VGI大数据和知识图谱及人类活动预测

姚尧 博士，副教授

地理与信息工程学院，地图制图学与地理信息工程

阿里巴巴集团，达摩院，访问学者

Email: yaoy@cug.edu.cn

办公地点：未来城校区地信楼522办公室

主要内容

- 1 VGI大数据简介
- 2 VGI大数据关键技术
- 3 POI向量化—— POI2Vec
- 4 轨迹向量化—— Traj2Vec
- 5 基于知识图谱的城市居民活动推断
- 6 OSM与中国城市交通布局评价

主要内容

- 1 VGI大数据简介
- 2 VGI大数据关键技术
- 3 POI向量化—— POI2Vec
- 4 轨迹向量化—— Traj2Vec
- 5 基于知识图谱的城市居民活动推断
- 6 OSM与中国城市交通布局评价

01 VGI大数据简介

VGI（Volunteered Geographic Information）**自发式地理信息**——任何人都可以通过移动设备和浏览器自发贡献地理数据。

数据来源：

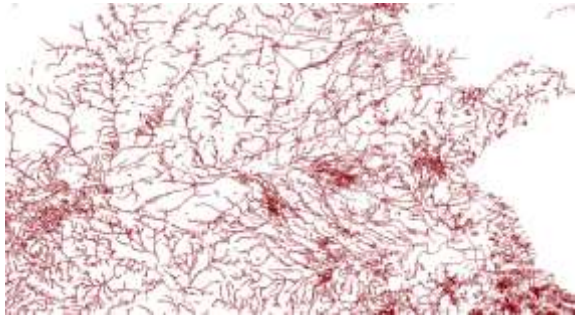
- VGI信息收集平台: OSM（openstreetmap）、 Wikimapia、Google Map Maker

例：

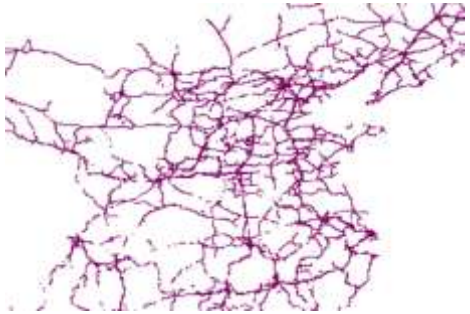
OSM

道路

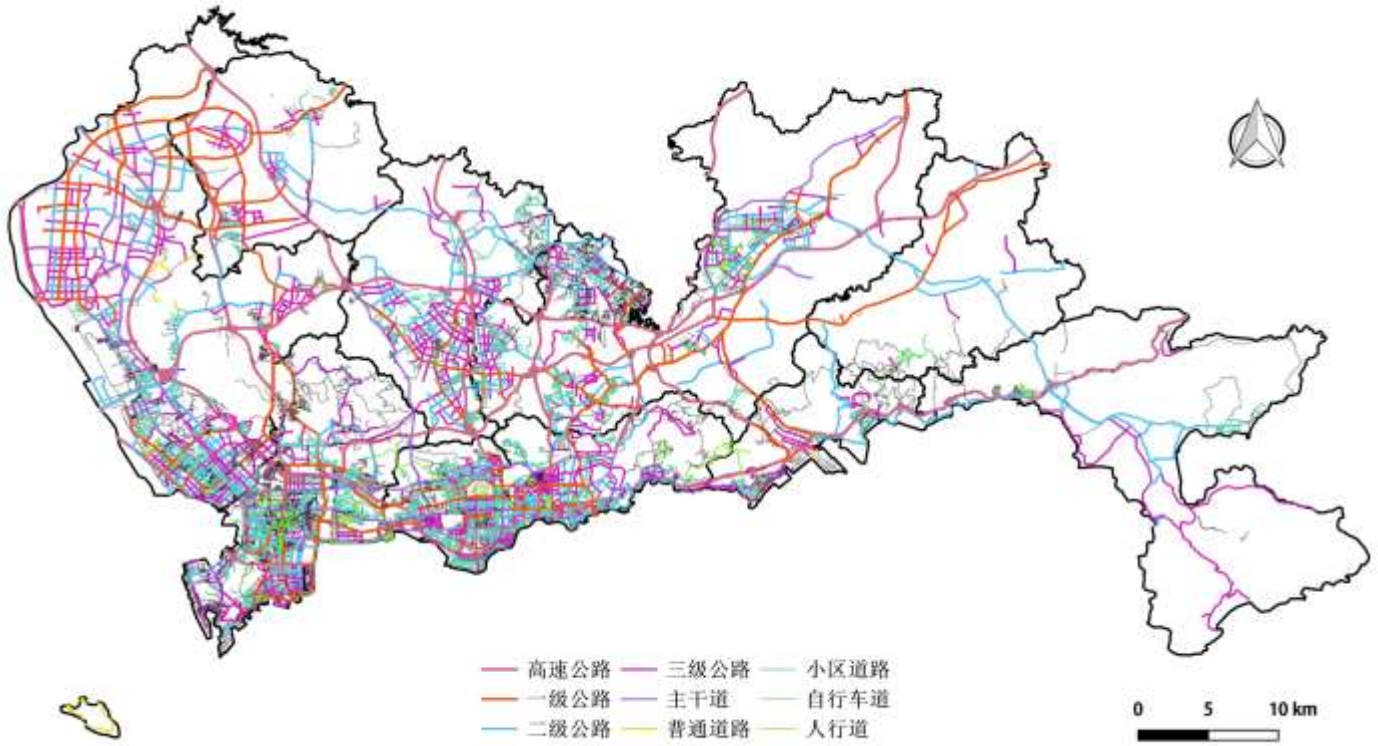
水系



铁路



01 VGI大数据简介



OpenStreetMap(OSM）路网数据来源：

http://www.openstreetmap.org

是由大众共同打造的免费开发和可编辑的地图服务

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

01 VGI大数据简介

■ OSM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

01 VGI大数据简介

- 除专门的VGI信息收集平台外，**带有地理坐标的社交媒体数据**也被认为是VGI数据，且社交媒体数据中包含大量文本、图片等，使得信息更加丰富

01 VGI大数据简介

- 以美团数据为例：

社交媒体数据特点：易获取、数据量大；数据源多样；实时更新等

01 手机信令大数据简介

VGI数据在城市计算中有什么优势?

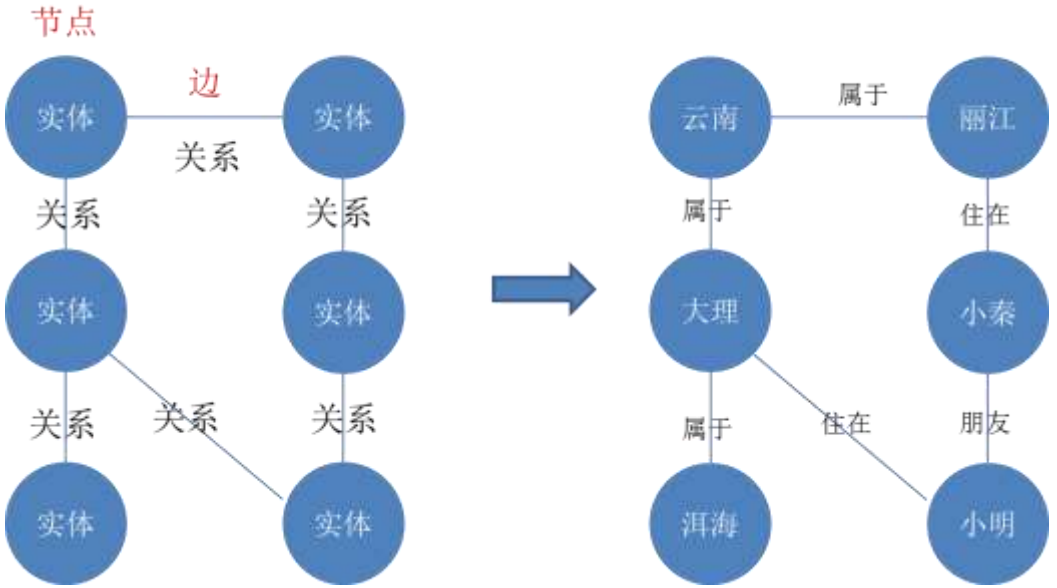
- ✓ 缩短了地理信息创建和传播的时间，降低了普通大众获取、分享和处理地理信息的门槛，**数据量大，易获取**；
- ✓ 补充地理框架数据的不足，为更好的描述现实世界提供丰富的**细节**；
- ✓ 具有随时编辑性，能尽可能的保证**现势性**。

主要内容

- 1 VGI大数据简介
- 2 VGI大数据关键技术
- 3 POI向量化—— POI2Vec
- 4 轨迹向量化—— Traj2Vec
- 5 基于知识图谱的城市居民活动推断
- 6 OSM与中国城市交通布局评价

知识图谱

知识图谱是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。

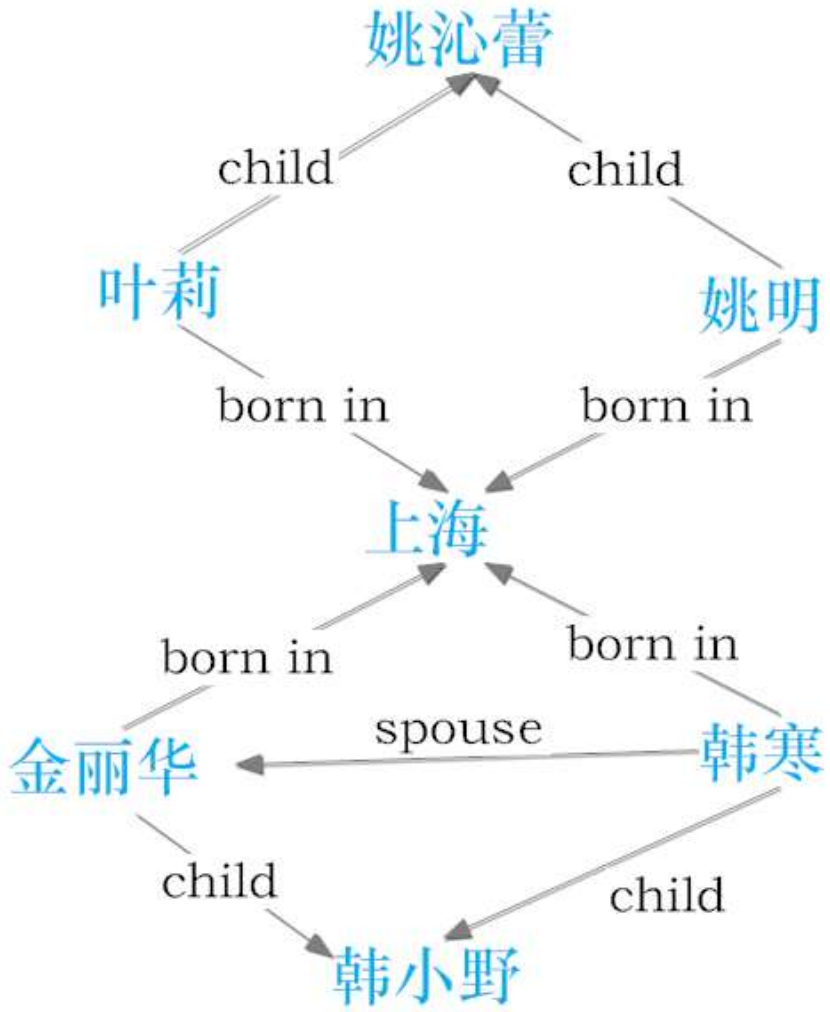


知识图谱

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

当前知识图谱中包含的**主要几种节点**有：

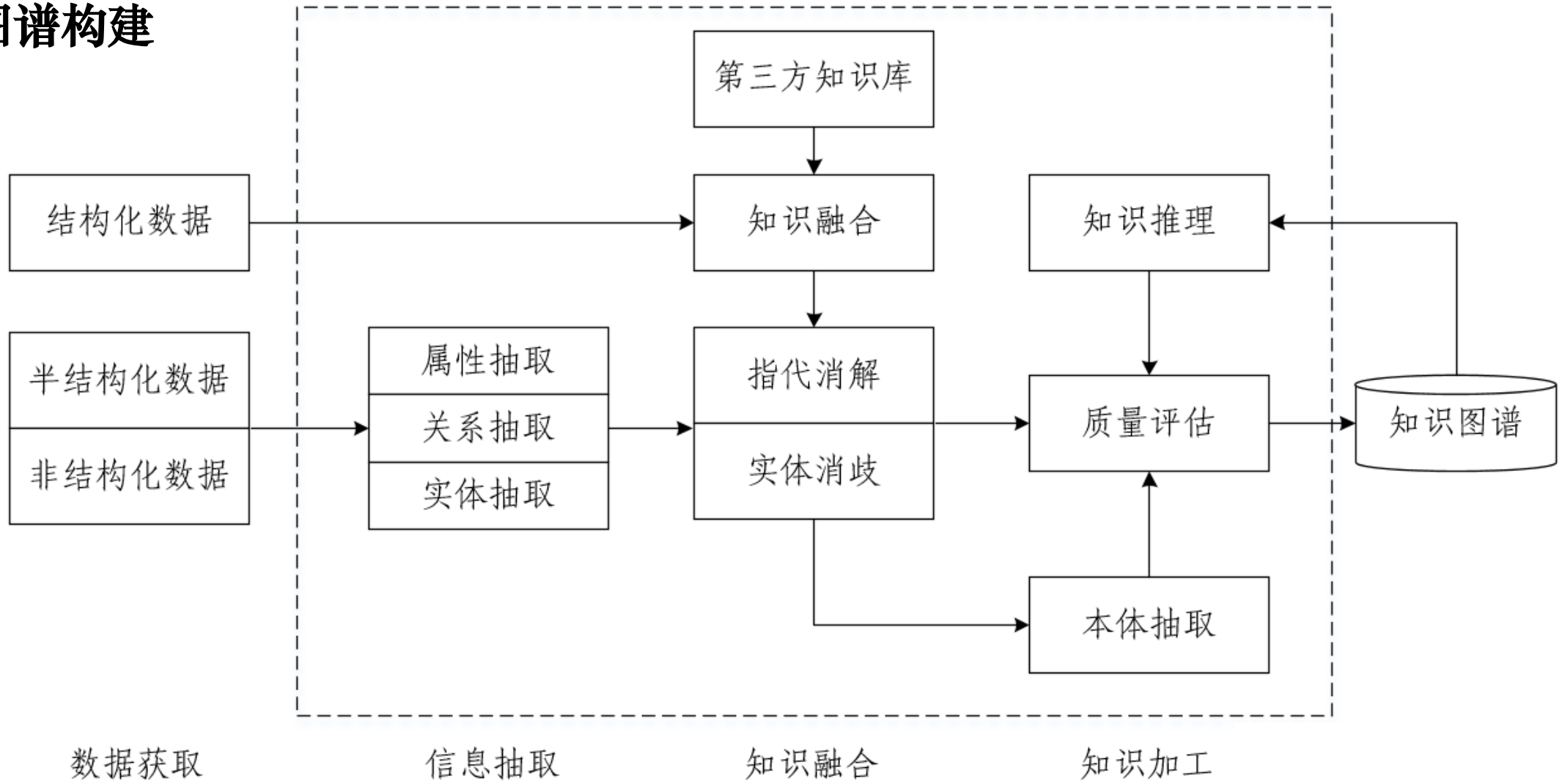
实体：指的是具有可区别性且独立存在的某种事物。如某一个人、某一座城市、某一种植物、某一件商品等等。实体是知识图谱中的最基本元素，不同的实体间存在不同的关系。

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。

02 VGI大数据关键技术

知识图谱构建



文本数据处理

例：数据中原始文本——潮汕秘制嫩牛腩非常好吃

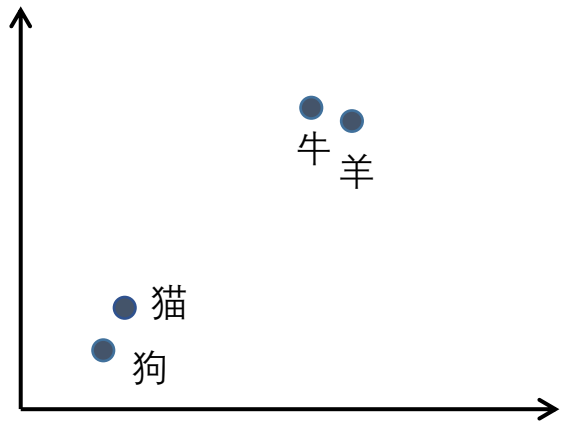
词语是文本的基本组成单元

1. 对文本进行分词：潮汕/秘制/嫩/牛腩/非常/好吃（**不可计算**）

2. 文本特征提取：将词语映射到高维向量空间中——通过**word2vec**进行词嵌入（**可计算**）

Word2vec

若两个词语词义相近，则二者在向量空间上也相近



对词向量进行计算时，有一个有趣的现象如下所示：

vector('queen') ~= vector('king') - vector('man') + vector('woman')

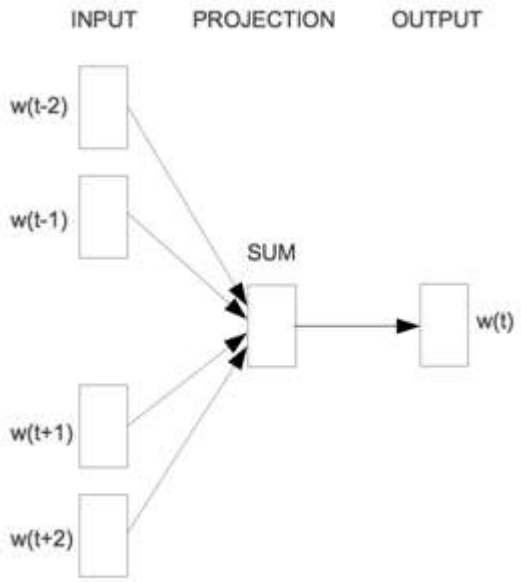
（女王~=国王-男人+女人）

那么，如何得到word2vec呢？

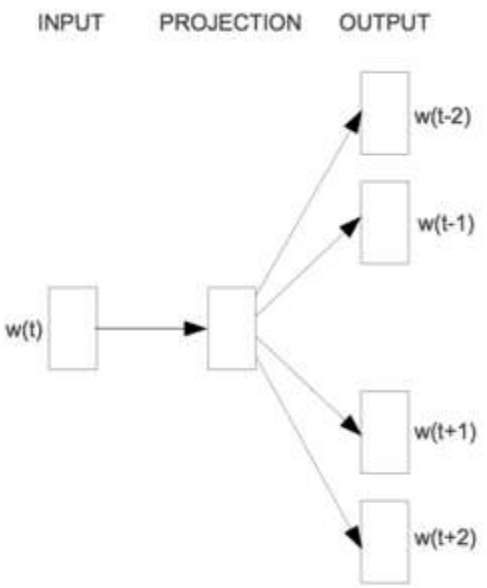
word2vec

常采用两种模型构建word2vec：**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



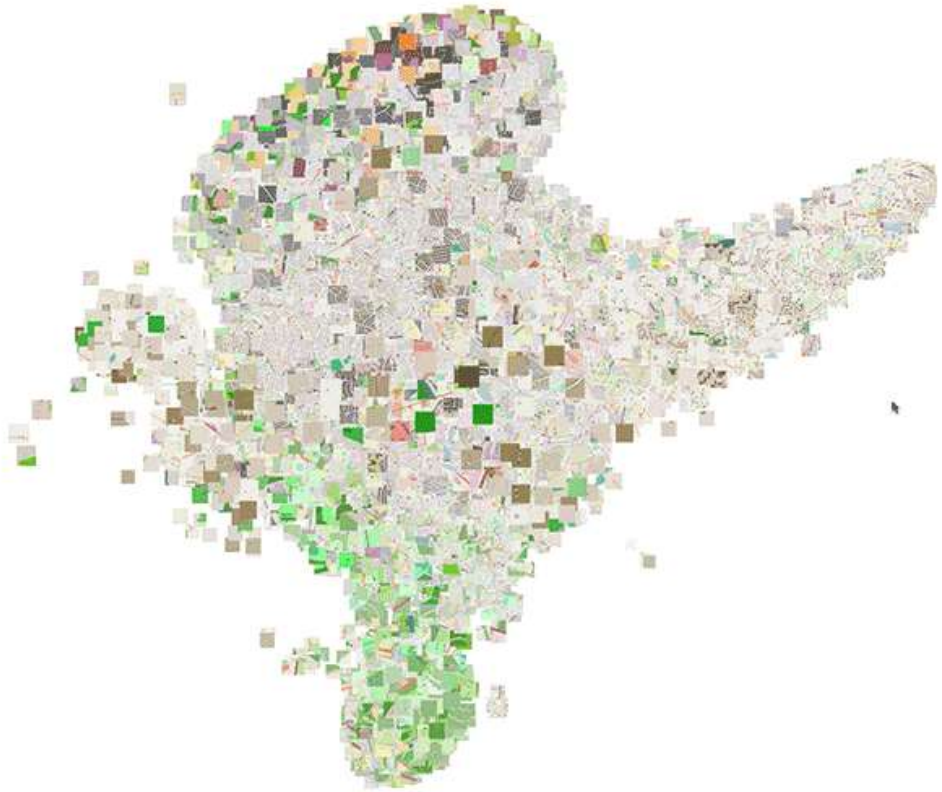
CBOW

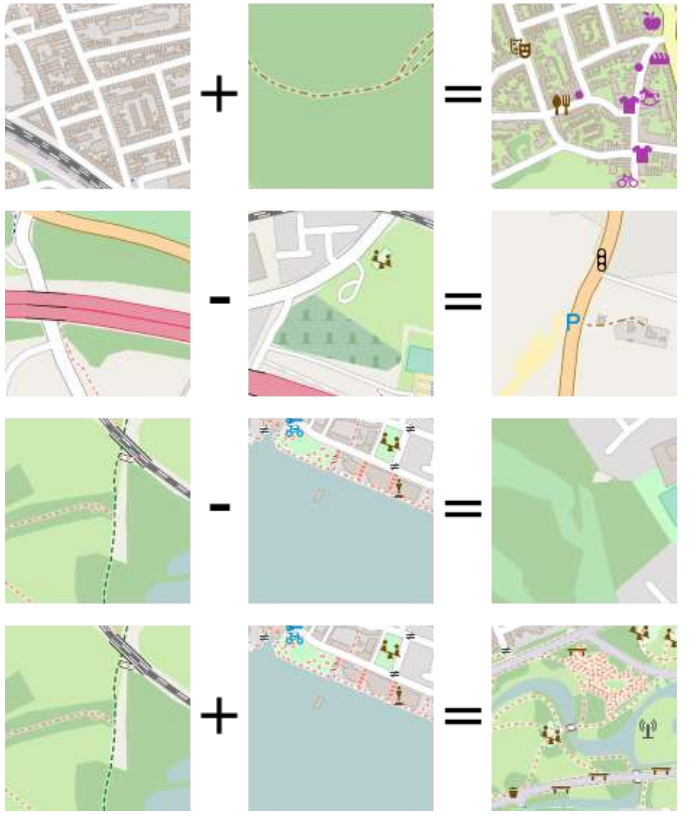


Skip-gram

Word2vec衍生

Word2vec思想在地理信息研究中也得到充分利用，如：POI2vec、Traj2Vec、StreetView2Vec、 Loc2Vec……





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。

  - 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)

主题模型

- **潜在语义分析(LSA)模型**

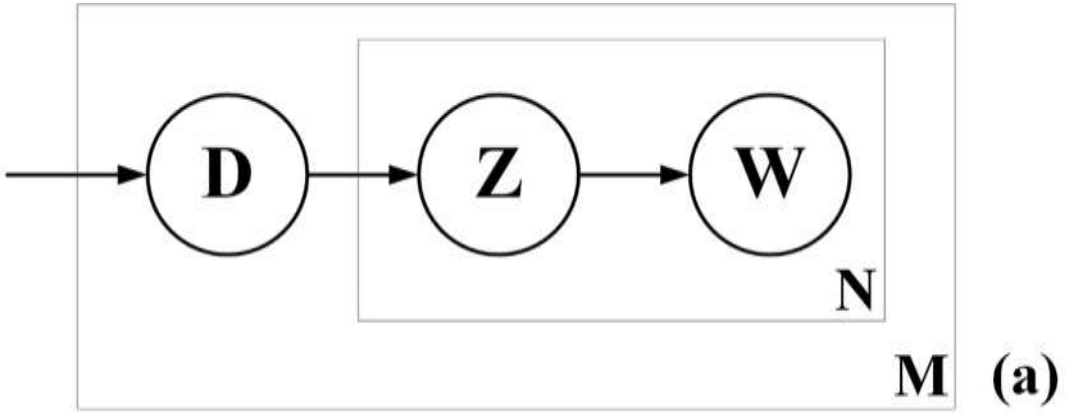
  是一种NLP技术，其核心思想是把我们所拥有的文档-术语矩阵分解成相互独立的文档-主题矩阵和主题-术语矩阵。通过将高维向量映射到潜在语义空间，提取与文档和词项有关的concepts，从而分析文档和词项之间的关系。LSA基于奇异值分解（SVD）的方法得到文档的主题，解决了传统向量空间模型无法处理一词多义或多次同义的问题，LSA 方法快速且高效，但它也有一些主要缺点：

  - 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

主题模型

- **概率潜在语义分析(PLSA)模型**

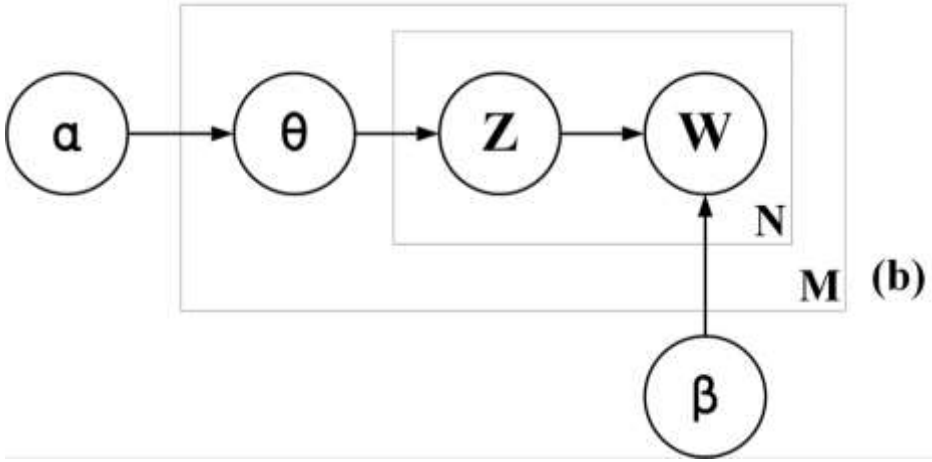
概率潜在语义分析，与LSA以共现表（矩阵w,d）的奇异值分解形式实现不同，pLSA是基于派生自LCM的混合矩阵分解，属于概率图模型中的生成模型，该类模型还有一元模型、混合一元模型等。其核心思想是找到一个潜在主题的概率模型，该模型可以生成在文档-术语矩阵中观察到的数据。是基于双模式和共现的数据分析方法延伸的经典统计学方法。



主题模型

- **潜在狄利克雷分配(LDA)模型**

  LDA 即潜在狄利克雷分布，是 pLSA 的贝叶斯版本。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。**本质上**，它回答了这样一个问题：“给定某种分布，我看到的实际概率分布可能是什么样子？”通过使用 LDA，可以从文档语料库中提取人类可解释的主题，其中每个主题都以与之关联度最高的词语作为特征。



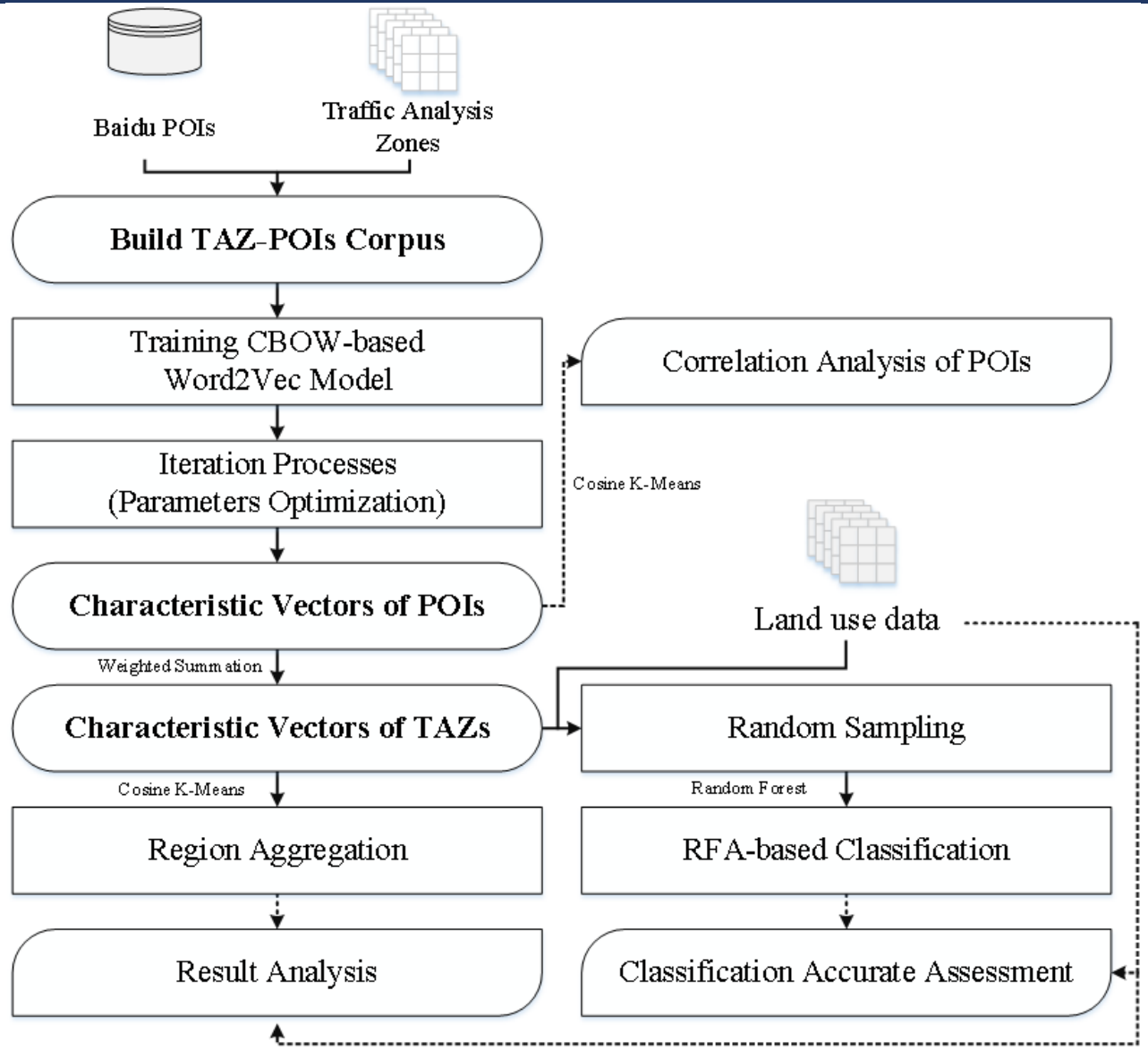
主要内容

- 1 VGI大数据简介
- 2 VGI大数据关键技术
- 3 POI向量化—— POI2Vec
- 4 轨迹向量化—— Traj2Vec
- 5 基于知识图谱的城市居民活动推断
- 6 OSM与中国城市交通布局评价

03 POI向量化—— POI2Vec

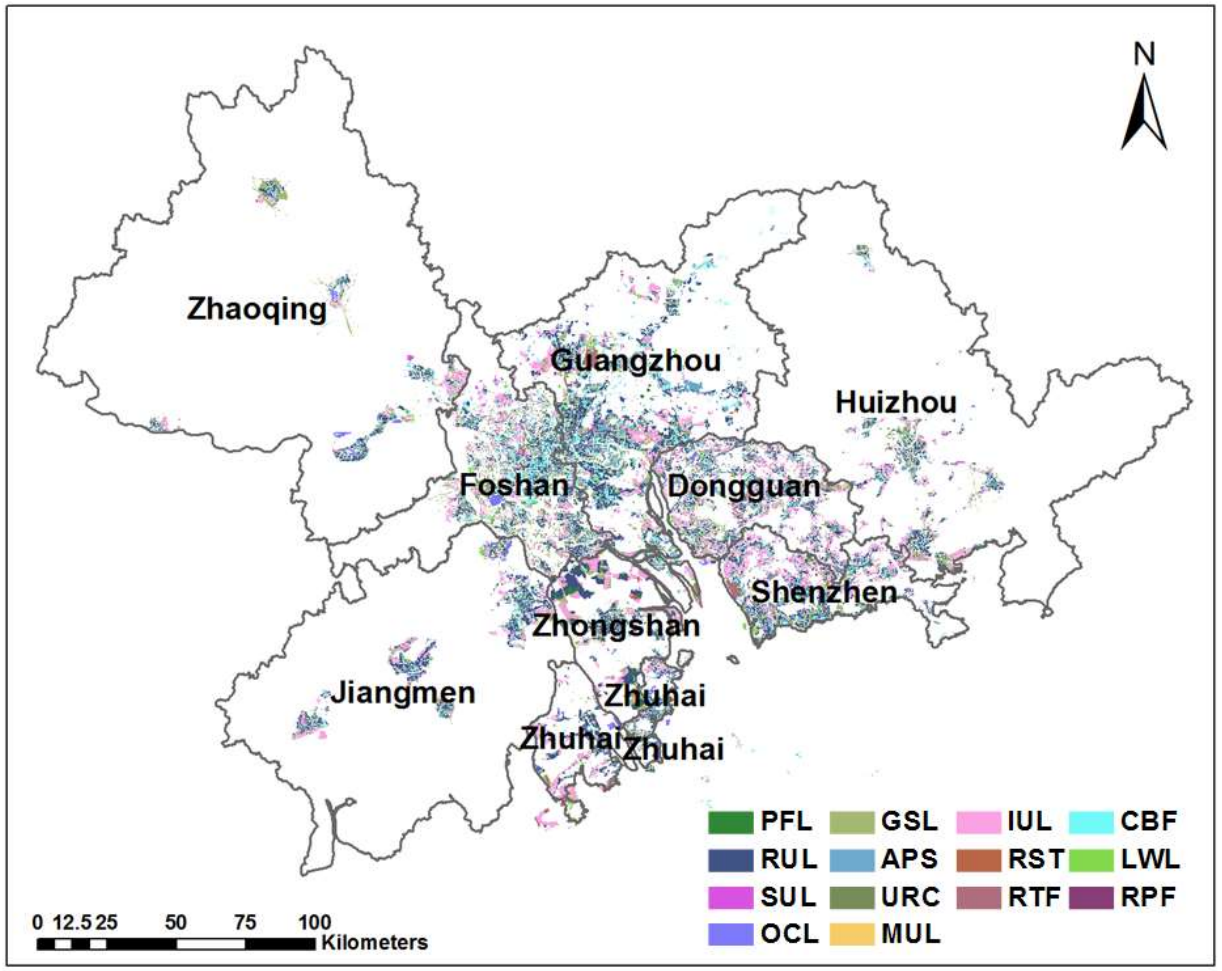
- 如何将POI转换为机器可以计算的向量形式，便于对其进行研究分析？

  ——POI2Vec



03 POI向量化—— POI2Vec

➢ 研究区域

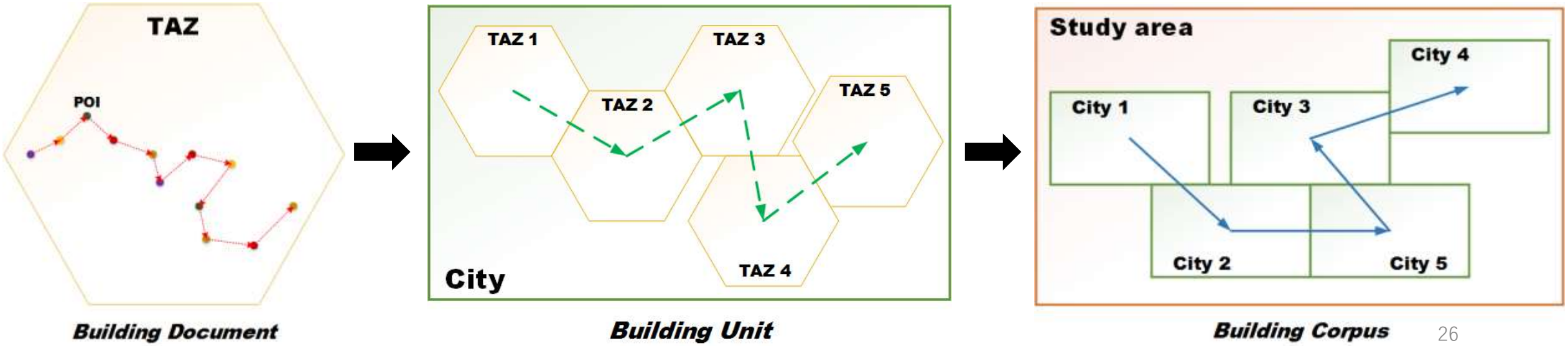


ID	Short Code	Land-use type	Proportions
1	PFL	Public facilities	3.97%
2	GSL	Green space and square	37.31%
3	IUL	Industrial land	8.19%
4	CBF	Commercial and business facilities	13.23%
5	RUL	Residential land	20.34%
6	APS	Administration and public services	7.52%
7	RST	Road, street and transportation	1.68%
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)

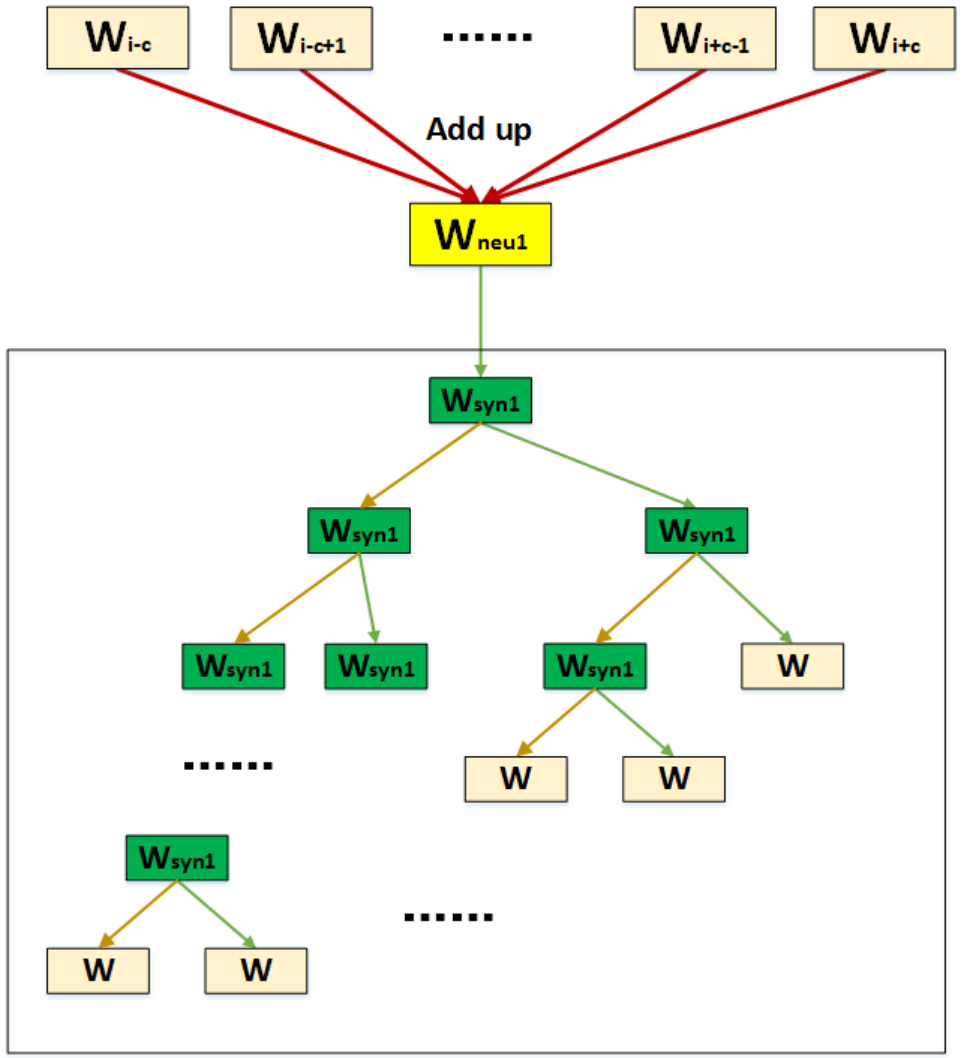
03 POI向量化—— POI2Vec

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



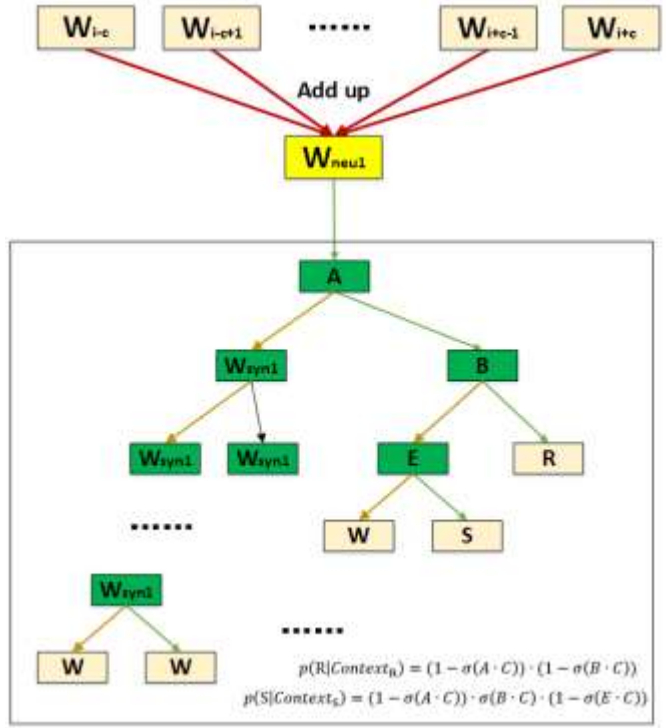
Building Document *Building Unit* *Building Corpus*

03 POI向量化—— POI2Vec



Hoffman Trees

* Number of POIs

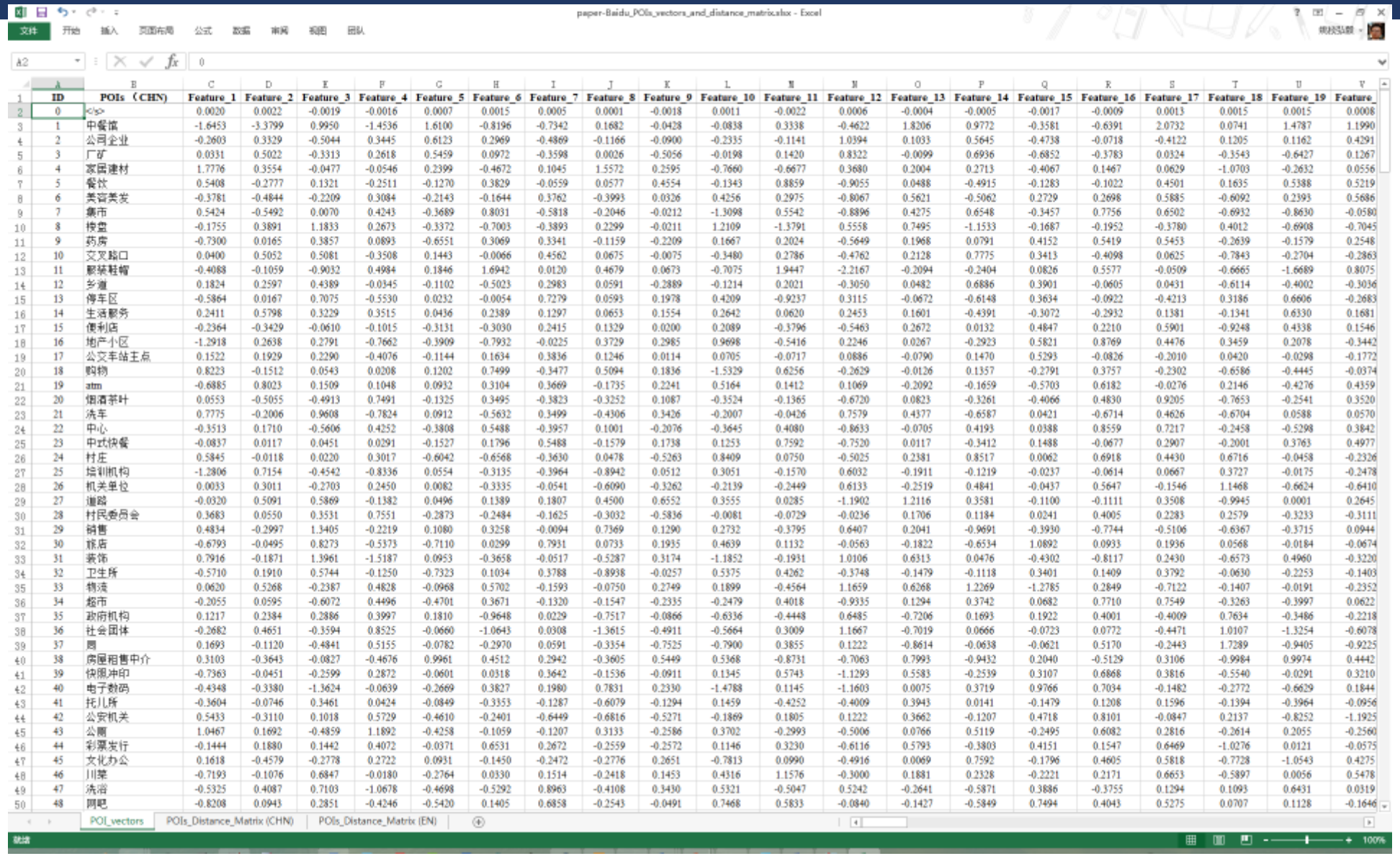


Maximum likelihood function

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

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

03 POI向量化—— POI2Vec

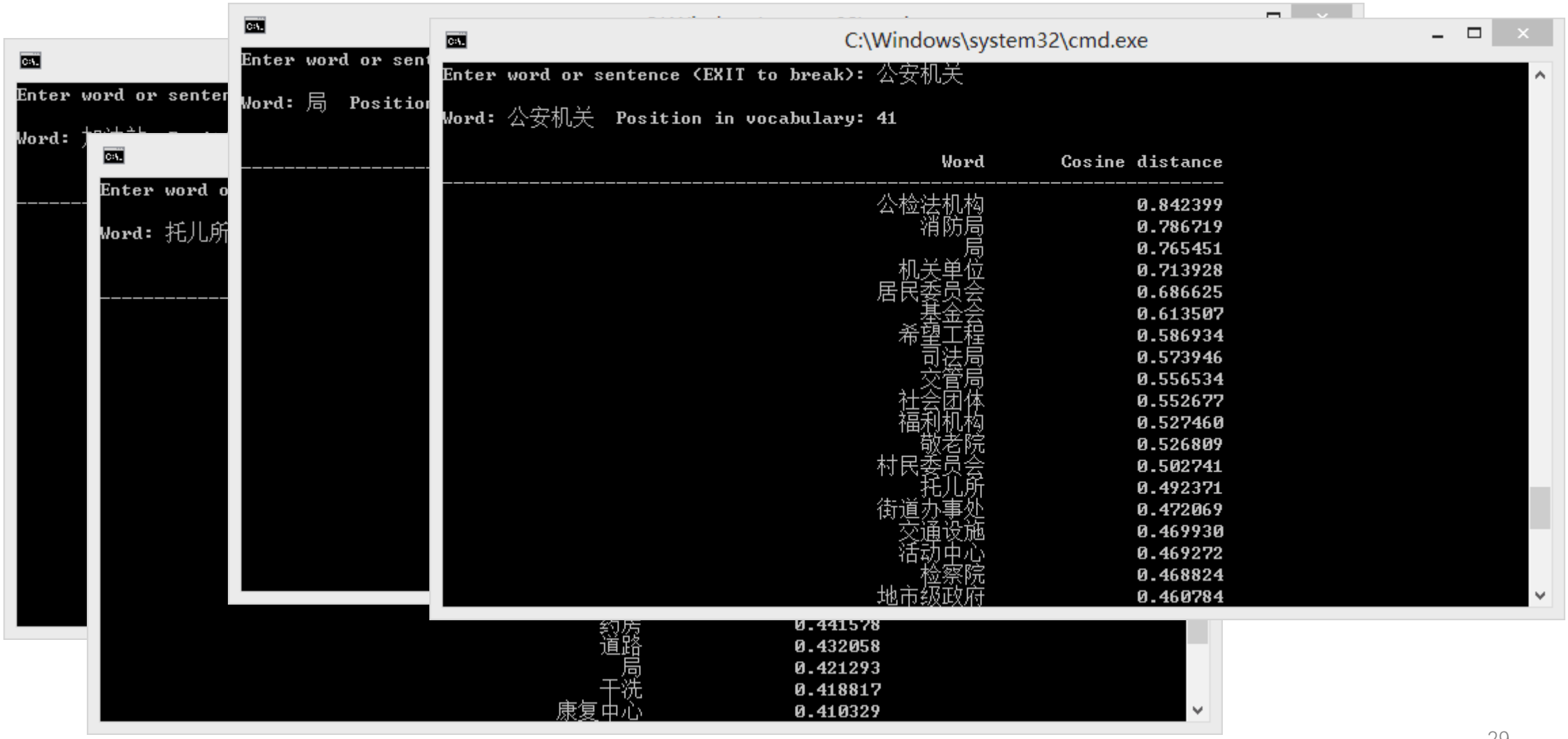


paper-Baidu_POIs_vectors_and_distance_matrix.xlsx - Excel

ID	POIs（CHN）	Feature_1	Feature_2	Feature_3	Feature_4	Feature_5	Feature_6	Feature_7	Feature_8	Feature_9	Feature_10	Feature_11	Feature_12	Feature_13	Feature_14	Feature_15	Feature_16	Feature_17	Feature_18	Feature_19	Feature_
0	</s>	0.0020	0.0022	-0.0019	-0.0016	0.0007	0.0015	0.0005	0.0001	-0.0018	0.0011	-0.0022	0.0006	-0.0004	-0.0005	-0.0017	-0.0009	0.0013	0.0015	0.0015	0.0008
1	中餐馆	-1.6453	-3.3799	0.9950	-1.4536	1.6100	-0.8196	-0.7342	0.1682	-0.0428	-0.0838	0.3338	-0.4622	1.8206	0.9772	-0.3581	-0.6391	2.0732	0.0741	1.4787	1.1990
2	公司企业	-0.2603	0.3329	-0.5044	0.3445	0.6123	0.2969	-0.4869	-0.1166	-0.0900	-0.2335	-0.1141	1.0394	0.1033	0.5645	-0.4738	-0.0718	-0.4122	0.1205	0.1162	0.4291
3	厂矿	0.0331	0.5022	-0.3313	0.2618	0.5459	0.0972	-0.3598	0.0026	-0.5056	-0.0198	0.1420	0.8322	-0.0099	0.6936	-0.6852	-0.3783	0.0324	-0.3543	-0.6427	0.1267
4	家居建材	1.7776	0.3554	-0.0477	-0.0546	0.2399	-0.4672	0.1045	1.5572	0.2595	-0.7660	-0.6677	0.3680	0.2004	0.2713	-0.4067	0.1467	0.0629	-1.0703	-0.2632	0.0556
5	餐饮	0.5408	-0.2777	0.1321	-0.2511	-0.1270	0.3829	-0.0559	0.0577	0.4554	-0.1343	0.8859	-0.9055	0.0488	-0.4915	-0.1283	-0.1022	0.4501	0.1635	0.5388	0.5219
6	美容美发	-0.3781	-0.4844	-0.2209	0.3084	-0.2143	-0.1644	0.3762	-0.3993	0.0326	0.4256	0.2975	-0.8067	0.5621	-0.5062	0.2729	0.2698	0.5885	-0.6092	0.2393	0.5686
7	集市	0.5424	-0.5492	0.0070	0.4243	-0.3689	0.8031	-0.5818	-0.2046	-0.0212	-1.3098	0.5542	-0.8896	0.4275	0.6548	-0.3457	0.7756	0.6502	-0.6932	-0.8630	-0.0580
8	按盘	-0.1755	0.3891	1.1833	0.2673	-0.3372	-0.7003	-0.3893	0.2299	-0.0211	1.2109	-1.3791	0.5558	0.7495	-1.1533	-0.1687	-0.1952	-0.3780	0.4012	-0.6908	-0.7045
9	药房	-0.7300	0.0165	0.3857	0.0893	-0.6551	0.3069	0.3341	-0.1159	-0.2209	0.1667	0.2024	-0.5649	0.1968	0.0791	0.4152	0.5419	0.5453	-0.2639	-0.1579	0.2548
10	交叉路口	0.0400	0.5052	0.5081	-0.3508	0.1443	-0.0066	0.4562	0.0675	-0.0075	-0.3480	0.2786	-0.4762	0.2128	0.7775	0.3413	-0.4098	0.0625	-0.7843	-0.2704	-0.2863
11	服装鞋帽	-0.4088	-0.1059	-0.9032	0.4984	0.1846	1.6942	0.0120	0.4679	0.0673	-0.7075	1.9447	-2.2167	-0.2094	-0.2404	0.0826	0.5577	-0.0509	-0.6665	-1.6689	0.8075
12	乡道	0.1824	0.2597	0.4389	-0.0345	-0.1102	-0.5023	0.2983	0.0591	-0.2889	-0.1214	0.2021	-0.3050	0.0482	0.6886	0.3901	-0.0605	0.0431	-0.6114	-0.4002	-0.3036
13	停车区	-0.5864	0.0167	0.7075	-0.5530	0.0232	-0.0054	0.7279	0.0593	0.1978	0.4209	-0.9237	0.3115	-0.0672	-0.6148	0.3634	-0.0922	-0.4213	0.3186	0.6606	-0.2683
14	生活服务	0.2411	0.5798	0.3229	0.3515	0.0436	0.2389	0.1297	0.0653	0.1554	0.2642	0.0620	0.2453	0.1601	-0.4391	-0.3072	-0.2932	0.1381	-0.1341	0.6330	0.1681
15	便利店	-0.2364	-0.3429	-0.0610	-0.1015	-0.3131	-0.3030	0.2415	0.1329	0.0200	0.2089	-0.3796	-0.5463	0.2672	0.0132	0.4847	0.2210	0.5901	-0.9248	0.4338	0.1546
16	地产小区	-1.2918	0.2638	0.2791	-0.7662	-0.3909	-0.7932	-0.0225	0.3729	0.2985	0.9698	-0.5416	0.2246	0.0267	-0.2923	0.5821	0.8769	0.4476	0.3459	0.2078	-0.3442
17	公交车站主点	0.1522	0.1929	0.2290	-0.4076	-0.1144	0.1634	0.3836	0.1246	0.0114	0.0705	-0.0717	0.0886	-0.0790	0.1470	0.5293	-0.0826	-0.2010	0.0420	-0.0298	-0.1772
18	购物	0.8223	-0.1512	0.0543	0.0208	0.1202	0.7499	-0.3477	0.5094	0.1836	-1.5329	0.6256	-0.2629	-0.0126	0.1357	-0.2791	0.3757	-0.2302	-0.6586	-0.4445	-0.0374
19	atm	-0.6885	0.8023	0.1509	0.1048	0.0932	0.3104	0.3669	-0.1735	0.2241	0.5164	0.1412	0.1069	-0.2092	-0.1659	-0.5703	0.6182	-0.0276	0.2146	-0.4276	0.4359
20	烟酒茶叶	0.0553	-0.5055	-0.4913	0.7491	-0.1325	0.3495	-0.3823	-0.3252	0.1087	-0.3524	-0.1365	-0.6720	0.0823	-0.3261	-0.4066	0.4830	0.9205	-0.7653	-0.2541	0.3520
21	洗车	0.7775	-0.2006	0.9608	-0.7824	0.0912	-0.5632	0.3499	-0.4306	0.3426	-0.2007	-0.0426	0.7579	0.4377	-0.6587	0.0421	-0.6714	0.4626	-0.6704	0.0588	0.0570
22	中心	-0.3513	0.1710	-0.5606	0.4252	-0.3808	0.5488	-0.3957	0.1001	-0.2076	-0.3645	0.4080	-0.8633	-0.0705	0.4193	0.0388	0.8559	0.7217	-0.2458	-0.5298	0.3842
23	中式快餐	-0.0837	0.0117	0.0451	0.0291	-0.1527	0.1796	0.5488	-0.1579	0.1738	0.1253	0.7592	-0.7520	0.0117	-0.3412	0.1488	-0.0677	0.2907	-0.2001	0.3763	0.4977
24	村庄	0.5845	-0.0118	0.0220	0.3017	-0.6042	-0.6568	-0.3630	0.0478	-0.5263	0.8409	0.0750	-0.5025	0.2381	0.8517	0.0062	0.6918	0.4430	0.6716	-0.0458	-0.2326
25	培训机构	-1.2806	0.7154	-0.4542	-0.8336	0.0554	-0.3135	-0.3964	-0.8942	0.0512	0.3051	-0.1570	0.6032	-0.1911	-0.1219	-0.0237	-0.0614	0.0667	0.3727	-0.0175	-0.2478
26	机关单位	0.0033	0.3011	-0.2703	0.2450	0.0082	-0.3335	-0.0541	-0.6090	-0.3262	-0.2139	-0.2449	0.6133	-0.2519	0.4841	-0.0437	0.5647	-0.1546	1.1468	-0.6624	-0.6410
27	道路	-0.0320	0.5091	0.5869	-0.1382	0.0496	0.1389	0.1807	0.4500	0.6552	0.3555	0.0285	-1.1902	1.2116	0.3581	-0.1100	-0.1111	0.3508	-0.9945	0.0001	0.2645
28	村民委员会	0.3683	0.0550	0.3531	0.7551	-0.2873	-0.2484	-0.1625	-0.3032	-0.5836	-0.0081	-0.0729	-0.0236	0.1706	0.1184	0.0241	0.4005	0.2283	0.2579	-0.3233	-0.3111
29	销售	0.4834	-0.2997	1.3405	-0.2219	0.1080	0.3258	-0.0094	0.7369	0.1290	0.2732	-0.3795	0.6407	0.2041	-0.9691	-0.3930	-0.7744	-0.5106	-0.6367	-0.3715	0.0944
30	旅店	-0.6793	-0.0495	0.8273	-0.5373	-0.7110	0.0299	0.7931	0.0733	0.1935	0.4639	0.1132	-0.0563	-0.1822	-0.6534	1.0892	0.0933	0.1936	0.0568	-0.0184	-0.0674
31	装饰	0.7916	-0.1871	1.3961	-1.5187	0.0953	-0.3658	-0.0517	-0.5287	0.3174	-1.1852	-0.1931	1.0106	0.6313	0.0476	-0.4302	-0.8117	0.2430	-0.6573	0.4960	-0.3220
32	卫生所	-0.5710	0.1910	0.5744	-0.1250	-0.7323	0.1034	0.3788	-0.8938	-0.0257	0.5375	0.4262	-0.3748	-0.1479	-0.1118	0.3401	0.1409	0.3792	-0.0630	-0.2253	-0.1403
33	物流	0.0620	0.5268	-0.2387	0.4828	-0.0968	0.5702	-0.1593	-0.0750	0.2749	0.1899	-0.4564	1.1659	0.6268	1.2269	-1.2785	0.2849	-0.7122	-0.1407	-0.0191	-0.2352
34	超市	-0.2055	0.0595	-0.6072	0.4496	-0.4701	0.3671	-0.1320	-0.1547	-0.2335	-0.2479	0.4018	-0.9335	0.1294	0.3742	0.0682	0.7710	0.7549	-0.3263	-0.3997	0.0622
35	政府机构	0.1217	0.2384	0.2886	0.3997	0.1810	-0.9648	0.0229	-0.7517	-0.0866	-0.6336	-0.4448	0.6485	-0.7206	0.1693	0.1922	0.4001	-0.4009	0.7634	-0.3486	-0.2218
36	社会团体	-0.2682	0.4651	-0.3594	0.8525	-0.0660	-1.0643	0.0308	-1.3615	-0.4911	-0.5664	0.3009	1.1667	-0.7019	0.0666	-0.0723	0.0772	-0.4471	1.0107	-1.3254	-0.6078
37	厕	0.1693	-0.1120	-0.4841	0.5155	-0.0782	-0.2970	0.0591	-0.3354	-0.7525	-0.7900	0.3855	0.1222	-0.8614	-0.0638	-0.0621	0.5170	-0.2443	1.7289	-0.9405	-0.9225
38	房屋租售中介	0.3103	-0.3643	-0.0827	-0.4676	0.9961	0.4512	0.2942	-0.3605	0.5449	0.5368	-0.8731	-0.7063	0.7993	-0.9432	0.2040	-0.5129	0.3106	-0.9984	0.9974	0.4442
39	快照冲印	-0.7363	-0.0451	-0.2599	0.2872	-0.0601	0.0318	0.3642	-0.1536	-0.0911	0.1345	0.5743	-1.1293	0.5583	-0.2539	0.3107	0.6868	0.3816	-0.5540	-0.0291	0.3210
40	电子数码	-0.4348	-0.3380	-1.3624	-0.0639	-0.2669	0.3827	0.1980	0.7831	0.2330	-1.4788	0.1145	-1.1603	0.0075	0.3719	0.9766	0.7034	-0.1482	-0.2772	-0.6629	0.1844
41	托儿所	-0.3604	-0.0746	0.3461	0.0424	-0.0849	-0.3353	-0.1287	-0.6079	-0.1294	0.1459	-0.4252	-0.4009	0.3943	0.0141	-0.1479	0.1208	0.1596	-0.1394	-0.3964	-0.0956
42	公安机关	0.5433	-0.3110	0.1018	0.5729	-0.4610	-0.2401	-0.6449	-0.6816	-0.5271	-0.1869	0.1805	0.1222	0.3662	-0.1207	0.4718	0.8101	-0.0847	0.2137	-0.8252	-1.1925
43	公厕	1.0467	0.1692	-0.4859	1.1892	-0.4258	-0.1059	-0.1207	0.3133	-0.2586	0.3702	-0.2993	-0.5006	0.0766	0.5119	-0.2495	0.6082	0.2816	-0.2614	0.2055	-0.2560
44	彩票发行	-0.1444	0.1880	0.1442	0.4072	-0.0371	0.6531	0.2672	-0.2559	-0.2572	0.1146	0.3230	-0.6116	0.5793	-0.3803	0.4151	0.1547	0.6469	-1.0276	0.0121	-0.0575
45	文化办公	0.1618	-0.4579	-0.2778	0.2722	0.0931	-0.1450	-0.2472	-0.2776	0.2651	-0.7813	0.0990	-0.4916	0.0069	0.7592	-0.1796	0.4605	0.5818	-0.7728	-1.0543	0.4275
46	川菜	-0.7193	-0.1076	0.6847	-0.0180	-0.2764	0.0330	0.1514	-0.2418	0.1453	0.4316	1.1576	-0.3000	0.1881	0.2328	-0.2221	0.2171	0.6653	-0.5897	0.0056	0.5478
47	洗浴	-0.5325	0.4087	0.7103	-1.0678	-0.4698	-0.5292	0.8963	-0.4108	0.3430	0.5321	-0.5047	0.5242	-0.2641	-0.5871	0.3886	-0.3755	0.1294	0.1093	0.6431	0.0319
48	网吧	-0.8208	0.0943	0.2851	-0.4246	-0.5420	0.1405	0.6858	-0.2543	-0.0491	0.7468	0.5833	-0.0840	-0.1427	-0.5849	0.7494	0.4043	0.5275	0.0707	0.1128	-0.1646

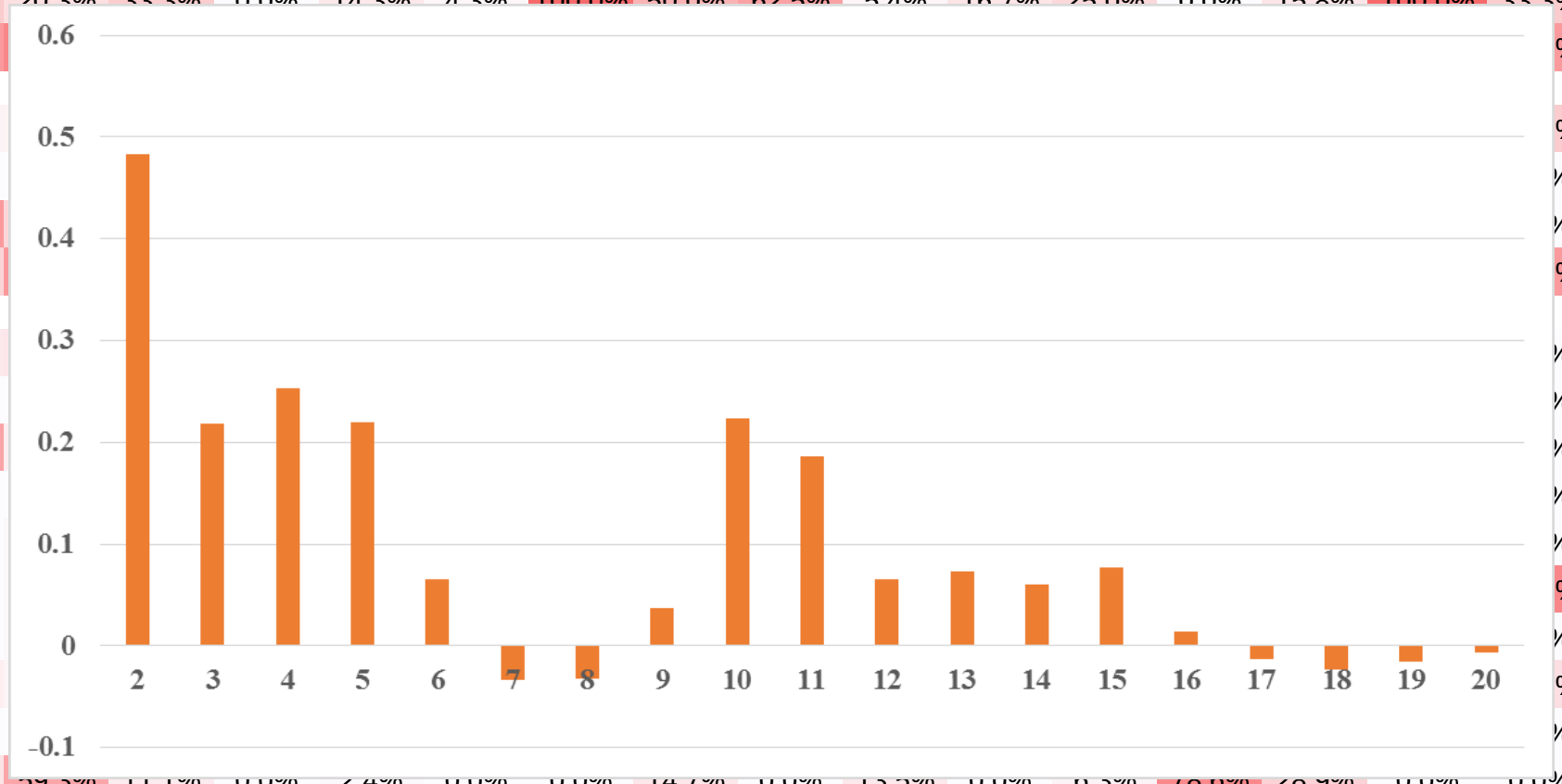
POI_vectors | POIs_Distance_Matrix (CHN) | POIs_Distance_Matrix (EN)

03 POI向量化—— POI2Vec



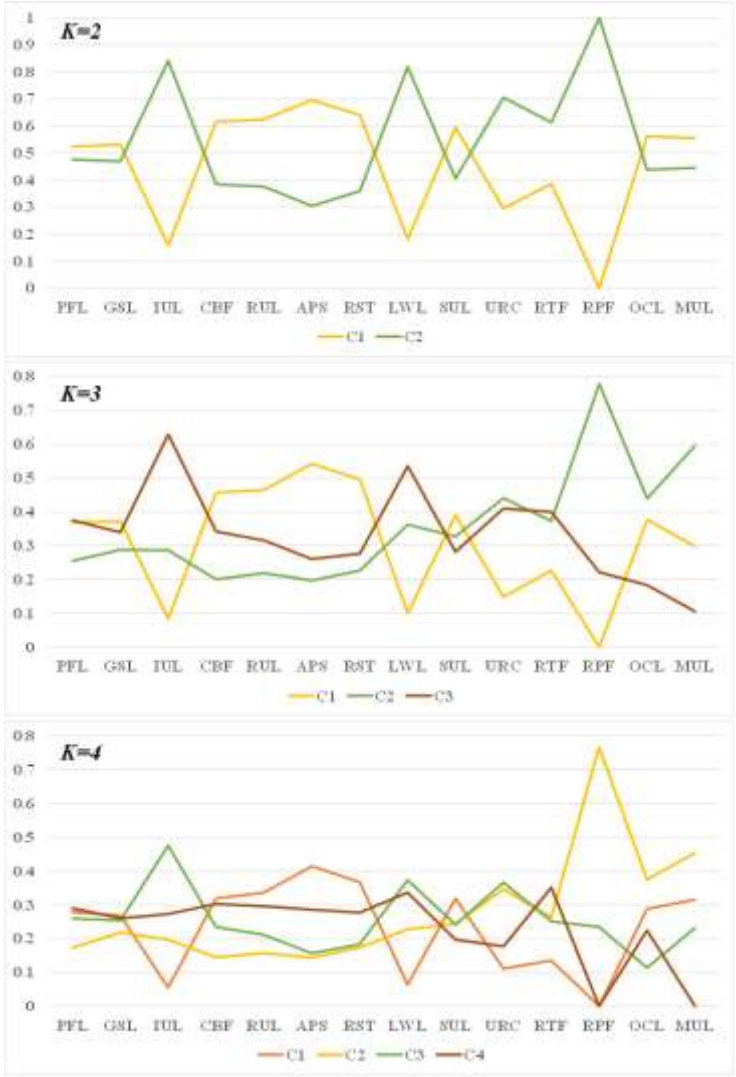
03 POI向量化—— POI2Vec

K	Class	COR	SHP	CAT	LIF	RSC	GOV	CLF	ROD	TRA	AMS	FII	ADL	EDU	HOT	ENT	LOC	BUB	MOU	SCE	GRE
2	C1	8.1%	33.3%	20.3%	33.3%	0.0%	14.3%	4.3%	100.0%	50.0%	62.5%	5.4%	16.7%	25.0%	0.0%	15.8%	100.0%	33.3%	100.0%	0.0%	50.0%
	C2	91.9%	66.7%															%	0.0%	100.0%	50.0%
4	C1	28.6%	6.7%															%	100.0%	0.0%	50.0%
	C2	21.4%	0.0%															%	0.0%	25.0%	0.0%
	C3	19.0%	66.7%															%	0.0%	0.0%	0.0%
	C4	31.0%	26.7%															%	0.0%	75.0%	50.0%
10	C1	11.9%	13.3%															%	100.0%	0.0%	0.0%
	C2	9.5%	0.0%															%	0.0%	0.0%	0.0%
	C3	14.3%	60.0%															%	0.0%	0.0%	0.0%
	C4	0.0%	0.0%															%	0.0%	8.3%	0.0%
	C5	2.4%	0.0%															%	0.0%	8.3%	0.0%
	C6	42.9%	0.0%															%	0.0%	16.7%	0.0%
	C7	7.1%	0.0%															%	0.0%	0.0%	0.0%
	C8	9.5%	10.0%															%	0.0%	25.0%	50.0%
	C9	2.4%	0.0%															%	0.0%	16.7%	0.0%
	C10	0.0%	16.7%	59.3%	11.1%	0.0%	2.4%	0.0%	0.0%	14.7%	0.0%	13.5%	0.0%	6.3%	78.6%	28.9%	0.0%	0.0%	0.0%	25.0%	50.0%



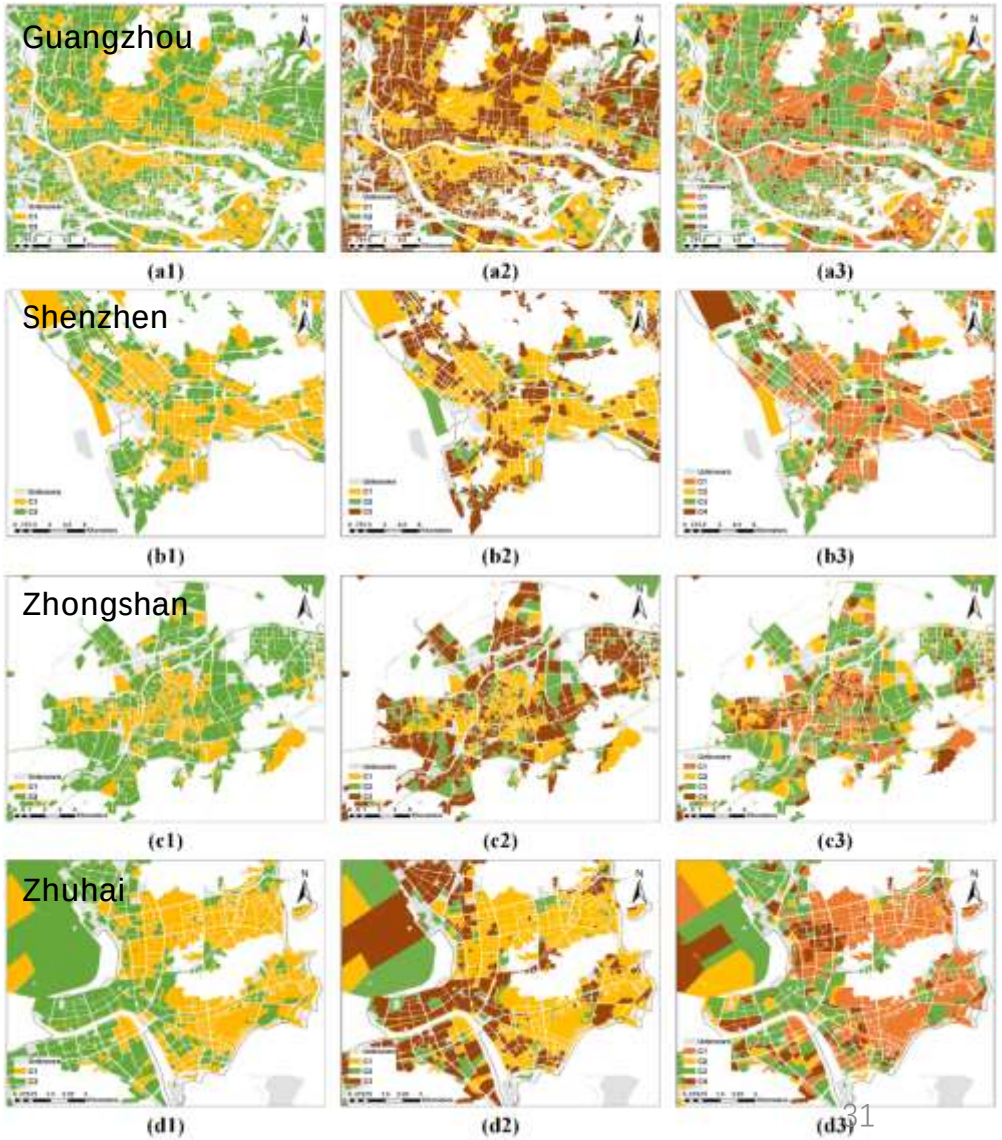
Types of POIs: COR = Corporation, SHP = Shopping, CAT = Catering, LIF = Life Service, RSC = Residential community, GOV = Government, CLF = Clinic facilitiy, ROD = Road, TRA = Traffic facility, AMS = Automobile service, FII = Financial industry, ADL = Administrative landmark, EDU = Education, HOT = Hotel,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

03 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)



03 POI向量化—— POI2Vec

Methods	Training Process		Predicting Process		Avg. Comp. Time
	OOB Avg. Error	OOB RMSE	OA	Kappa	Unit: seconds
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

TFIDF	PFL	GSL	IUL	CBF	RUL	APS	RST	LWL	SUL	URC	RTF	RPF	OCL	MUL
PFL	0.00	0.51	0.03	0.06	0.37	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
GSL	0.00	0.69	0.02	0.04	0.23	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
IUL	0.00	0.32	0.41	0.02	0.24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CBF	0.00	0.28	0.03	0.39	0.28	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RUL	0.00	0.20	0.02	0.03	0.74	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
APS	0.00	0.27	0.02	0.04	0.29	0.37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RST	0.00	0.39	0.03	0.04	0.24	0.02	0.27	0.00	0.00	0.00	0.01	0.00	0.00	0.00
LWL	0.01	0.30	0.08	0.03	0.30	0.00	0.00	0.27	0.00	0.01	0.01	0.00	0.00	0.00
SUL	0.01	0.21	0.04	0.04	0.37	0.01	0.00	0.00	0.32	0.00	0.00	0.00	0.00	0.00
URC	0.00	0.39	0.04	0.03	0.26	0.00	0.00	0.00	0.00	0.27	0.00	0.00	0.00	0.00
RTF	0.00	0.33	0.04	0.06	0.31	0.01	0.00	0.00	0.00	0.00	0.25	0.00	0.00	0.00
RPF	0.00	0.43	0.14	0.00	0.14	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.00
OCL	0.00	0.32	0.04	0.03	0.29	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.30	0.00
MUL	0.00	0.42	0.00	0.00	0.42	0.00	0.00	0.00	0.00	0.17	0.00	0.00	0.00	0.00

(b) TF-IDF

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

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

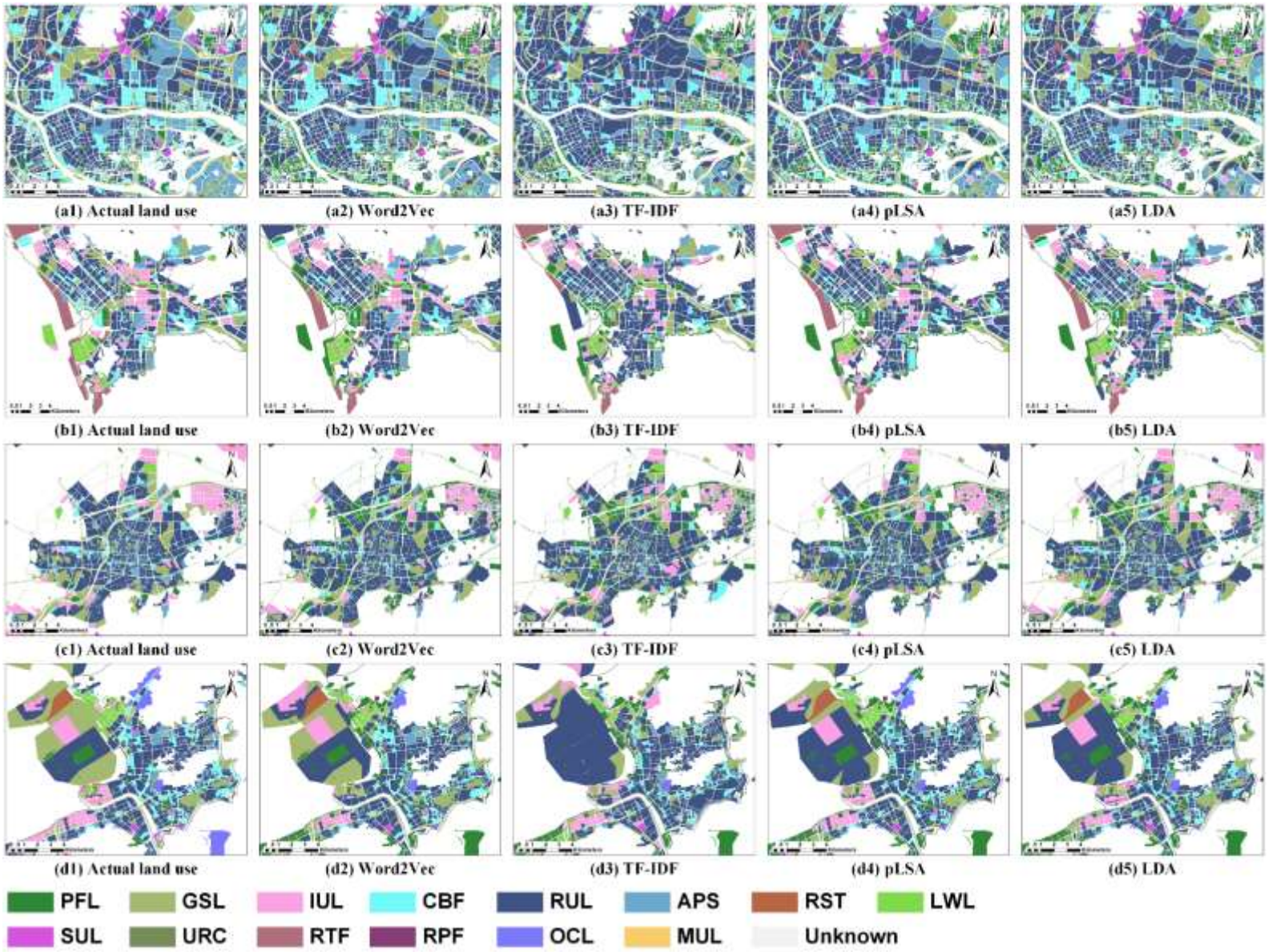
03 POI向量化—— POI2Vec

Guangzhou

Shenzhen

Zhongshan

Zhuhai



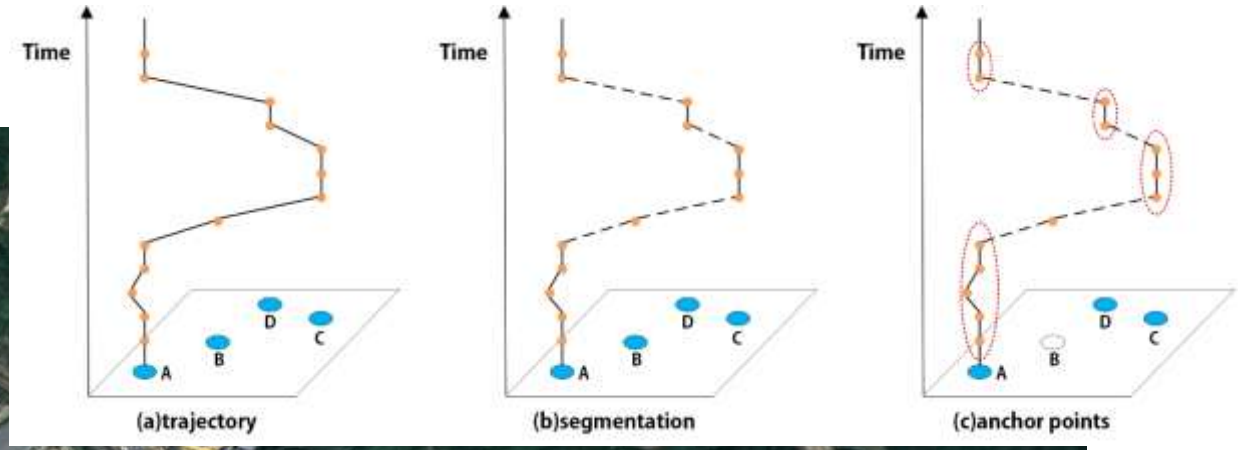
Land-use types:

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)

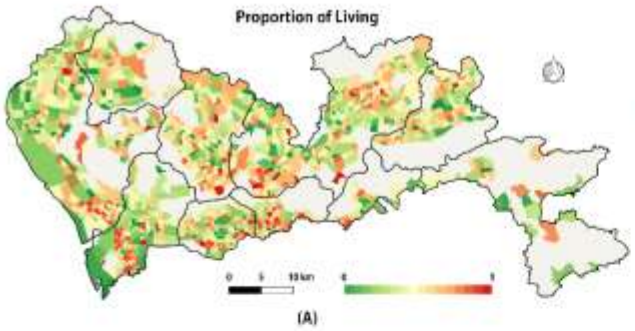
主要内容

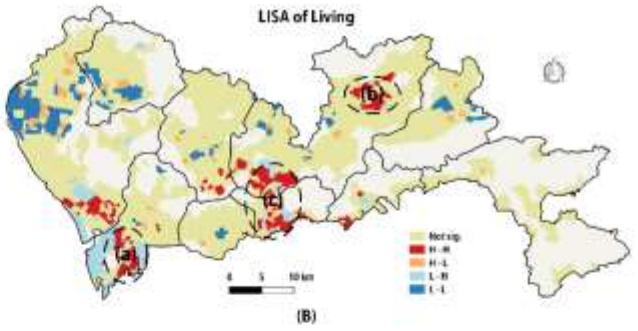
- 1 VGI大数据简介
- 2 VGI大数据关键技术
- 3 POI向量化—— POI2Vec
- 4 轨迹向量化—— Traj2Vec
- 5 基于知识图谱的城市居民活动推断
- 6 OSM与中国城市交通布局评价

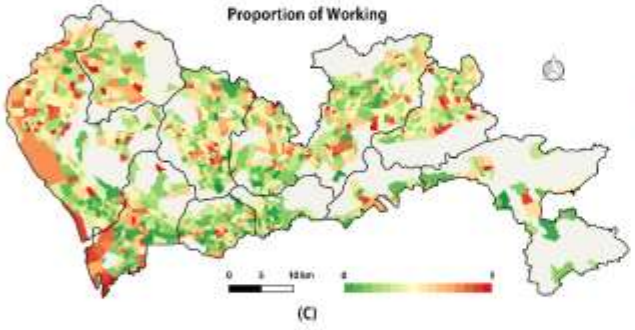
04 轨迹向量化—— Traj2Vec

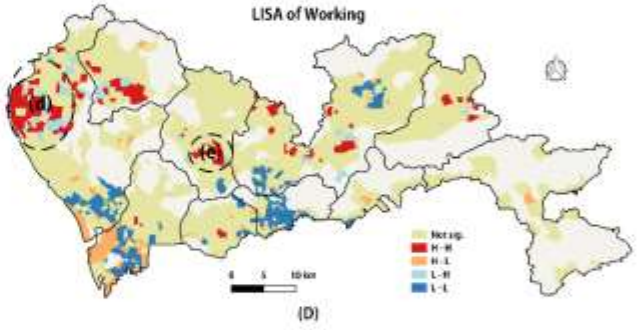


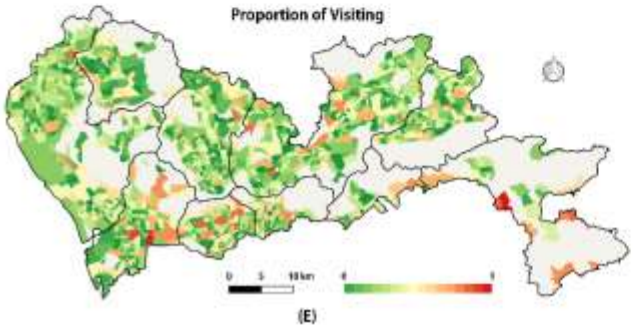
04 轨迹向量化—— Traj2Vec

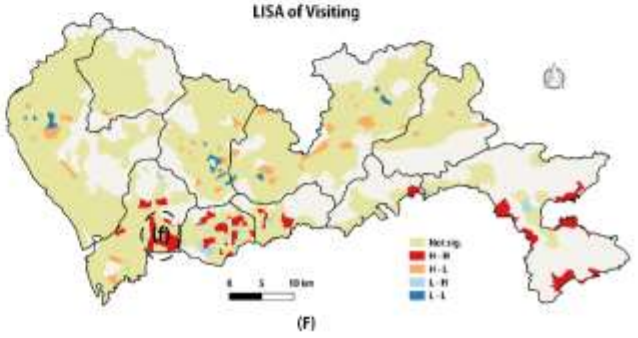


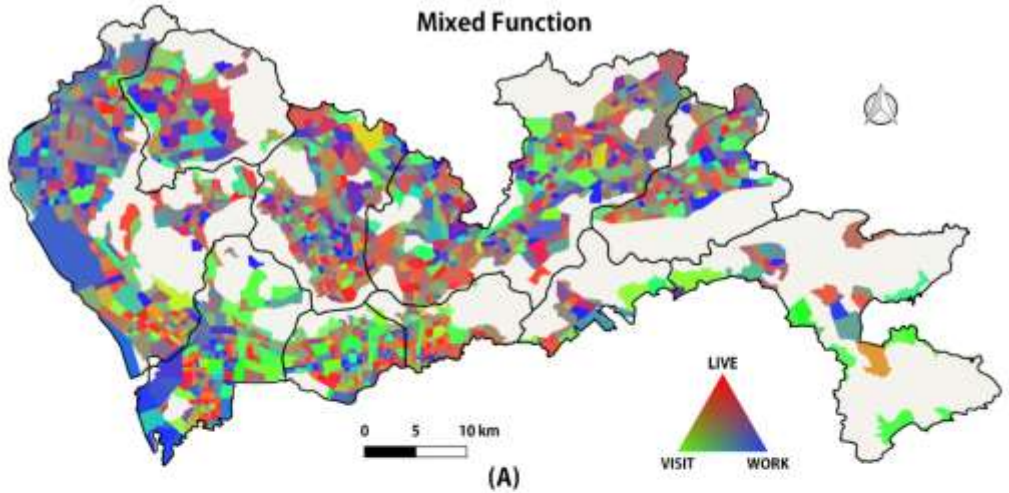


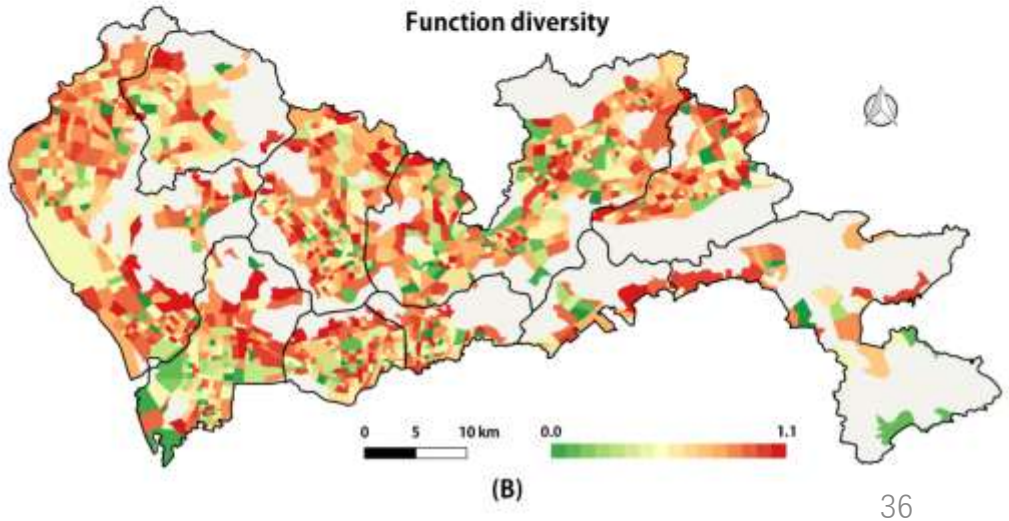












主要内容

- 1 VGI大数据简介
- 2 VGI大数据关键技术
- 3 POI向量化—— POI2Vec
- 4 轨迹向量化—— Traj2Vec
- 5 基于知识图谱的城市居民活动推断
- 6 OSM与中国城市交通布局评价

05 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？这对于交通、金融、医疗等方面都有重要意义。社交媒体数据，尤其是点评类社交媒体数据中包含的人类活动信息非常丰富。

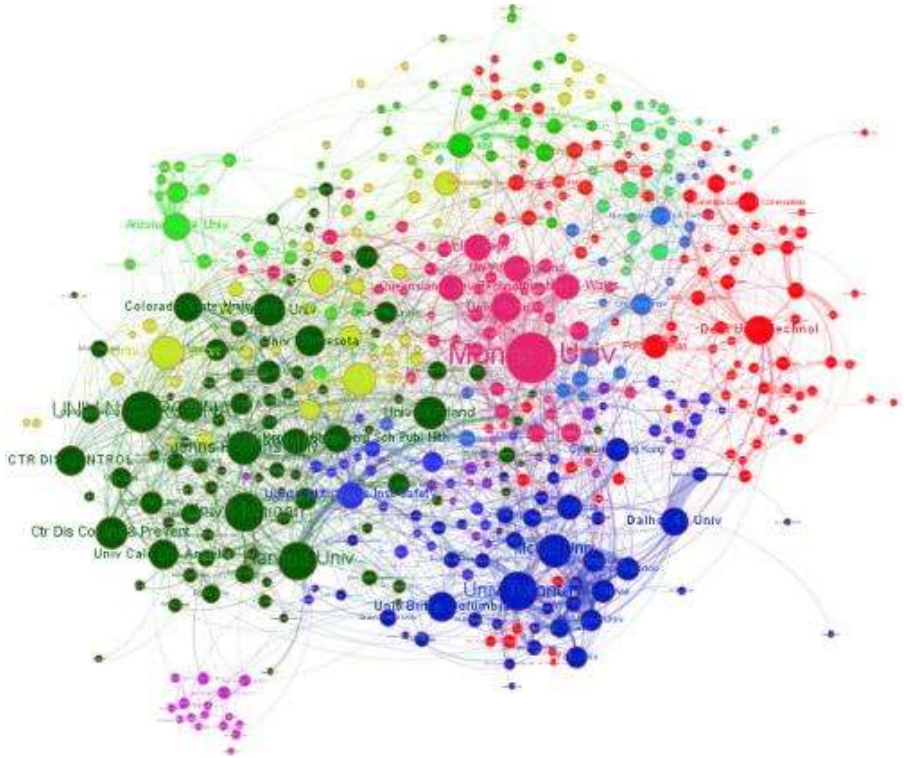
采用数据：美团店铺数据及评论数据

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

05 基于知识图谱的城市居民活动推断

- 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。



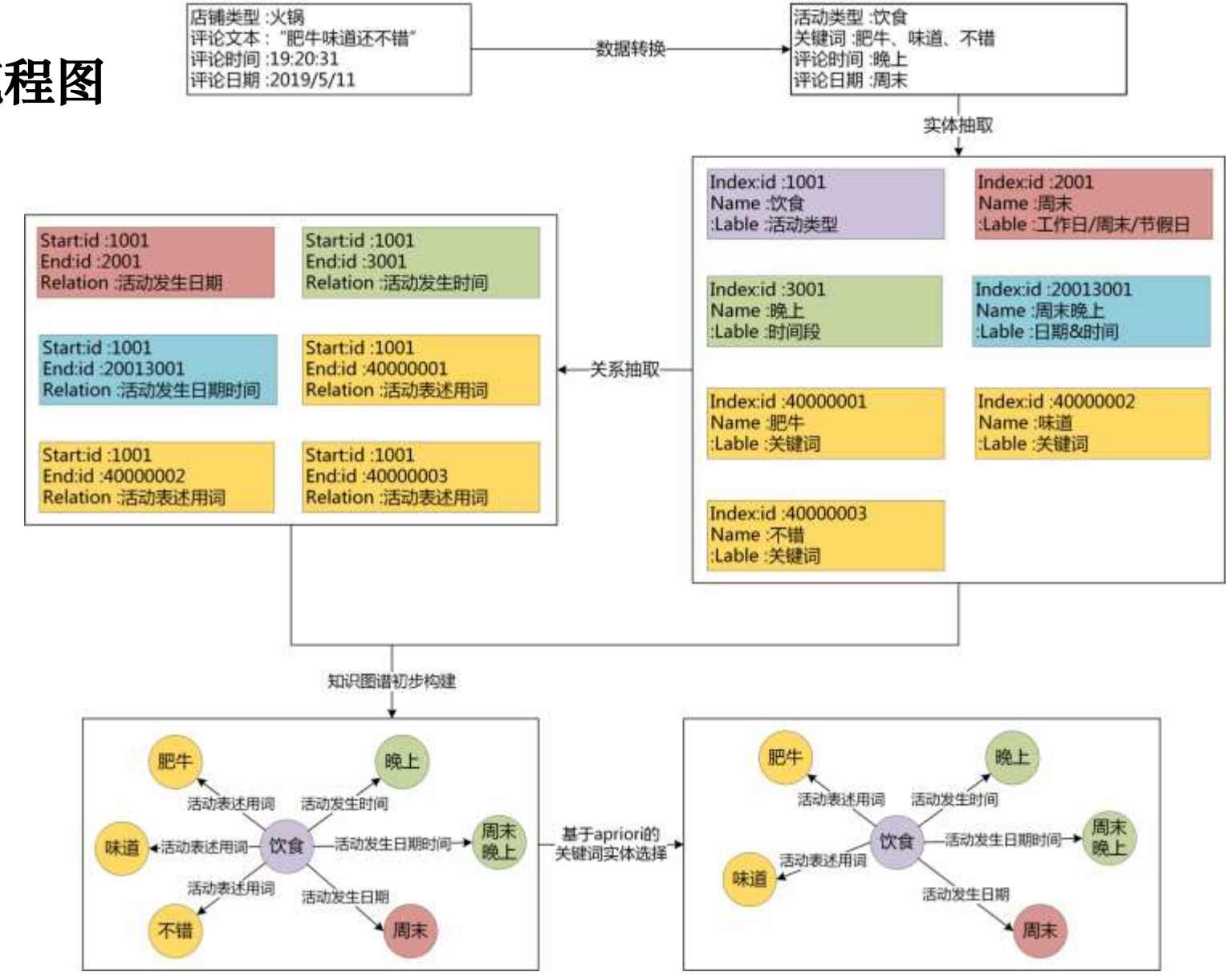
05 基于知识图谱的城市居民活动推断

➢ 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

05 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



05 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

- **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

05 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

- **时间转时间段**

  时间划分为时间段，如：凌晨，清晨，上午，中午，……

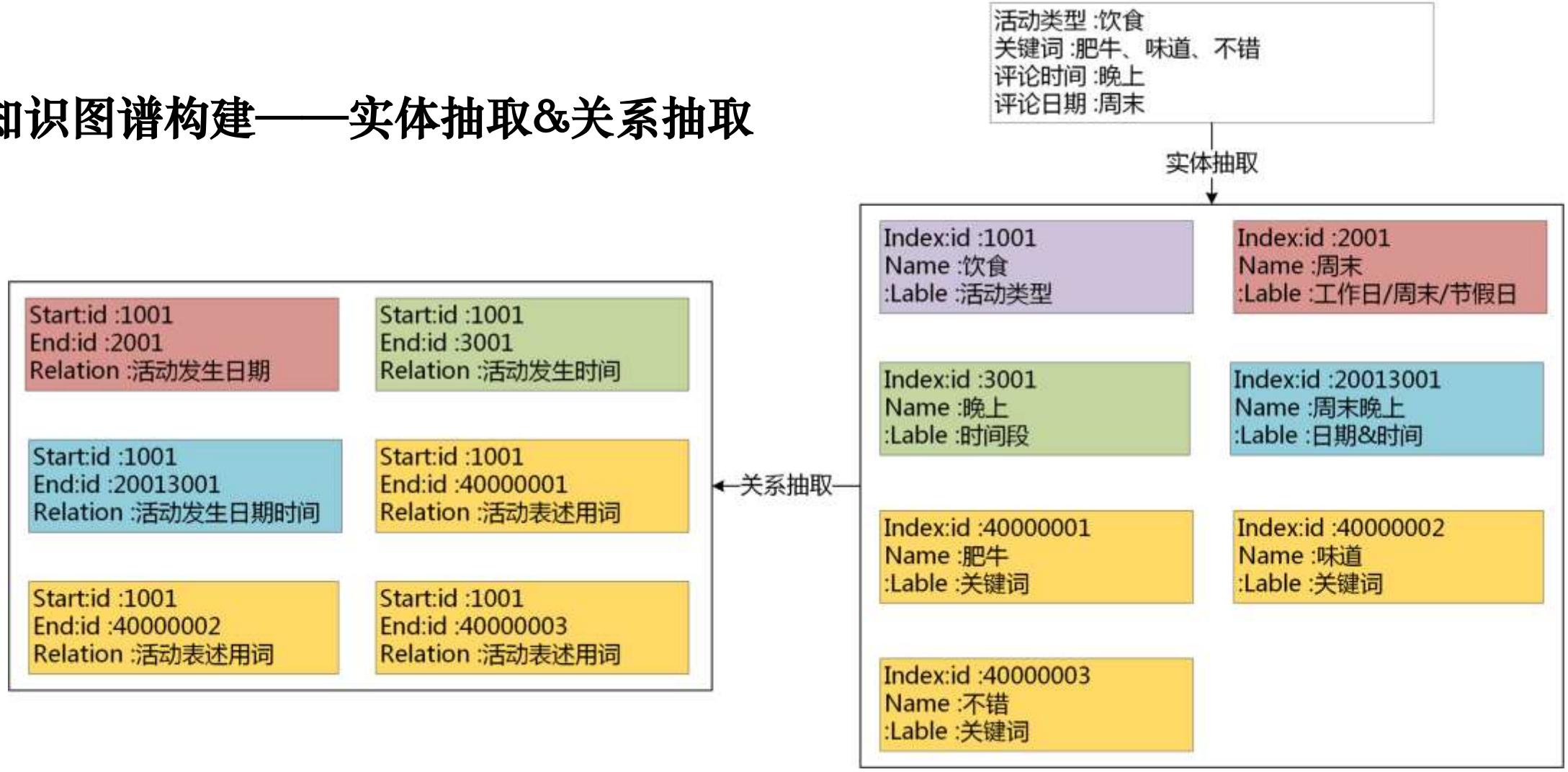
00:00:00-04:59:59	凌晨	12:00:00-13:59:59	中午
05:00:00-06:59:59	清晨	14:00:00-17:59:59	下午
07:00:00-08:59:59	早上	18:00:00-18:59:59	傍晚
09:00:00-11:59:59	上午	19:00:00-23:59:59	晚上

- **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

05 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



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

$$\mathrm{PMI}(word_1, word_2) = \frac{P(word_1, word_1)}{P(word_1) * P(word_2)}$$

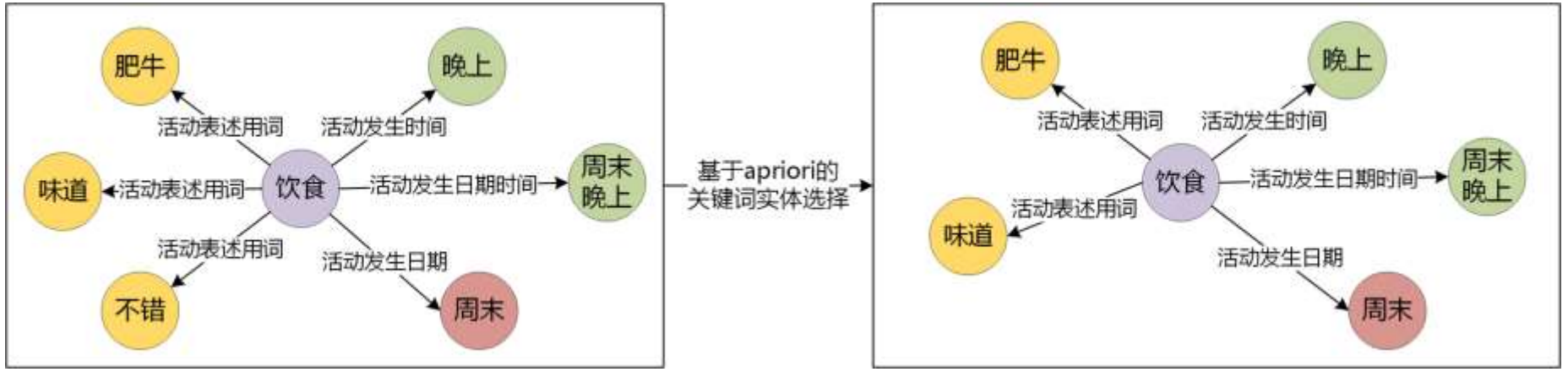
知识图谱构建——实体抽取&关系抽取

- 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support(A)=P(A), support(A=>B) = P(A∪B)

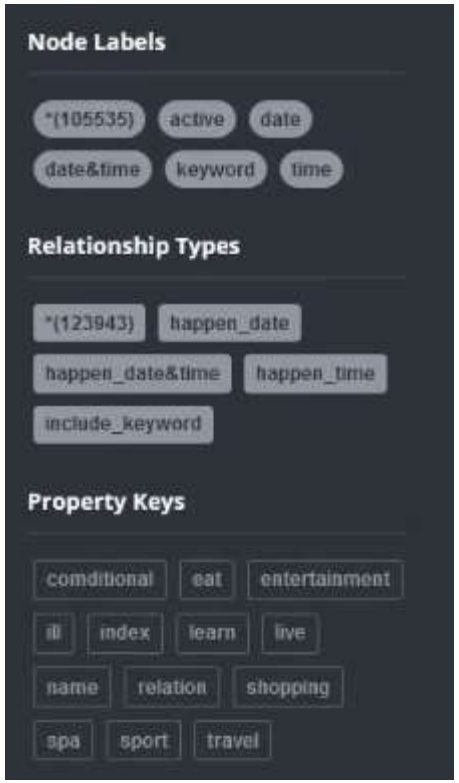
置信度confidence (A=>B) = $\frac{support(A=>B)}{support(A)}$

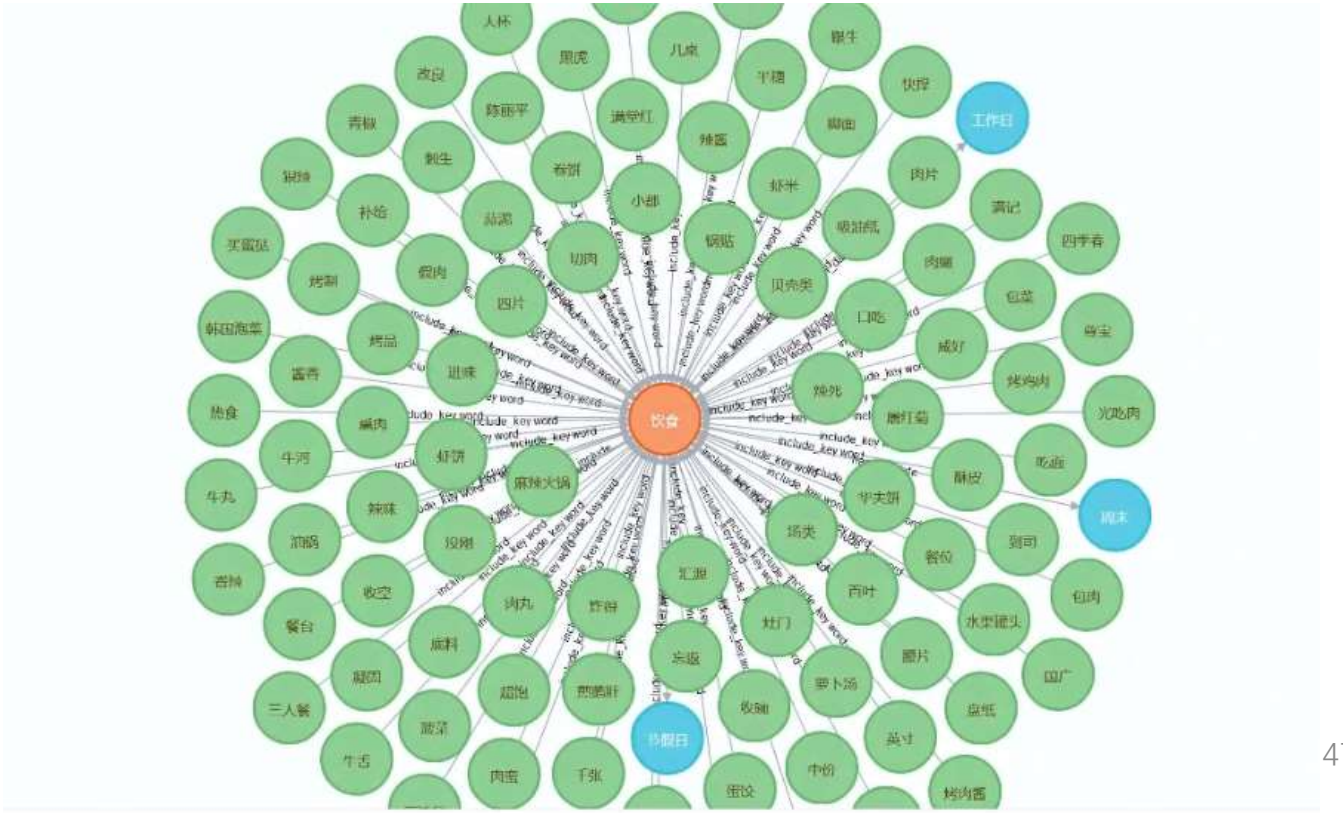


05 基于知识图谱的城市居民活动推断

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(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。

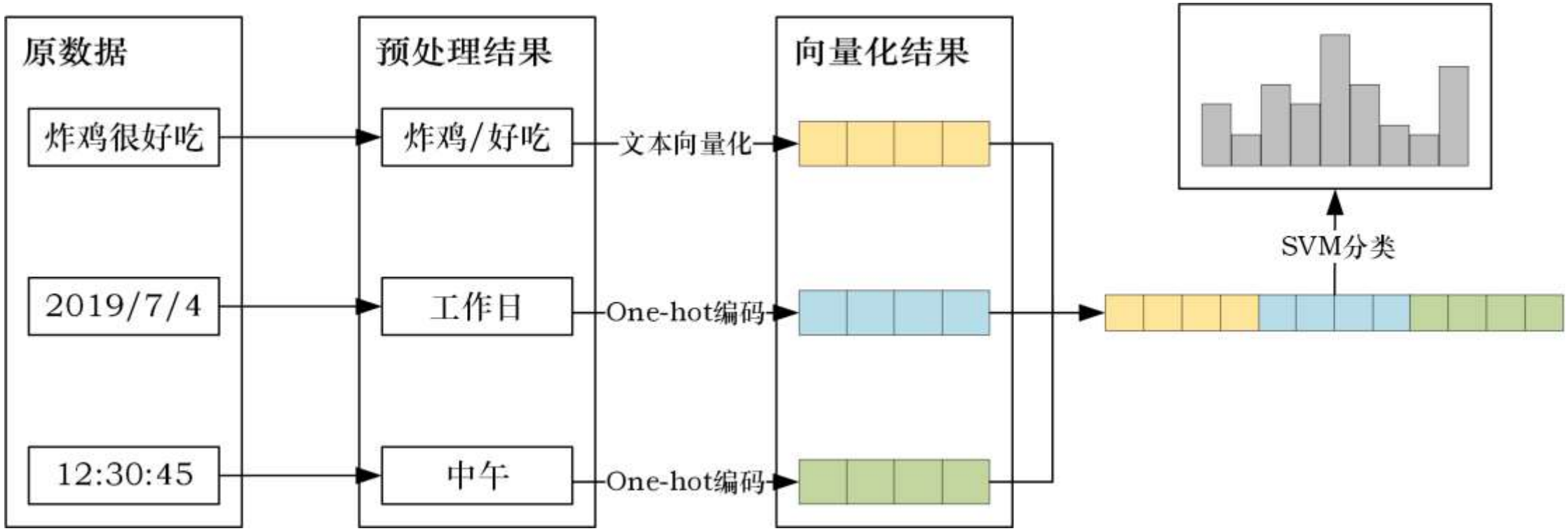




05 基于知识图谱的城市居民活动推断

基于svm进行的人类活动判定

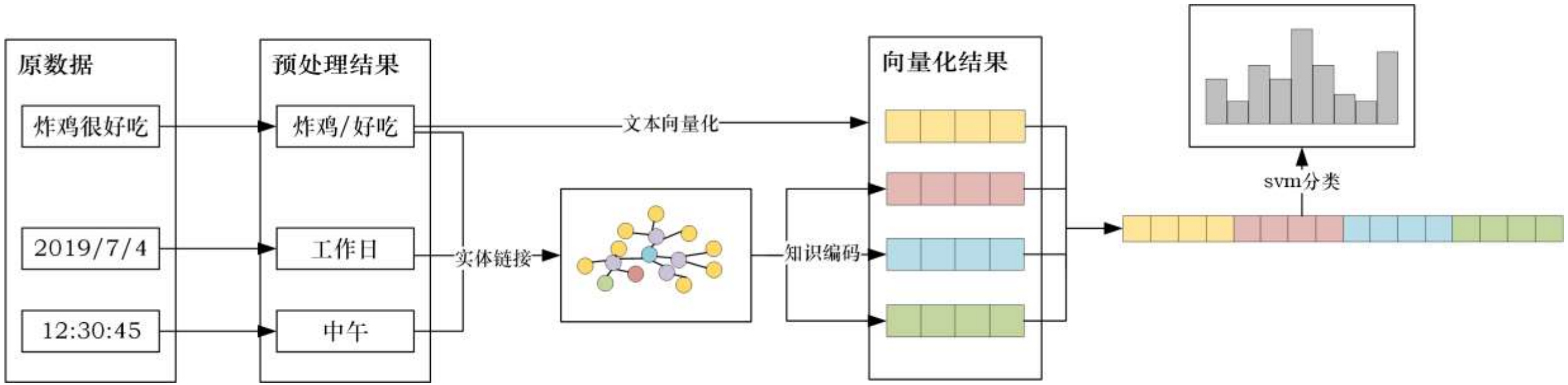
- 不加入知识图谱



05 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



05 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

05 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

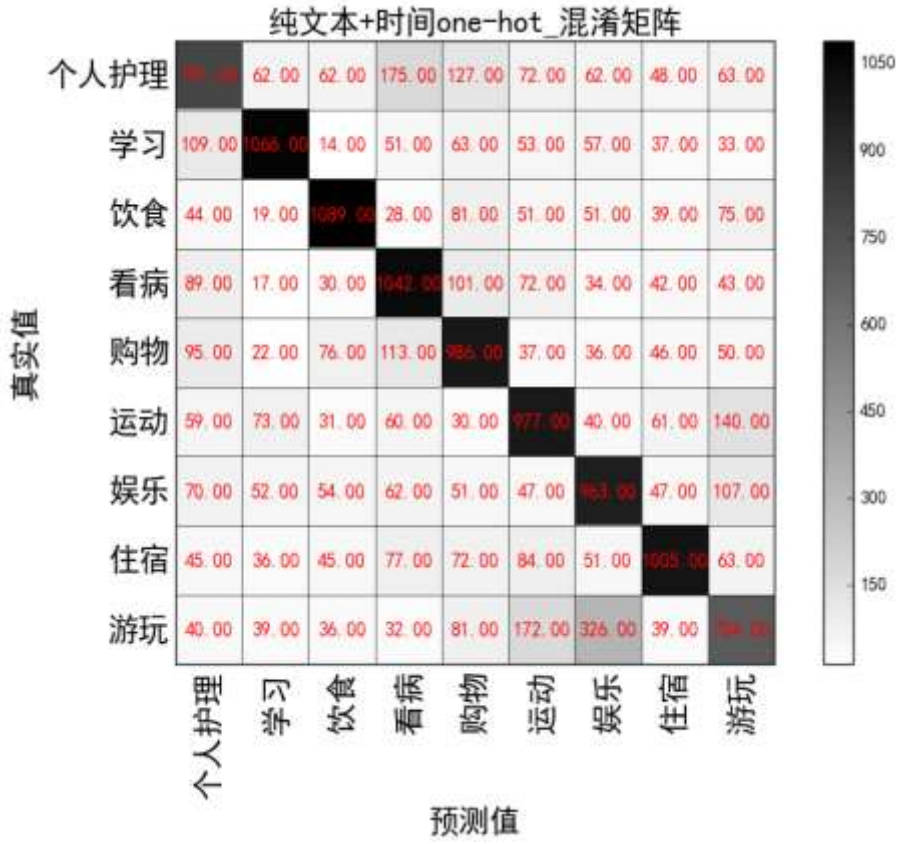
- 加入知识图谱

	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

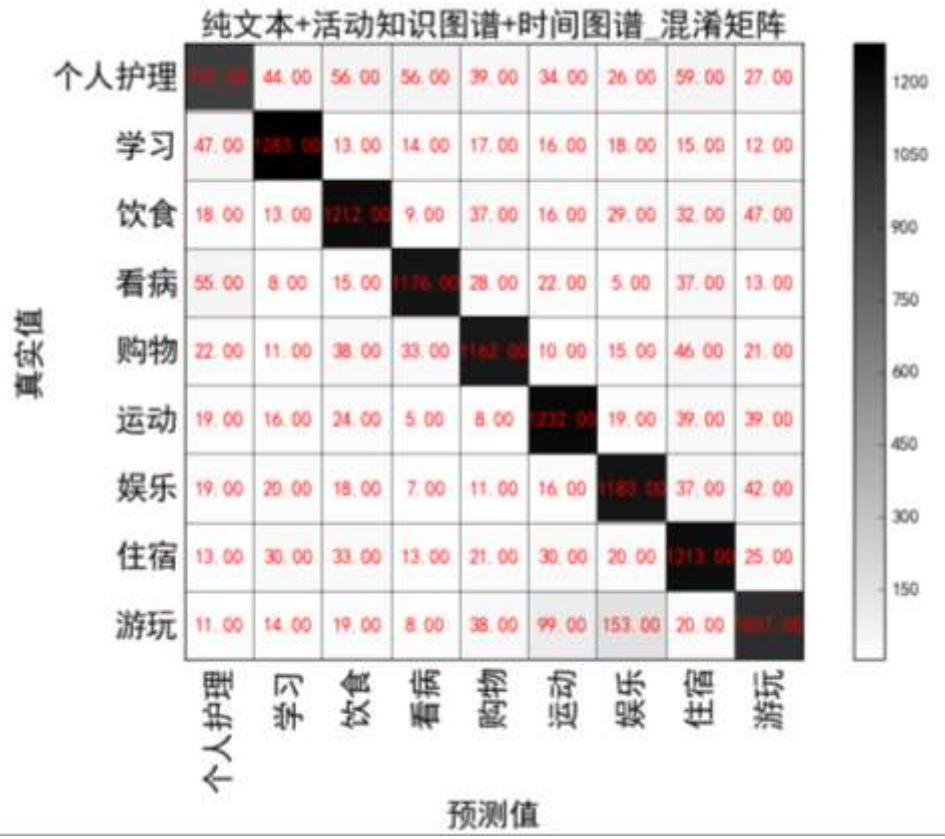
05 基于知识图谱的城市居民活动推断

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

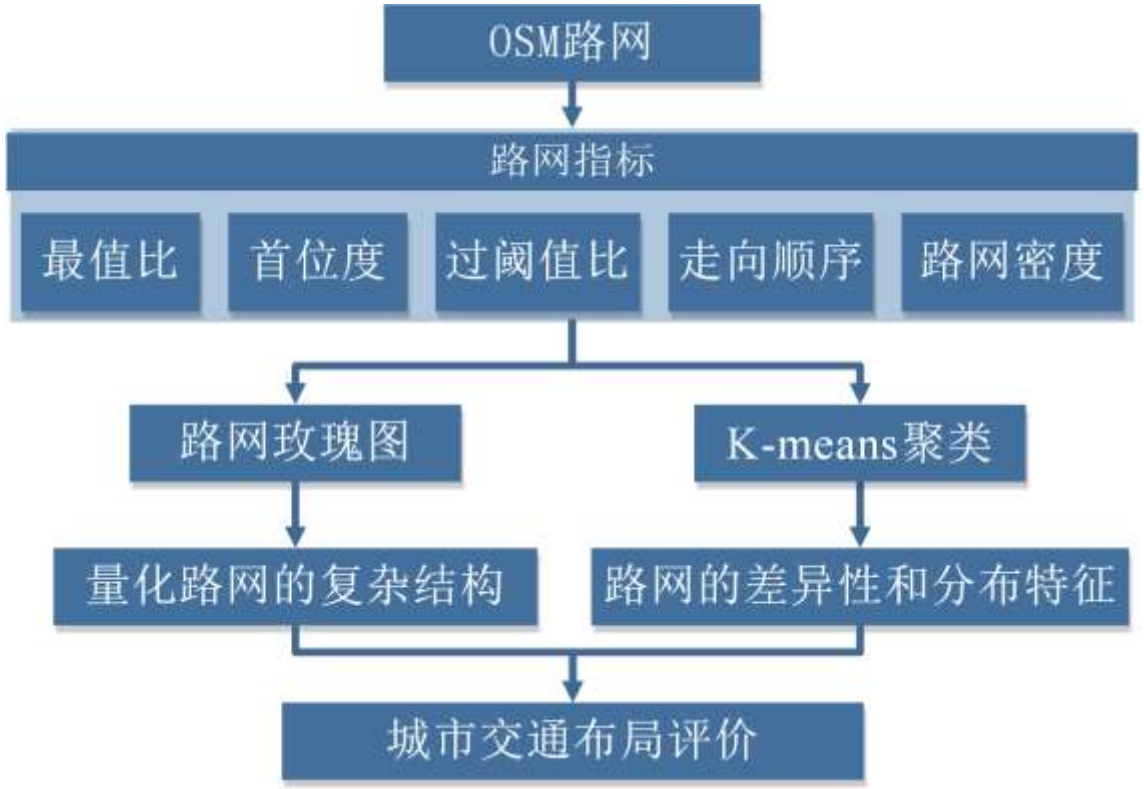


主要内容

- 1 VGI大数据简介
- 2 VGI大数据关键技术
- 3 POI向量化—— POI2Vec
- 4 轨迹向量化—— Traj2Vec
- 5 基于知识图谱的城市居民活动推断
- 6 OSM与中国城市交通布局评价

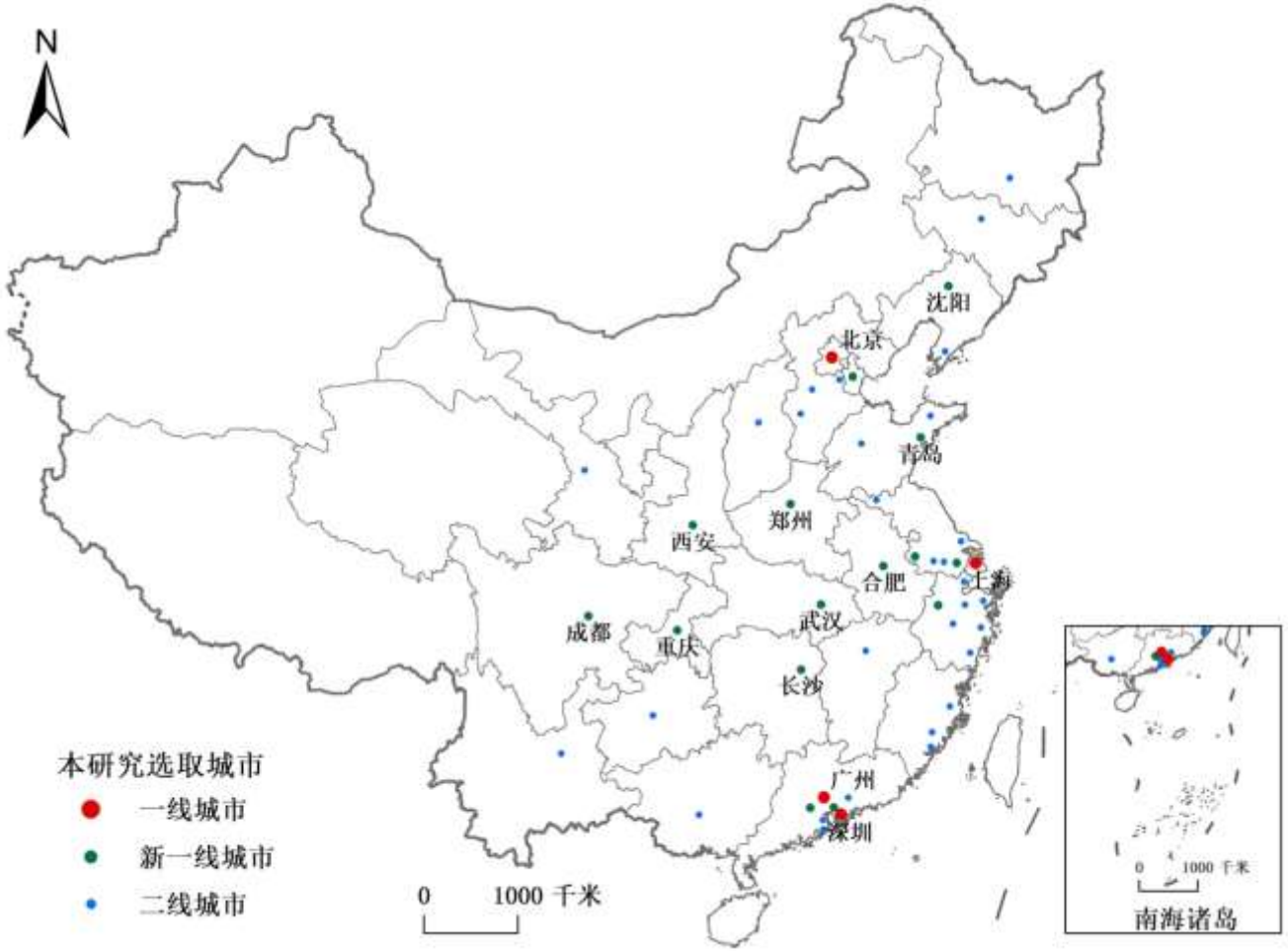
06 VGI大数据与中国城市交通布局评价

- 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？



06 VGI大数据与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

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

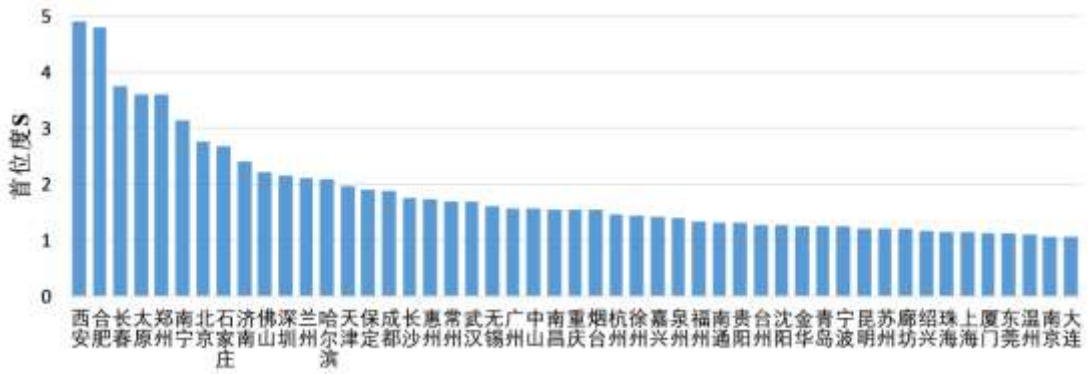
$$H_0=-\sum_{i=1}^{n}P(O_i)\ln(P(O_i)) \qquad \varphi=1-\left(\frac{H_o-H_g}{H_{max}-H_g}\right)^2$$

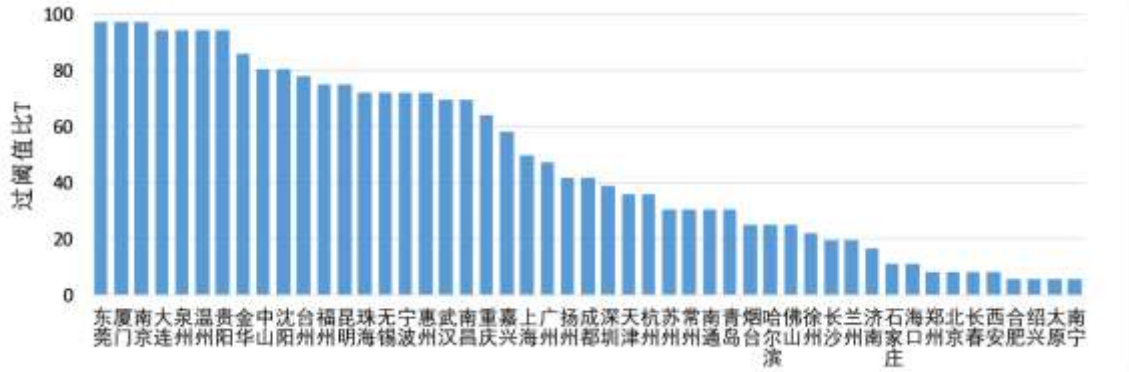
（5）**路网密度δ**：城市道路总长度与城市面积之比，从长度上描述了不同城市规模应有的道路发展水平

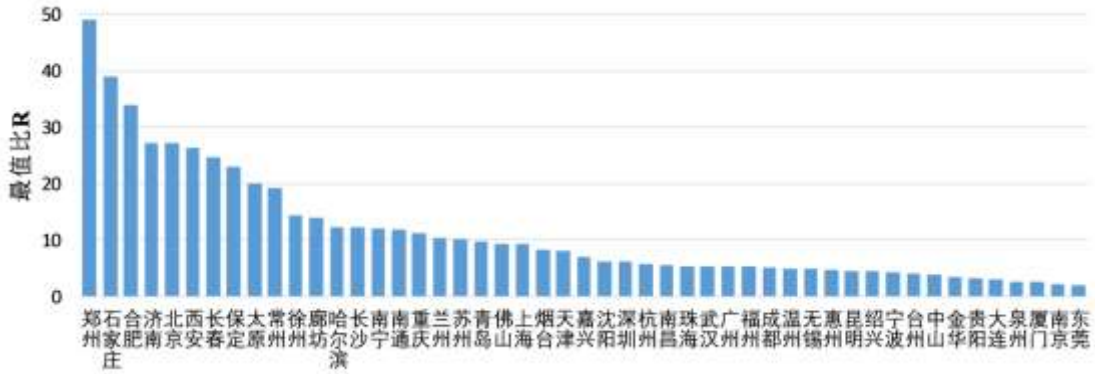
$$\delta=\frac{\sum x}{F}$$

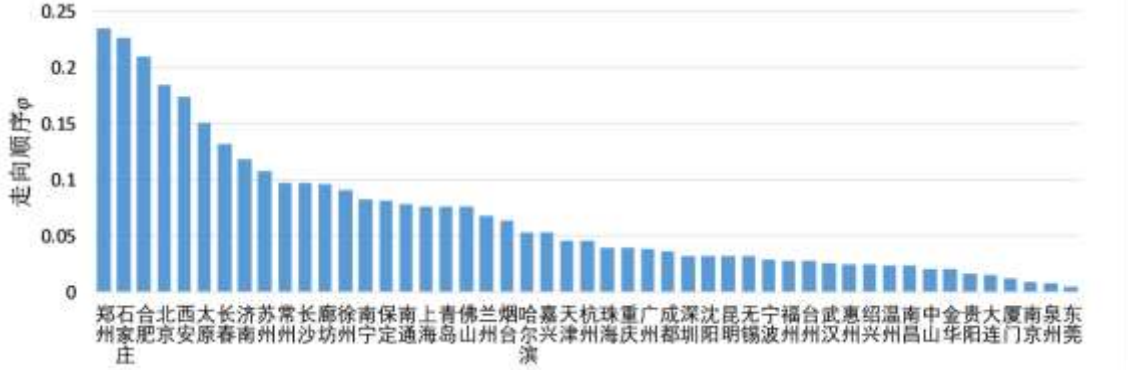
06 VGI大数据与中国城市交通布局评价

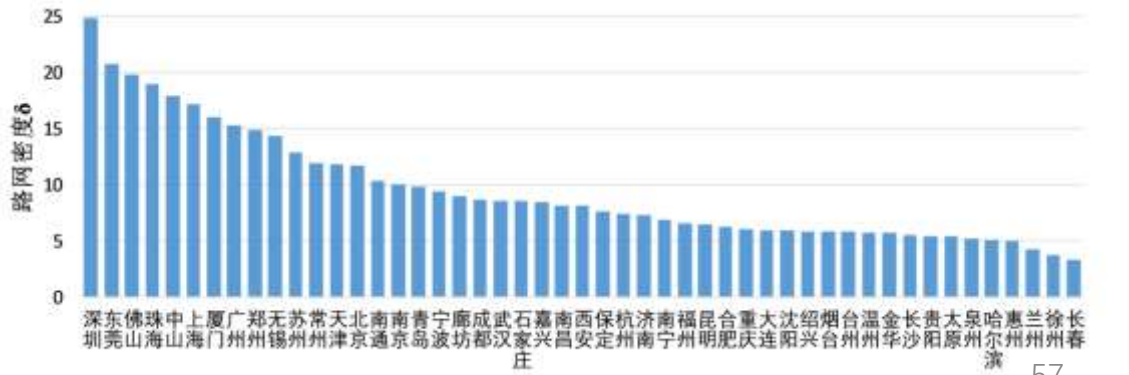
■ 路网指标数值分布







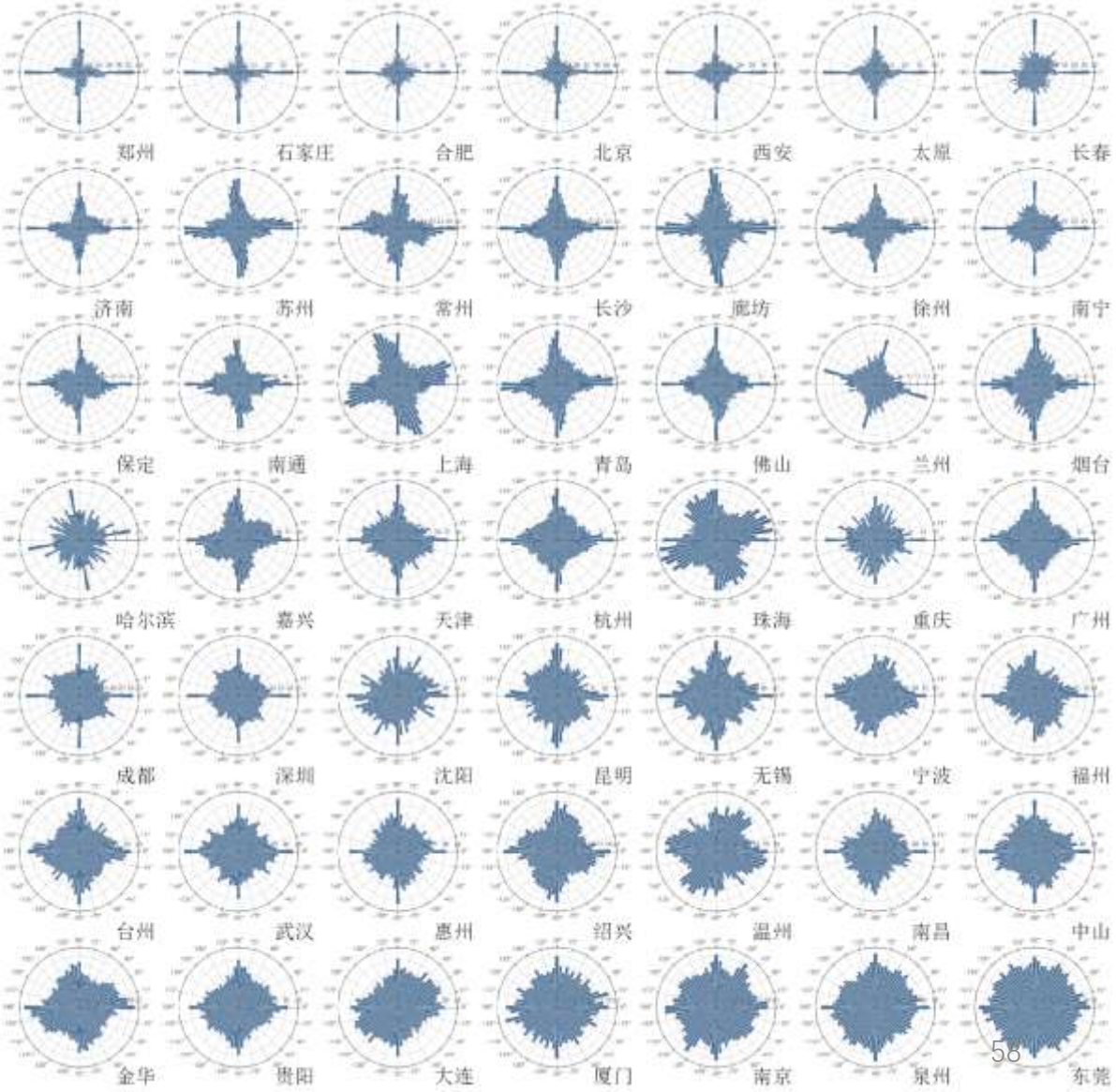




06 VGI大数据与中国城市交通布局评价

■ 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

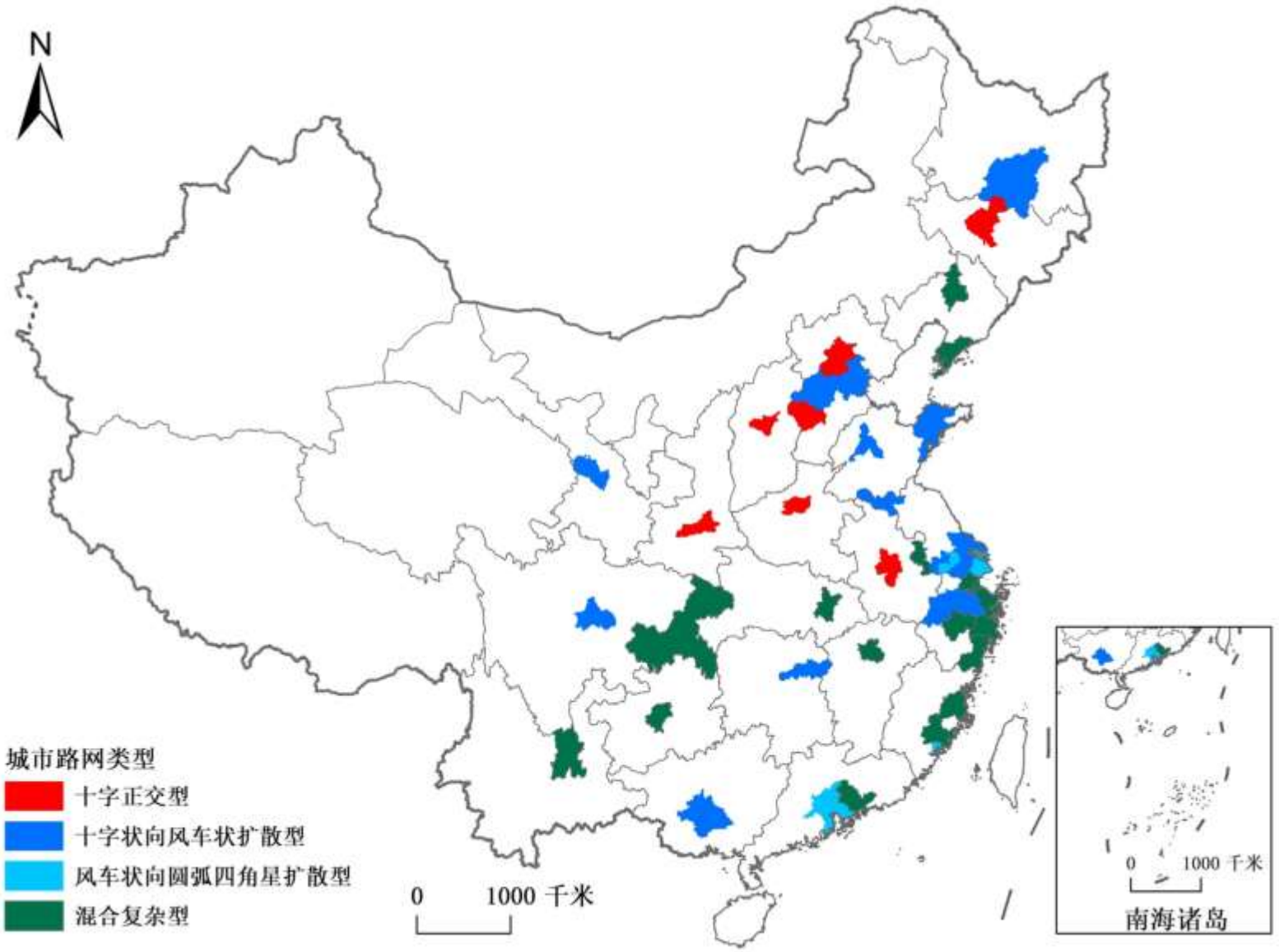


06 VGI大数据与中国城市交通布局评价

- 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

06 VGI大数据与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

06 VGI大数据与中国城市交通布局评价

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

A. 北京

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

B. 常州

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

C. 珠海

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

D. 贵阳

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

本章总结

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、大众化等的特点。

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城市居民活动、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。