59 早上

09:00:00-11:59:59 上午

12:00:00-13:59:59 中午

14:00:00-17:59:59 下午

18:00:00-18:59:59 傍晚

19:00:00-23:59:59 晚上

- 日期转工作日/节假日/周末

日期转换为其所对应的工作日/节假日/周末

3.3 基于知识图谱的城市居民活动推断



➤ 知识图谱构建——实体抽取&关系抽取

活动类型 :饮食
关键词 :肥牛、味道、不错
评论时间 :晚上
评论日期 :周末

实体抽取

Index:id :1001
Name :饮食
:Lable :活动类型

Index:id :2001
Name :周末
:Lable :工作日/周末/节假日

Index:id :3001
Name :晚上
:Lable :时间段

Index:id :20013001
Name :周末晚上
:Lable :日期&时间

Index:id :40000001
Name :肥牛
:Lable :关键词

Index:id :40000002
Name :味道
:Lable :关键词

Index:id :40000003
Name :不错
:Lable :关键词

关系抽取

Start:id :1001
End:id :2001
Relation :活动发生日期

Start:id :1001
End:id :3001
Relation :活动发生时间

Start:id :1001
End:id :20013001
Relation :活动发生日期时间

Start:id :1001
End:id :40000001
Relation :活动表述用词

Start:id :1001
End:id :40000002
Relation :活动表述用词

Start:id :1001
End:id :40000003
Relation :活动表述用词

3.3 基于知识图谱的城市居民活动性分析



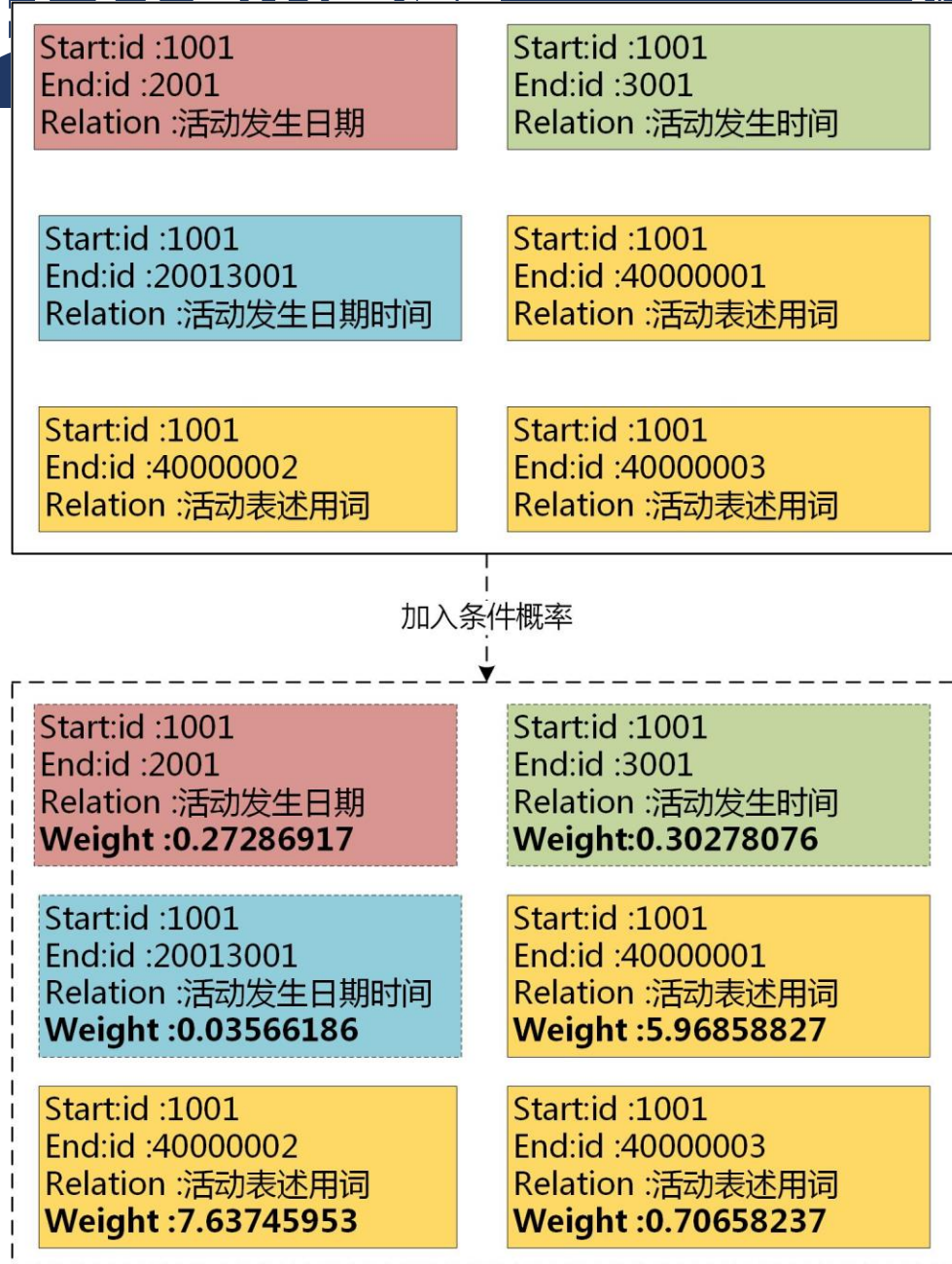
➤ 知识图谱构建——实体抽取&关系抽取

➤ 关系权重计算

知识图谱中关系间的权重由点间互信息（PMI）进行计算，PMI主要用于计算词语间的语义相似度。

基本思想是统计两个词语在文本中同时出现的概率，如果概率越大，其相关性就越紧密，关联度越高。两个词语word₁与word₂的PMI值计算公式如下式所

$$PMI(\text{word}_1, \text{word}_2) = \frac{P(\text{word}_1, \text{word}_2)}{P(\text{word}_1) * P(\text{word}_2)}$$



3.3 基于知识图谱的城市居民活动推断



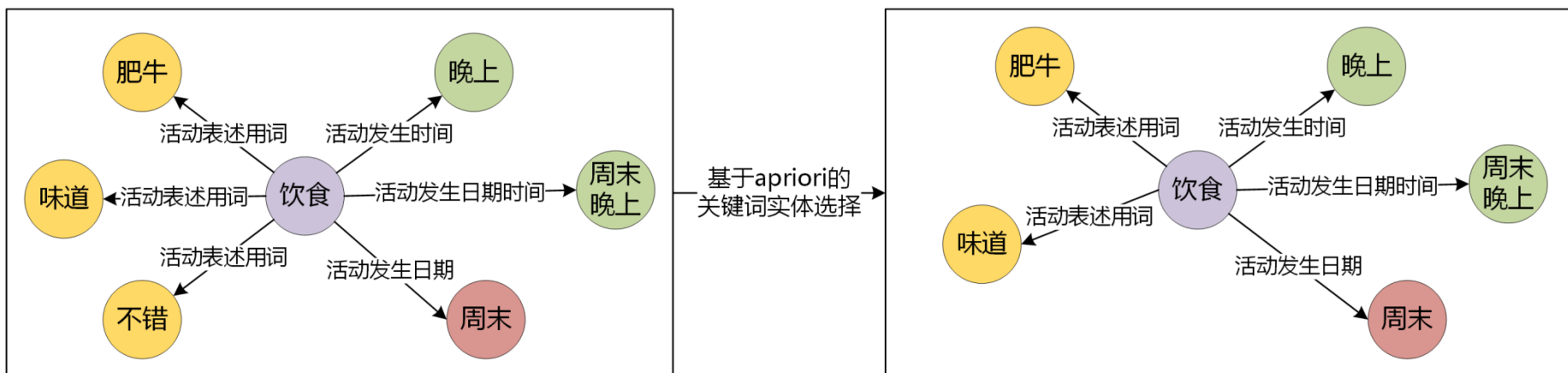
➤ 知识图谱构建——实体抽取&关系抽取

- 基于Apriori算法进行实体筛选

Apriori是一种关联分析算法，通过计算支持度（support）并选择支持度阈值，获得频繁项集，然后通过支持度计算关系间的置信度（confidence）并选择置信度阈值，制定关联规则，保留关联强度恰当的关系对。其中：

支持度 $\text{support}(A)=P(A)$, $\text{support}(A \Rightarrow B) = P(A \cup B)$

置信度 $\text{confidence}(A \Rightarrow B) = \frac{\text{support}(A \Rightarrow B)}{\text{support}(A)}$

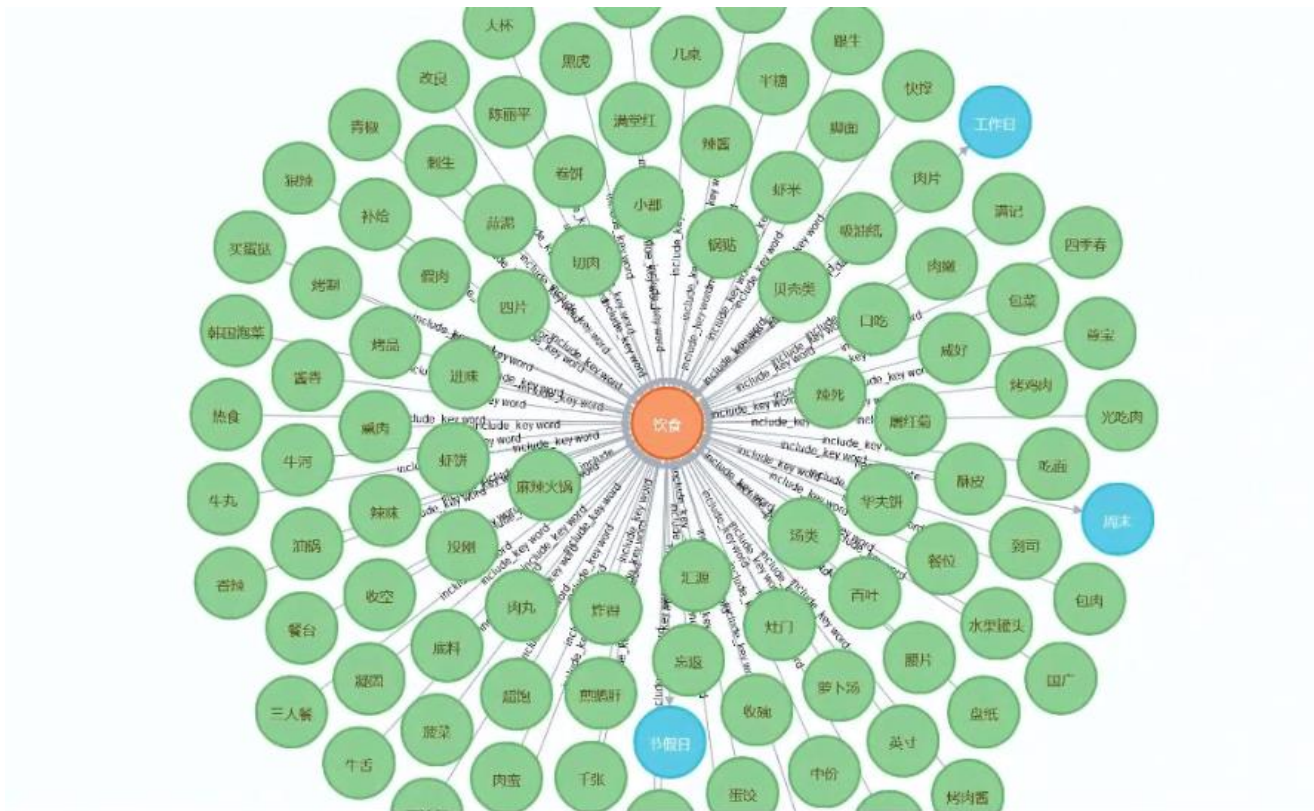


3.3 | 基于知识图谱的城市居民活动推断



➤ 知识图谱构建——Neo4j存储

- Neo4j是一个高性能的,NOSQL图形数据库, 它将结构化数据存储在网络上而不是表中。它是一个嵌入式的、基于磁盘的、具备完全的事务特性的Java持久化引擎, 但是它将结构化数据存储在网络上(从数学角度叫做图)上而不是表中。Neo4j也可以被看作是一个高性能的图引擎, 该引擎具有成熟数据库的所有特性。

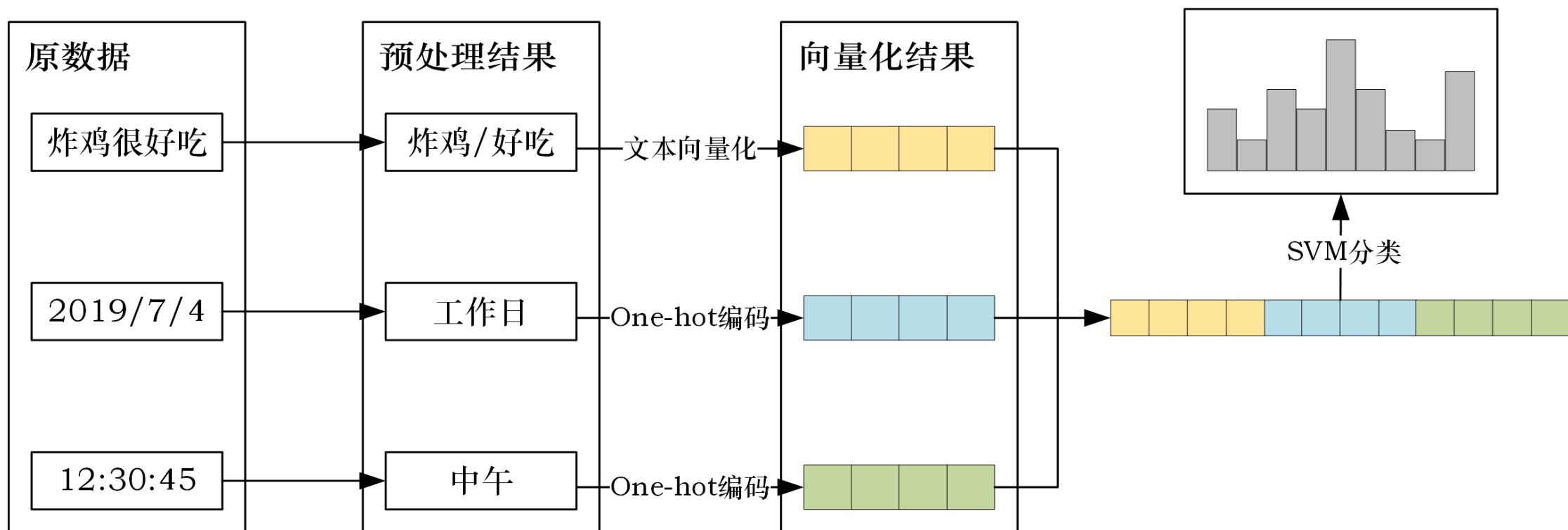


3.3 | 基于知识图谱的城市居民活动推断



➤ 基于svm进行的人类活动判定

- 不加入知识图谱

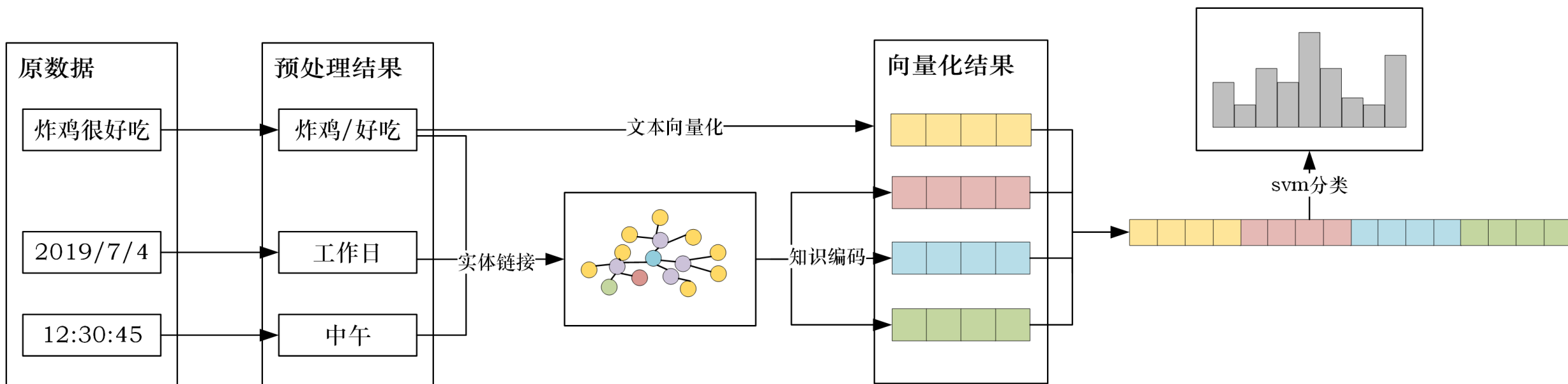


3.3 | 基于知识图谱的城市居民活动推断



➤ 基于svm进行的人类活动判定

- 加入知识图谱



3.3 | 基于知识图谱的城市居民活动推断



➤ 基于svm进行的人类活动判定

- 训练数据

	个人护理	学习	饮食	看病	购物	运动	娱乐	住宿	游玩	总计
数据量 (条)	3000	3000	3000	3000	3000	3000	3000	3000	3000	27000

- 测试数据

	个人护理	学习	饮食	看病	购物	运动	娱乐	住宿	游玩	总计
数据量 (条)	1500	1500	1500	1500	1500	1500	1500	1500	1500	13500

3.3 | 基于知识图谱的城市居民活动推断



➤ 基于svm进行的人类活动判定

- 不加入知识图谱

	个人护理	学习	饮食	看病	购物	运动	娱乐	住宿	游玩
PRECISION	0.58942	0.76912	0.75783	0.63537	0.61935	0.62428	0.59444	0.73680	0.55086
REALL	0.54104	0.71881	0.73731	0.70884	0.67488	0.66417	0.66277	0.67997	0.47924
F1	0.56419	0.74312	0.74743	0.67010	0.64592	0.64361	0.62675	0.70725	0.51256

- 加入知识图谱

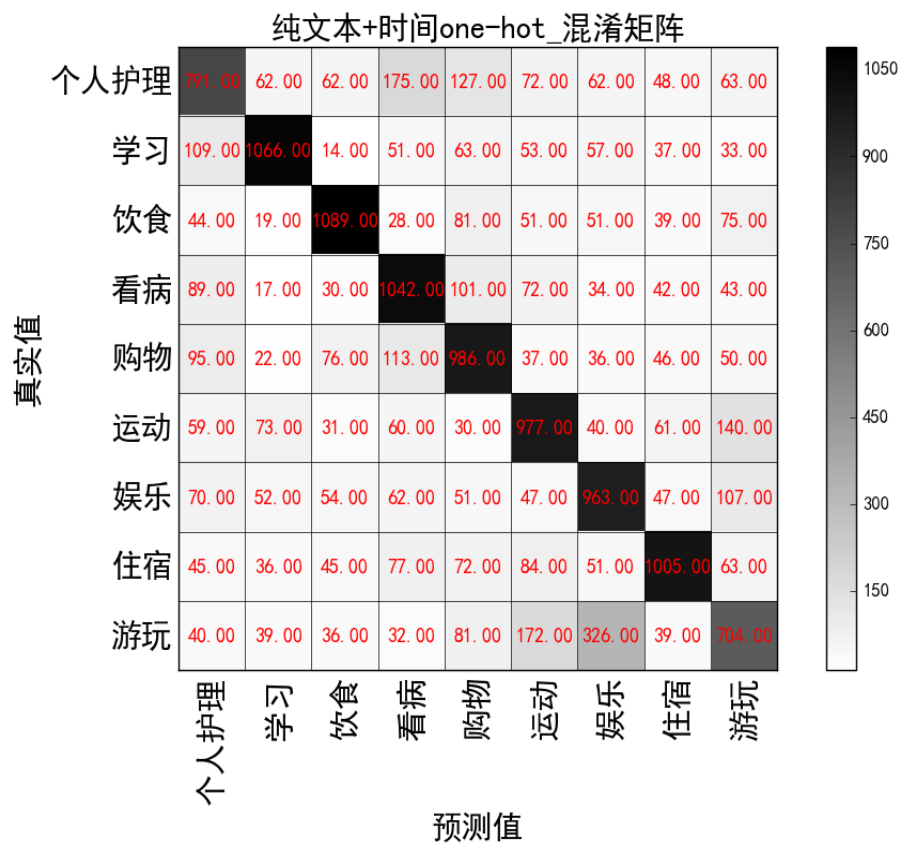
	个人护理	学习	饮食	看病	购物	运动	娱乐	住宿	游玩
PRECISION	0.82653	0.89159	0.84874	0.89024	0.85378	0.83525	0.80586	0.80975	0.82385
REALL	0.74029	0.89408	0.85775	0.86534	0.85567	0.8793	0.87435	0.86767	0.74489
F1	0.78104	0.89283	0.85322	0.87761	0.85473	0.85675	0.83871	0.83771	0.78238

3.3 基于知识图谱的城市居民活动推断

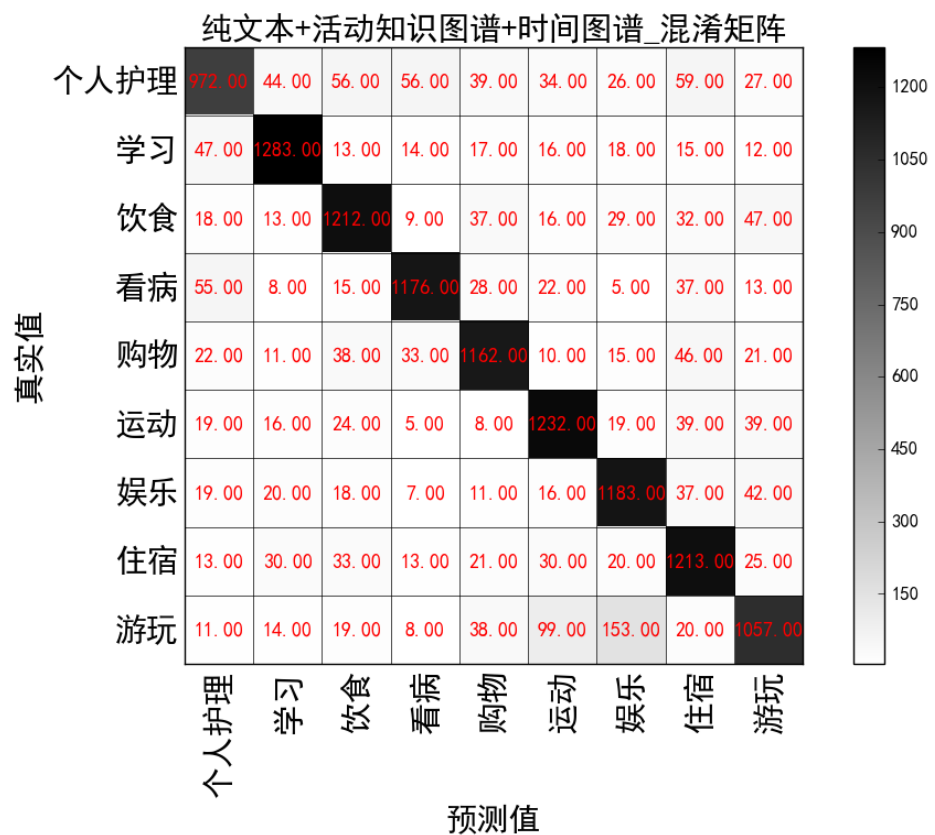


➤ 基于svm进行的人类活动判定

- 不加入知识图谱



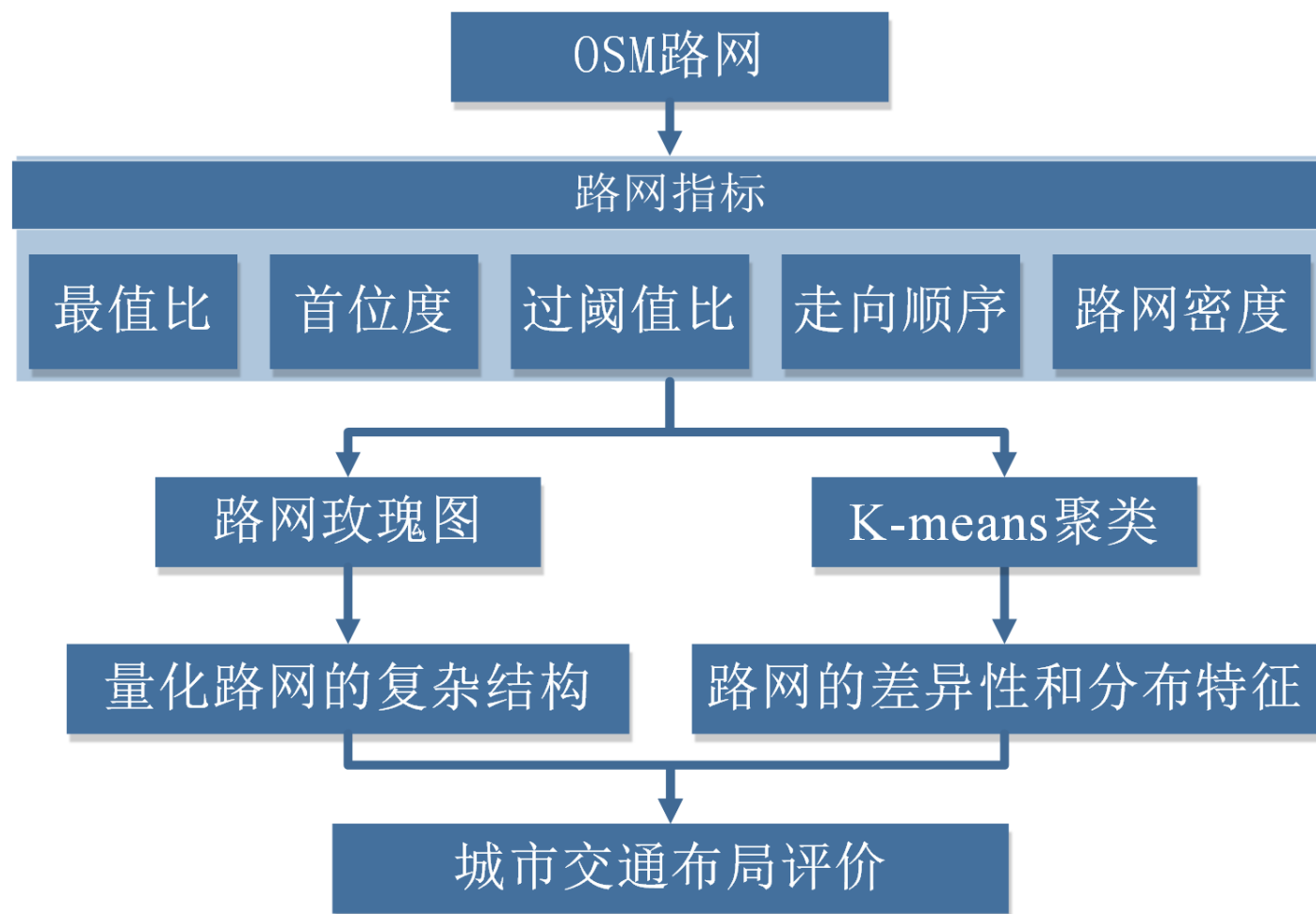
- 加入知识图谱



3.4 | 基于OSM的中国城市交通布局评价



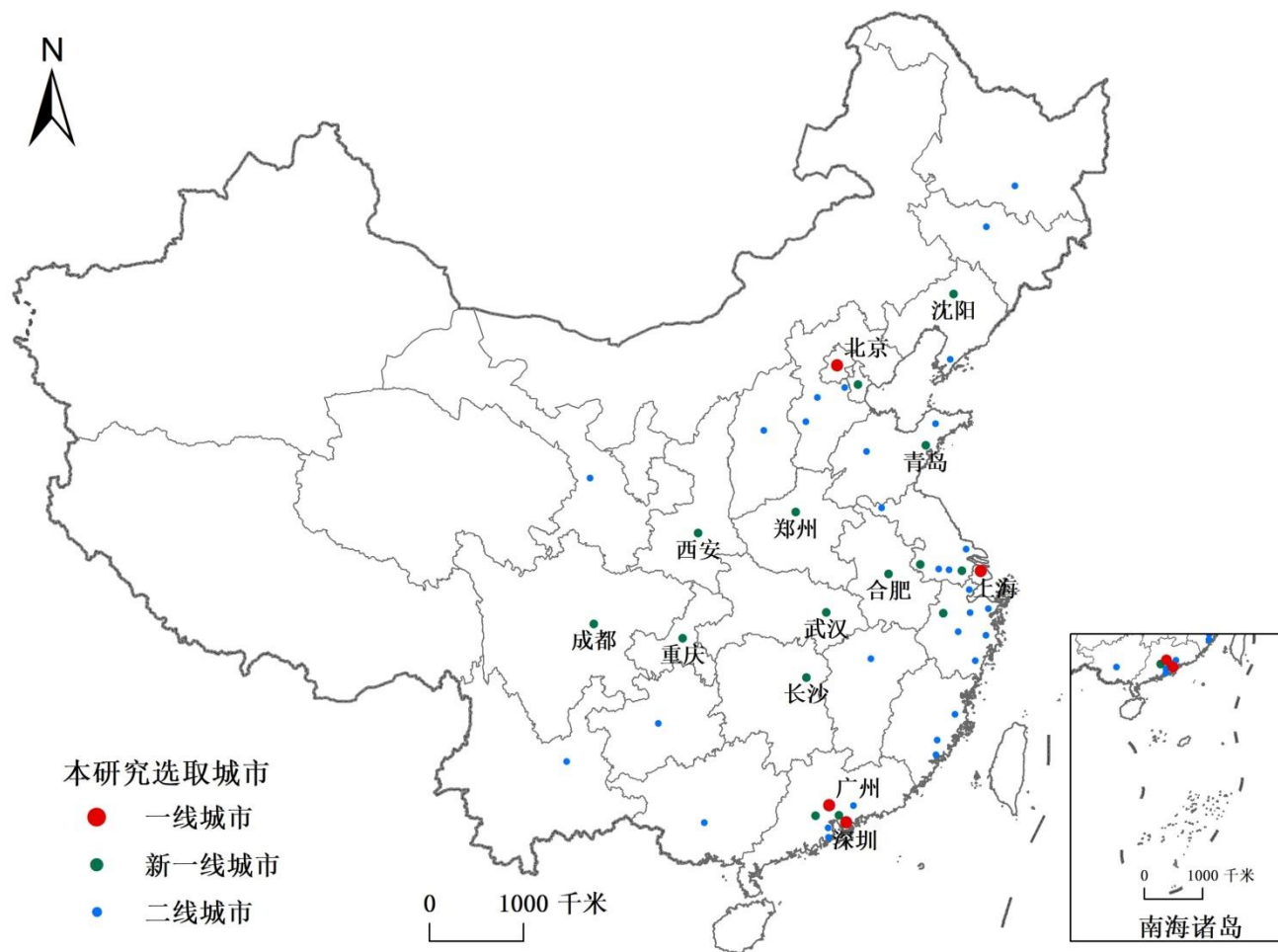
如何基于OSM路网数据和路网指标，利用玫瑰图和聚类方法探究国内主要城市路网的差异性和分布特征，对城市交通布局进行系统性评价？



3.4 | 基于OSM的中国城市交通布局评价



➤ 数据介绍



- 研究区：49个一、二线及新一线城市
- 数据：OSM路网数据
- 获取时间：2020年 2月 20 日 数据格式：shp 矢量数据
- 投影坐标系：
- WGS_1984_Mercator 坐标系

3.4 基于OSM的中国城市交通布局评价



➤ 路网指标选取

(1) **最值比R**: 道路长度的最大值和最小值之比, a 和 b 为道路长度的最小值和最大值, 公式如下:

$$g(x) = a' + \frac{b' - a'}{b - a} (f(x) - a)$$

(2) **首位度S**: 研究城市之间相互关系、衡量城市规模合理性的理论概念, 该指标用于突出道路走向的集中程度, 公式如下

$$S = \frac{P_2}{P_1}$$

(3) **过阈值比T**: 选择道路最大长度的某比值作为阈值, 然后对大于这一阈值的数据进行计数并统计其占比作为该项指标数值, 计算公式如下

$$T = \frac{\text{countif}(x > T_v)}{n} \times 100\%$$

(4) **走向顺序 φ** : 外国学者Boeing在路网研究中提出的计算指标, 用于量化城市街道网络遵循单个网格排序的程度

$$H_0 = -\sum_{i=1}^n P(O_i) \ln(P(O_i)) \quad \varphi = 1 - \left(\frac{H_0 - H_g}{H_{\max} - H_g} \right)^2$$

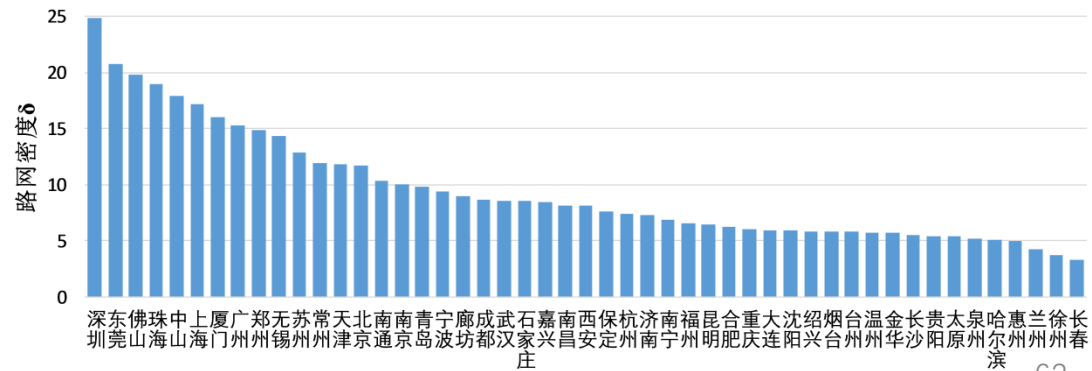
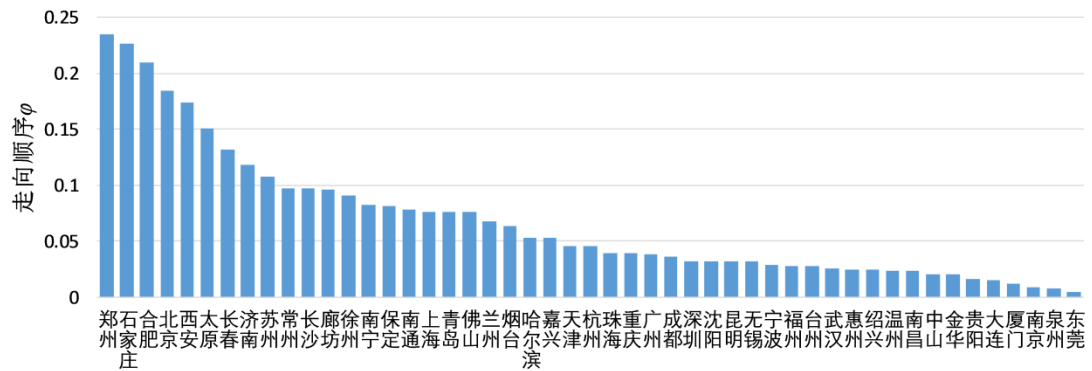
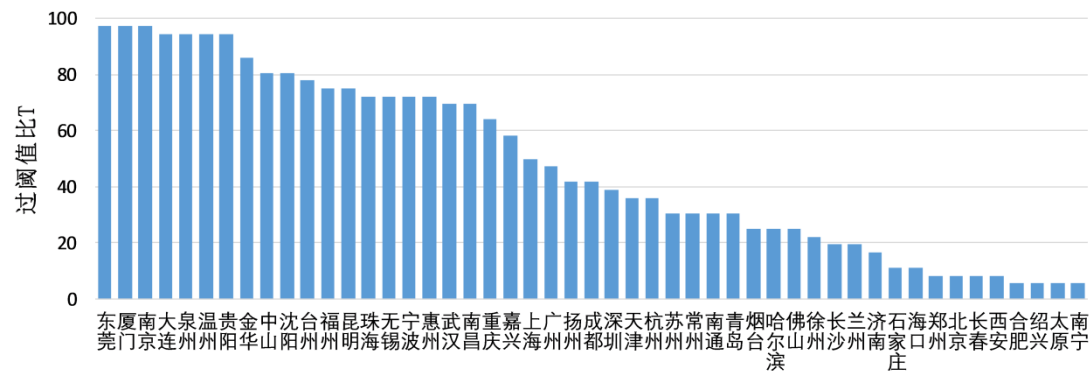
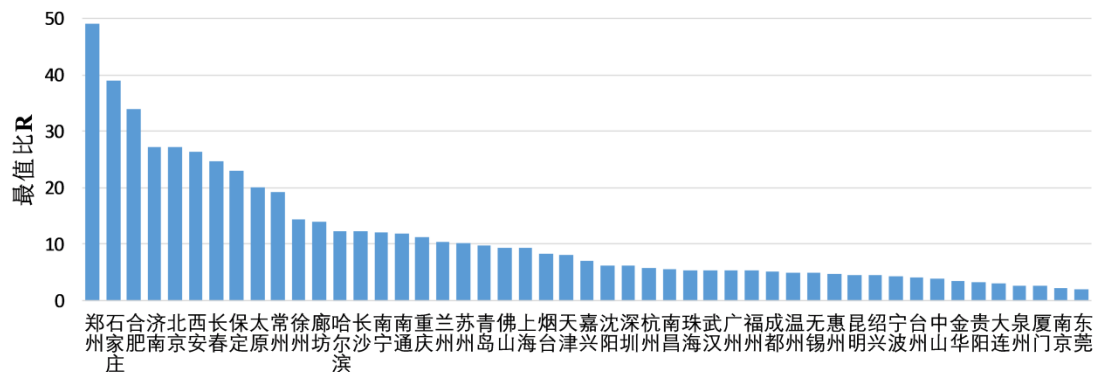
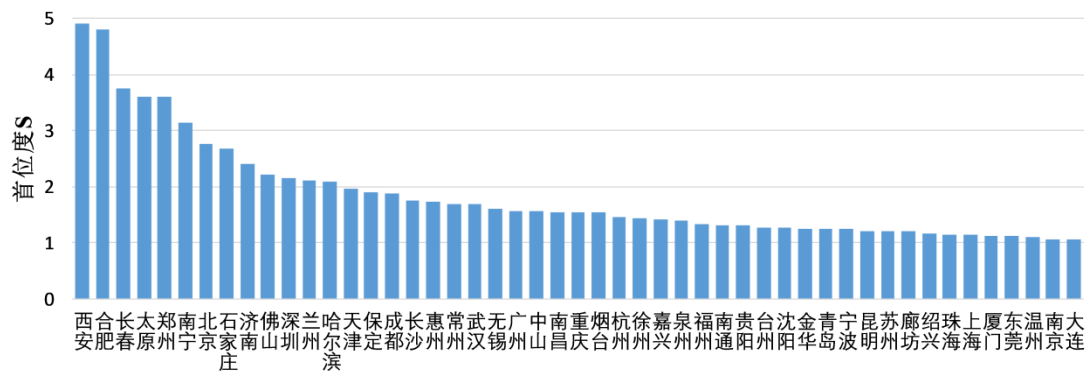
(5) **路网密度 δ** : 城市道路总长度与城市面积之比, 从长度上描述了不同城市规模应有的道路发展水平

$$\delta = \frac{\sum x}{F}$$

3.4 基于OSM的中国城市交通布局评价



➤ 路网指标数值分布

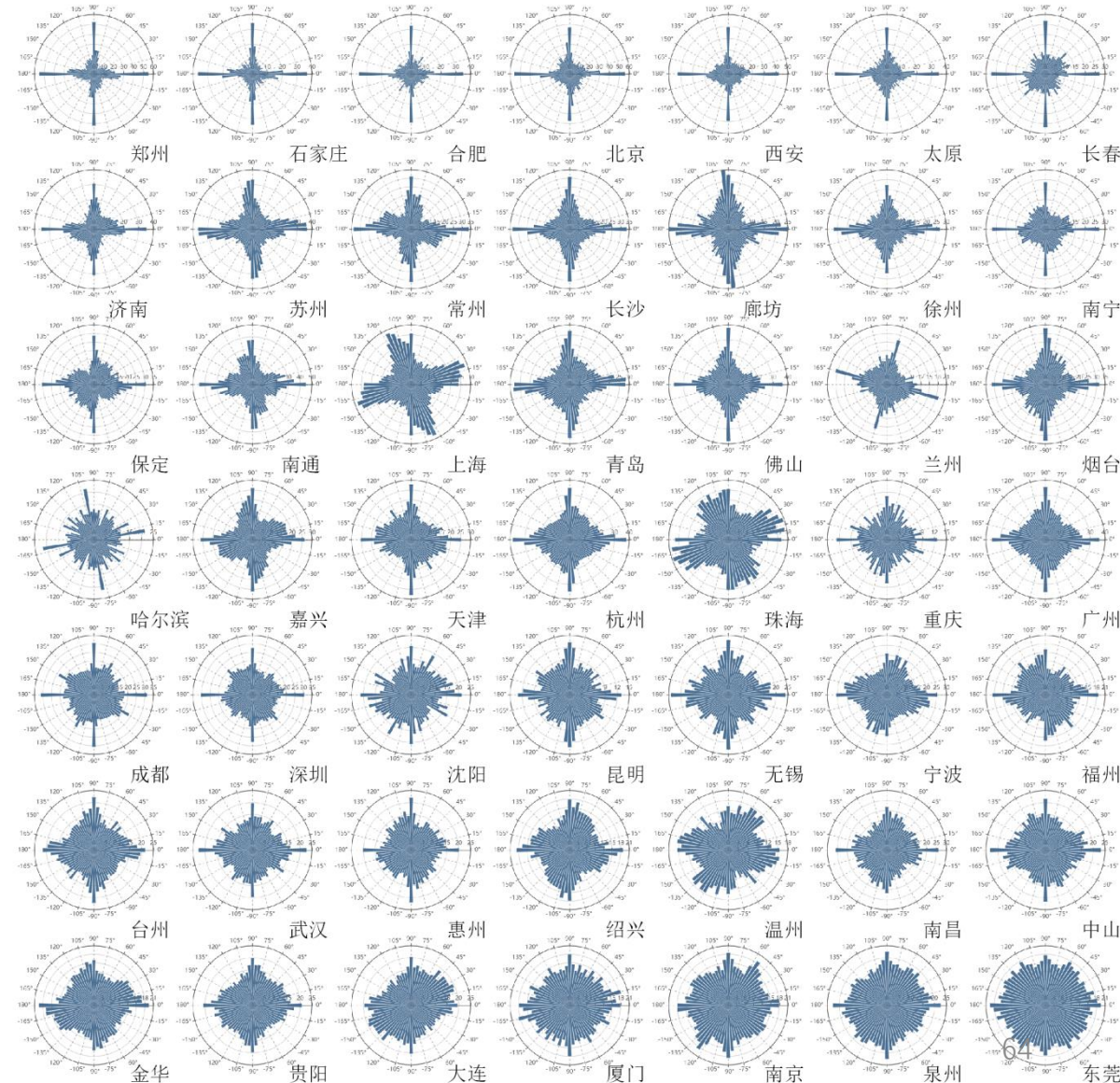


3.4 基于OSM的中国城市交通布局评价



➤ 路网玫瑰图

- 玫瑰图是表示二维方向数据频率分布的图形方法，善于表达方向性数据。玫瑰图的条带方向表示道路的方位，条带长度表示与这些方位的相对频率。
- 计算：以正东方向开始，沿顺时针和逆时针方向各展开二分之一分带作为第一个区间。此后在逆时针方向上以一个分带作为一个统计区间，对每个方向道路长度进行统计



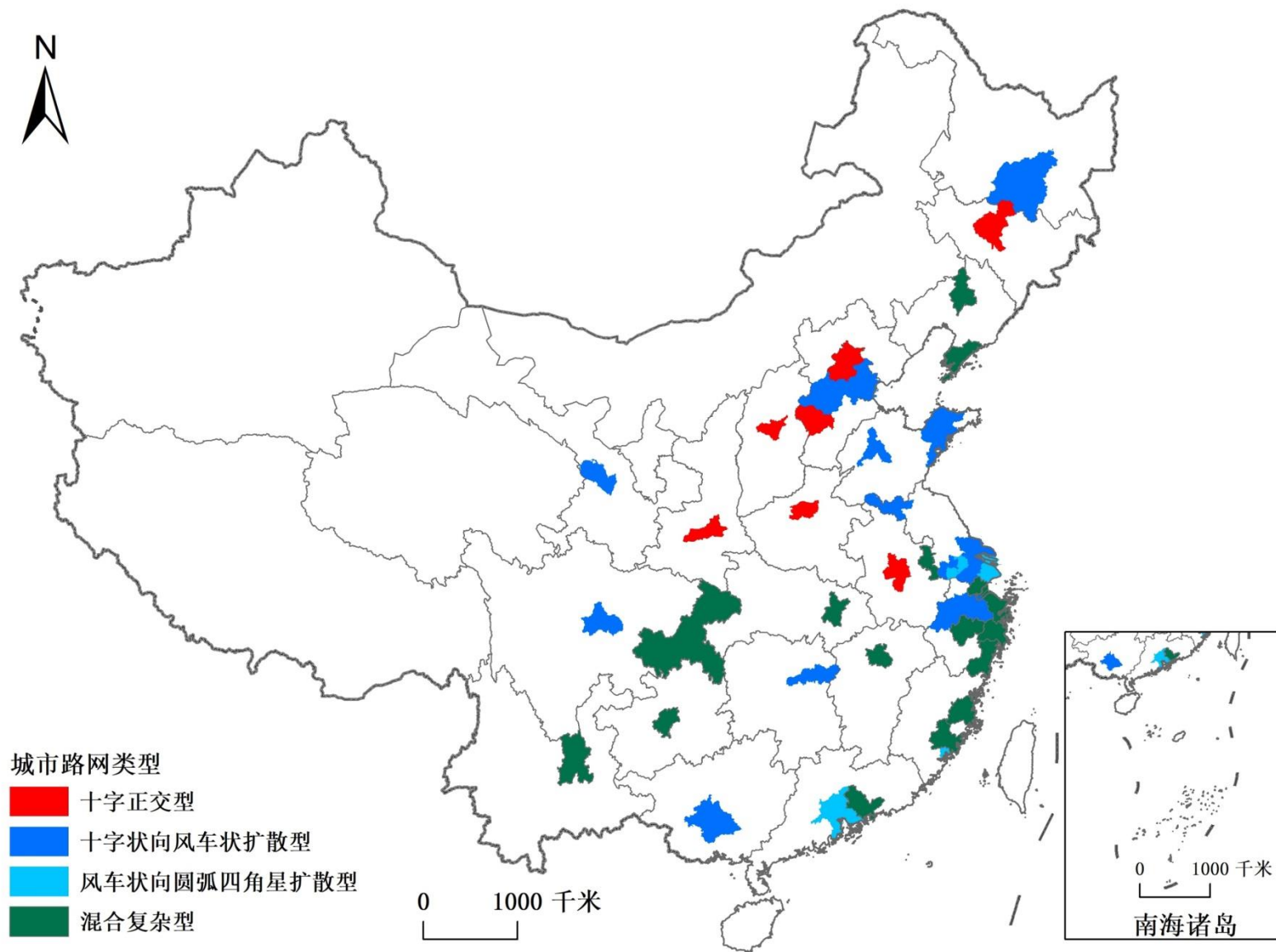
3.4 | 基于OSM的中国城市交通布局评价



- 采用K-Means聚类对研究的5个路网指标进行聚类分析
- 基于路网的4种典型基本形式，同时参考评价聚类效果好坏的轮廓系数，选取聚类数K值为4，最终将49个一、二线城市分为4类

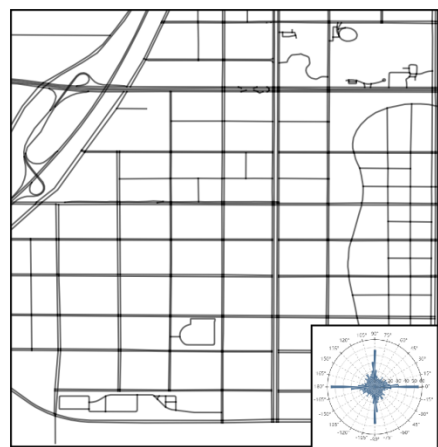
类别	城市
I	郑州、北京、石家庄、西安、合肥、太原、长春
II	苏州、常州、天津、南通、青岛、成都、保定、廊坊、杭州 济南、绍兴、长沙、烟台、兰州、徐州、哈尔滨、南宁
III	深圳、东莞、佛山、珠海、上海、中山、厦门、广州、无锡
IV	南京、宁波、武汉、嘉兴、南昌、福州、台州、金华、大连 沈阳、温州、贵阳、泉州、惠州、昆明、重庆

3.4 | 基于OSM的中国城市交通布局评价



- 十字正交型和十字状向风车状扩散型路网大多分布在内陆地区，而风车状向圆弧四角星扩散型和混合复杂型路网大多分布在沿海地区
- 聚类结果形成这样一个空间分布与我国各个城市的地形地势密切相关，比如十字正交型路网基本位于我国地势平坦的平原地区

3.4 基于OSM的中国城市交通布局评价



A. 北京

均值±标准差:
R: 31.459 ± 9.956
S: 3.724 ± 0.878
T: 7.539 ± 2.099
 ϕ : 0.187 ± 0.038
 δ : 8.317 ± 3.932



B. 常州

均值±标准差:
R: 12.257 ± 5.934
S: 1.739 ± 0.501
T: 25.490 ± 10.272
 ϕ : 0.074 ± 0.025
 δ : 7.887 ± 2.684



C. 珠海

均值±标准差:
R: 5.479 ± 2.590
S: 1.515 ± 0.435
T: 64.506 ± 25.601
 ϕ : 0.036 ± 0.025
 δ : 18.346 ± 3.209



D. 贵阳

均值±标准差:
R: 4.902 ± 2.132
S: 1.341 ± 0.202
T: 80.034 ± 12.838
 ϕ : 0.025 ± 0.011
 δ : 6.004 ± 1.607

- A. 十字正交型：道路网为十字正交型或者接近十字正交型网络，道路整齐有序，主导干道以外方向上的道路较少
- B. 十字状向风车状扩散型：适当增加其他方向的道路，降低了道路网的旅行距离，保证城市道路网的通畅有序
- C. 风车状向圆弧四角星扩散型：正交网络变少，交叉道路的角度变大。城市的无序性变得更强，路网更加复杂
- D. 混合复杂型：因地势起伏等地理因素导致其城市道路规模形态较为复杂，这类城市是最为复杂和无序的



总结



本章介绍了VGI大数据的**数据主要来源、数据特点及优势**以及**VGI大数据关键技术**以及其在**城市功能、城市居民活动判定、城市交通布局**等方面的应用。

VGI大数据主要具有**易获取、数据量大、覆盖面广、更新速度快、大众化**等特点。

VGI大数据主要形式为**VGI信息收集平台数据、带有地理位置的社交媒体数据**。

VGI大数据中带有地理位置的社交媒体数据在**细粒度理解城市社会经济特征、构建细粒度居民活动模式**等方面有广泛的应用。

基于VGI信息收集平台所得数据（如：**OSM**）为**城市规划、评估城市交通布局状况**等提供决策支持。



资料下载



HPSCIL@CUG城市计算小组: www.urbancomp.net

微信公众号: **iGEODATA**

HPSCIL@CUG城市计算小组: www.urbancomp.net

首页 团队新闻 研究成果 授权专利 在研课题 专题展示 软件下载 资料下载 团队成员 联系我们

VecLI v2.0.0: 基于真实地块的矢量景观指数计算与分析系统 发布于 2021-08-17

在VecLI v1.0.0发布后, 我们随即开始了VecLI v2.0.0的开发与测试。考虑到原始版本中实现的景观指数种类较少, 难以充分满足相关领域研究者和城市规划人员的研究需求。为此, 我们在新版本的软件中设计了一套全新的景观指数计算模块, 并提供了类型丰富、种类更多的景观指数和完善的说明文档。Our

TraceVideo v1.0: 带有位置信息的视频流数据采集软件 发布于 2021-06-23

TraceVideo v1.0是基于手机视频和传感器, 获取位置信息和街景视频流的Android手机软件。TraceVideo v1.0会要求手机的摄像头、GPS和存储权限, 在用户移动的过程中, 记录视频, 及视频的起始、终点时间, 和视频记录时间全过程的位置信息。开发者团队研究生 王宇同学: 9164910

StreetViewRatings v1.0: 支持人机对抗的图片打分软件 发布于 2021-06-10

StreetViewRatings Version 1.0是我们HPSCIL团队自主研发的支持人机对抗的图片打分软件(单机版)。StreetViewRatings Version 1.0 is an image scoring software (Standalone) that supports

<https://www.urbancomp.net/categories/thematicdisplay>



团队成

发布于

投下的第

行选址

热度 0 条评论

投放位置的选择, 营销市场具有十分数据难获取, ...



切换主题 | SCHEME TOOL

发布于 2021-07-06

Analyzing the correlation between visual space and residential space

86 热度 0 条评论

ABSTRACTAlthough previous studies have assessed the relationships between visual space indicators and urban residential space, the results are inconsistent. This study aims to explore the correlation between visual space indicators and urban residential space. The study area is the urban area of Wuhan, China. The study uses a combination of visual space indicators and urban residential space data to analyze the correlation. The results show that the correlation between visual space indicators and urban residential space is significant. The study provides a new perspective on the relationship between visual space and urban residential space.

切换主题 | SCHEME TOOL

发布于 2021-07-06

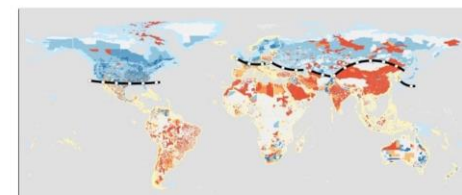


切换主题 | SCHEME TOOL

9:49

iGEODATA 推送汇总

iGEODATA



最新进展 | 全球城市发展模式的南北划分线

论文报告 开放研究 资料公开 推荐稿件

最新进展 | 城市视觉空间和居民心理感知的关系及影响机制

原创:iGEODATA



最新进展 | 倡导高效化还是侧重均等化? 以西安市文化设施空间分布为例

原创:iGEODATA



最新进展 | 全球城市发展模式的南北划分线

原创:张亚清



最新进展 | 基于路网数据和复杂图论的中国城市交通布局评价



最新进展 | cuFSDAE: 基于GPU并行





Discussion!

姚尧 博士, 副教授

中国地质大学 (武汉) 地理与信息工程学院

阿里巴巴集团

Email: yaoy@cug.edu.cn

UrbanComp团队主页: <https://www.urbancomp.net>

