



Urban region representation learning with human trajectories: a multi-view approach incorporating transition, spatial, and temporal perspectives

基于人类轨迹的城市区域表征学习：一种包含过渡、空间和时间视角的多视角方法



1 Introduction

研究目的？

- 人类轨迹所表现出的时空结构和规律与城市空间的许多社会经济方面有着密切的联系。
- 例如，不同城市区域之间的集体流动模式和相互作用表明了城市功能的空间分布(Wang et al. 2023; Yuan, Zheng, and Xie 2012)
- 各种数据挖掘技术被应用于城市规划和管理中的各种应用：土地利用分类或聚类(e.g. Liu et al. 2012; Zhang et al. 2021;) 人口估计 (e.g. Chen et al. 2018; Douglass et al. 2015), 房价预测 (e.g. Kang et al. 2021; Wang and Li 2017), 社会事件检测(e.g. Zheng et al. 2013), 交通信号控制 (e.g. Lin et al. 2023), 以及交通流量预测 (e.g. Qu et al. 2022; Zhang, Gong, Zhang, et al. 2023).

当前研究的局限性？

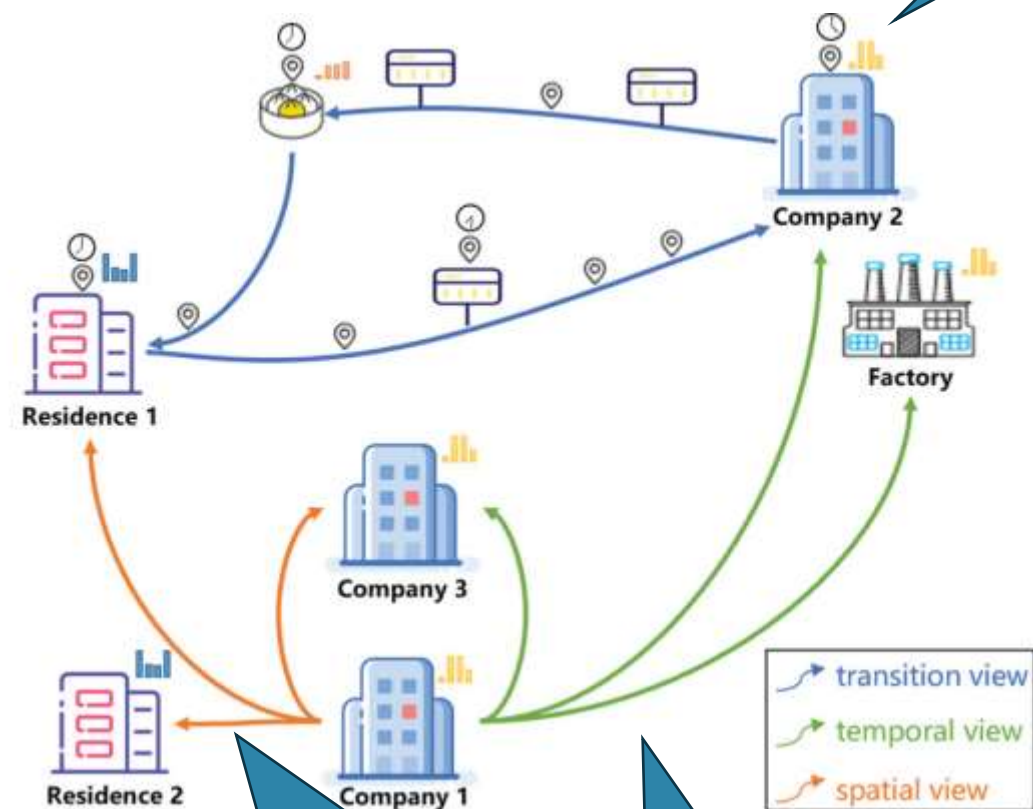
1. 人类迁移模式反映城市区域之间的远距离(多跳)依赖性往往被忽略
2. 大多研究依赖于监督学习完成特定任务，对于处理不可用或稀疏的真值数据有时是不切实际的(土地利用通常与人口分布相关，但这一特定任务所学习到的模型和特征并不一定适用于其他任务)

解决方案？

- 从人类轨迹中学习有效区域嵌入；哪些潜在信息需要被纳入下游应用

1 Introduction

■ 三种信息视角用于反映城市区域语义



人类轨迹中的多跳转换是有意义的(明确的人类运动)

没有人类运动所表达的明确的连通性，远距离以来很难在城市区域的粒度上成立

捕获时间接近度无法在长期时间相关性上成立

- **迁移视图(Transition View):** 人类在不同城市位置或区域之间的迁移是轨迹携带信息的主要来源。它们反映了城市区域之间的连通性和相关性。人类的地理位置迁移通常需要不同地点之间的紧密联系，可以显示出长期依赖关系(例如：家庭和工作场所之间的远程通勤)
- **空间视图(Spatial View):** 相邻的城市区域或地点在许多社会经济因素(人口密度、房价)具有相似性和相关性。明确建模空间视图可以加强相邻城市区域之间的相关性，并在很大程度上填补未捕获的迁移视图 (Wang and Li 2017).
- **时间视图(Temporal View):** 能够揭示城市区域(住宅区、工业区)的时间规律，被认为具有城市功能的指示性 (Wan et al. 2021)



1 Introduction

■ MTE(Multi-view Trajectory Embedding)

- 使用Skip-Gram(Mikolov et al. 2013; as used in Word2Vec)建模迁移视图
- 基于图扩散和infomax的GNN建模时间和空间视图
- 将人类轨迹的三种信息视图整合到区域嵌入中，利用不同视图特征表达进行表示学习

■ 模型应用

- 在中国深圳研究区域将生成的区域嵌入用于三个下游任务：
 - 土地利用分类、人口密度估计和房价预测



2 Related Work



2 用于表征城市空间的轨迹挖掘

■ (早期)特征工程方法

- 从轨迹中手工提取特征

■ (近年)机器(深度)学习技术的转变

➤ Hu et.al(2021)

- 结合Skip-Gram和GCN的框架用于道路级别的城市功能分类
- Skip-Gram利用出租车轨迹获得路段嵌入，并在邻接图中进行图卷积
- 忽略了轨迹中的时间信息

➤ Sun et.al(2022)

- 利用深度卷积自编码器从手机使用中重建时间特征；用于聚类分析和土地利用分类
- 深入挖掘了每个城市区域的时间模式
- 缺乏空间模式和人类活动的相关性探索

■ (日益兴盛)利用人类流动性数据学习区域表示

□ 无监督

- 没有来自下游任务的基础真值数据的先验知识
- 学习多任务表示

➤ Wang and Li et.al(2017)

- 考虑出租车流量数据区域之间的时间相关性、多跳转换、空间邻接性
- 局限：远距离空间依赖性可以很容易捕获，导致遥远区域的嵌入较为平滑
- 每个区域在每个时间段较为重复，构建的图很难应用于大城市

2 数据

■ 数据来源

- 2012年3月22日和23日两天内的匿名手机位置数据
- 记录了5818个信号塔中超过1600万个手机用户的位置变化
- 每条轨迹为信号塔之间的一系列迁移序列
- 利用该数据集学习不同下游任务的区域嵌入

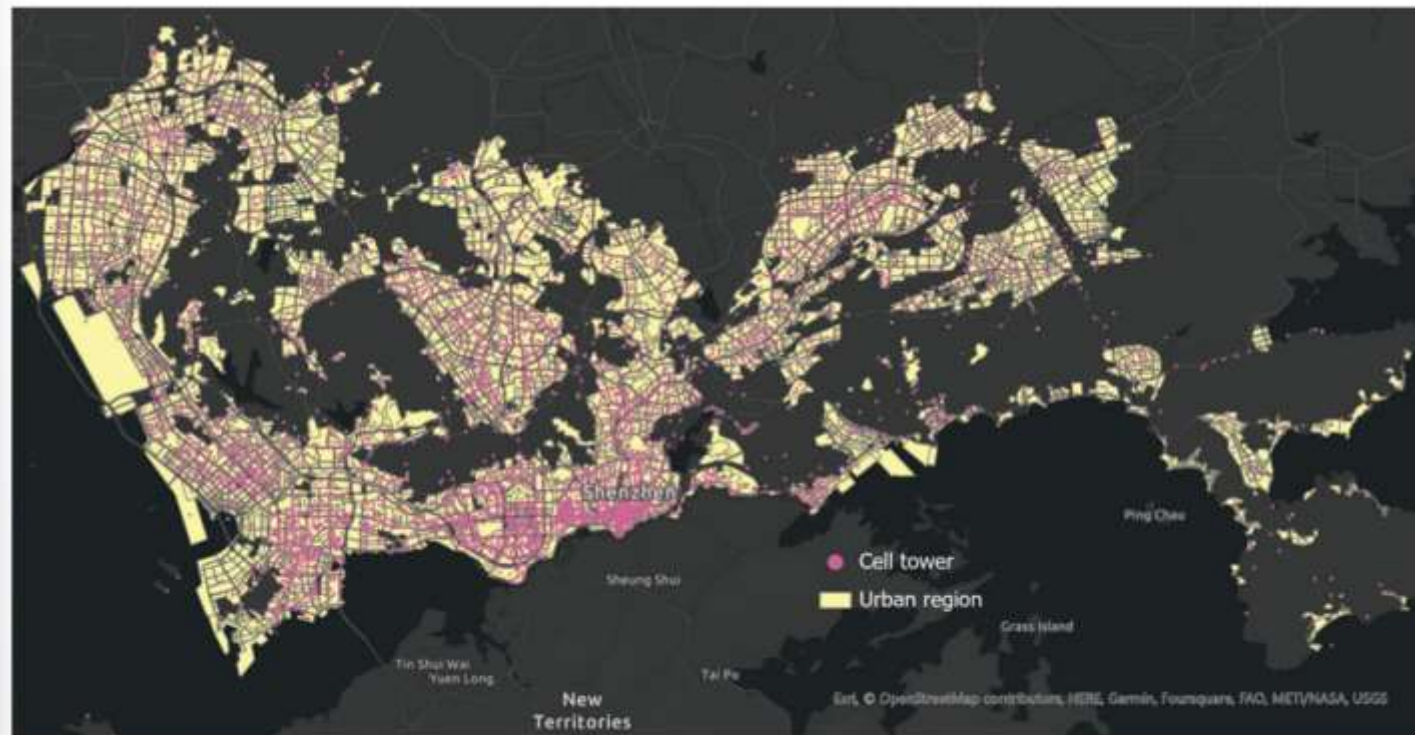
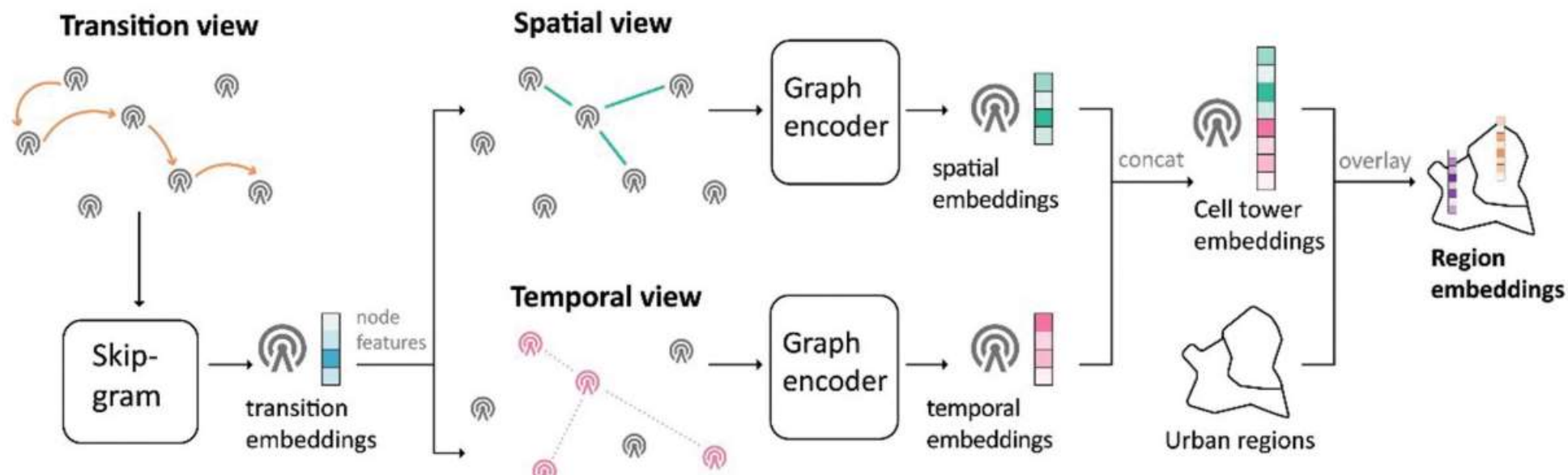


Figure 2. The study area, Shenzhen, encompassing 5,818 cell towers and 6,890 distinct regions.

3 Methodology

3 MTE模型图



1. Skip-Gram:

将人类在信号塔之间转移的轨迹序列输入Skip-Gram模型中以学习信号塔的转移嵌入
捕获大量人类轨迹所带来的远距离(多跳)依赖关系

2. 转换为时空图

空间图: 信号塔的空间邻接性
时间图: 信号塔时间模式的相似性

3. MVGRL无监督图表示学习模型

学习时间和空间视图的信号塔嵌入
通过图扩散、局部和全局表示之间的相互信息最大化来学习蜂窝塔嵌入

4. 利用Voronoi将多视图信号塔嵌入映射到下游任务中

3.1 学习迁移视图的信号塔嵌入

■ 将轨迹输入Skip-Gram

- 将轨迹数据处理为一个信号塔标识符序列，可以将它视为一个由几个单词组成的句子
- 设定窗口大小为 w ，对于轨迹中的每个信号塔，在窗口中检索所有上下文信号塔；在信号塔之间形成共现对，捕获它们之间的多跳依赖关系；

$$\mathcal{L}_{transition} = \sum_{c \in \mathcal{C}} \sum_{c_q \in N_{R(c)}} - \left(\log \left(\sigma(\mathbf{c}_i^{rT} \mathbf{c}_q') \right) - \sum_{i=1}^k \log \left(\sigma(\mathbf{c}_i^{rT} \mathbf{c}_{n_i}') \right) \right)$$

包含共现对和负采样过程；通过最小化目标函数获得迁移视图中的信号塔嵌入

\mathbf{c}_i^r : 表示迁移视图中信号塔 \mathbf{c}_i 目标嵌入

$N_{R(c)}$: 表示给定窗口内 \mathbf{c}_i 的上下文信号塔集

\mathbf{c}_q' : 表示上下文信号塔集的嵌入

σ : 激活函数

\mathbf{c}_n^i : 负采样过程得到的信号塔

3.2 构建时空图

■ 空间视图

- 使用k-近邻(KNN)算法构建边

$$w_s(c_i, c_j) = \log \left((1 + D^{1.5}) / (1 + sd_{c_i c_j}^{1.5}) \right)$$

根据公式为每条边分配一个非归一化的权值 w_s

D : 表示信号塔最小边界矩形的对角线长

$sd_{c_i c_j}$: 表示两个信号塔之间的空间距离

最后将所有边权重线性缩放为[0,1]

$$G_s = (\mathbf{C}_r, \mathbf{A}_s)$$

■ 时间视图

- 总结每个信号塔按比例的小时访问分布；构成一个信号塔的时间坐标 $tc_i = \{tp_1, tp_2, \dots, tp_{24}\}$
- tp_j 表示该小时内占全天的访问比例；
- 因此满足 $tp_j \in [0,1]$ and $\sum_k tp_j = 1$

对时间坐标首先进行归一化；

使用L2距离捕获信号塔之间的时间相似性

使用KNN策略构造具有非归一化边根据公式为每条边分配一个非归一化的权的时间图

$$G_t = (\mathbf{C}_r, \mathbf{A}_t)$$

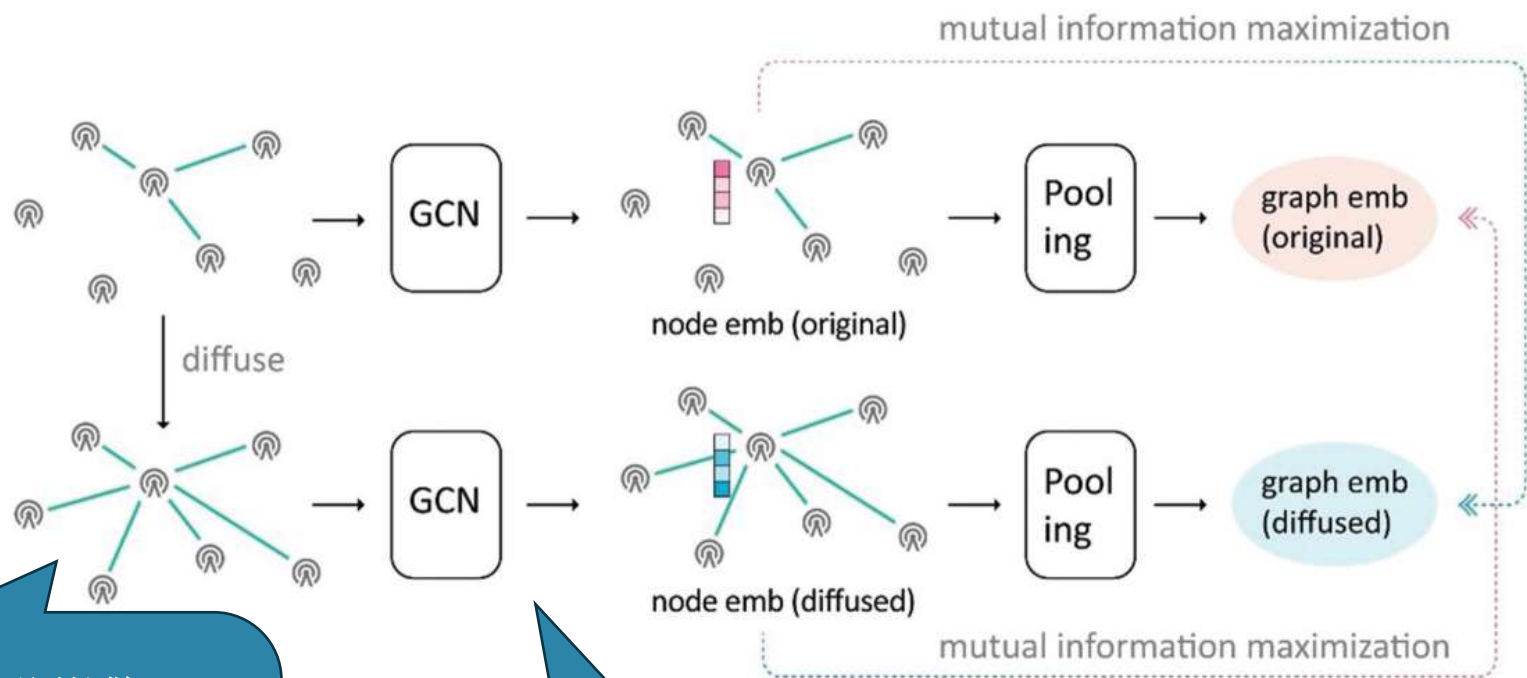
原理上，这两个图可以被视为一个具有两种边(时间边和空间边)的多路图，将其保留为两个图的原因：

时间视图和空间视图在不同的下游任务中具有不同程度的有效性
并且完全混合所有信息视图有时并不能保证最佳性能

3.3 用无监督图学习时间 和空间嵌入

- 设计目标：以完全无监督的方式学习多任务区域表示

图表示学习对时空视图是有效的：其中的直接和单跳邻居之间的相关性建模比多跳和远程依赖关系更有意义



3. (Infomax信息最大化原则) 通过最大化一个版本的节点嵌入和另一个版本的图嵌入(通过池化获得)之间的互信息来训练模型

1. 利用图扩散 (Klicpera, Weissenberger, and Günnemann 2019) 创建原始图的另一个版本

2. 原始版本和扩散版本都进行图卷积以获得节点嵌入

➤ 通过这种方式，节点嵌入既具有局部相关性又具有全局相关性，并携带了图的不同结构形式所包含的丰富的连通性信息。

3.3 用无监督图学习时间 and 空间嵌入

1 图扩散技术

- 发现了图的更全局的结构形式，以捕获未连接的信号塔之间的细微联系

通过PageRank算法实现；给定图 G_s ，将其邻接矩阵 A_s 变换为 A_s^d

$$A_s^d = \beta \left(\mathbf{I} - (1 - \beta) \mathbf{D}_s^{-\frac{1}{2}} \mathbf{A}_s \mathbf{D}_s^{-\frac{1}{2}} \right)$$

β : 可调的传送概率

D_s : 邻接矩阵 A_s 的度矩阵

2 图卷积生成节点嵌入

$$\mathbf{c}_s^o = \tau \left(\hat{\mathbf{D}}_s^{-\frac{1}{2}} \hat{\mathbf{A}}_s \hat{\mathbf{D}}_s^{-\frac{1}{2}} \mathbf{c}_r \Theta_s^o \right)$$

$$\mathbf{c}_s^d = \tau \left(\hat{\mathbf{D}}_s^{d-\frac{1}{2}} \hat{\mathbf{A}}_s^d \hat{\mathbf{D}}_s^{d-\frac{1}{2}} \mathbf{c}_r \Theta_s^d \right)$$

$\hat{\mathbf{A}}$: 包含自环的邻接矩阵

$\hat{\mathbf{D}}$: 邻接矩阵 A_s 的度矩阵
得到原始时空图和扩散版本的节点嵌入

3 池化生成图嵌入

$$\mathbf{g}_s^o = \sigma \left(\frac{1}{n_c} \sum_{i=1}^{n_c} \mathbf{c}_{s,i}^o \right)$$

$$\mathbf{g}_s^d = \sigma \left(\frac{1}{n_c} \sum_{i=1}^{n_c} \mathbf{c}_{s,i}^d \right)$$

提供全局级别的监督信号

3.3 用无监督图表示学习 学习时间和空间嵌入

4 infomax最大化图的节点嵌入和图嵌入的互信息来优化

$$\mathcal{L}_{spatial} = -\frac{1}{n_c} \sum_{i=1}^{n_c} \left(\mathcal{D}(\mathbf{c}_{s,i}^o, \mathbf{g}_s^d) + \mathcal{D}(\mathbf{c}_{s,i}^d, \mathbf{g}_s^o) \right)$$

\mathcal{D} : 基于Jensen-Shannon精度的噪声对比型目标建模判别:

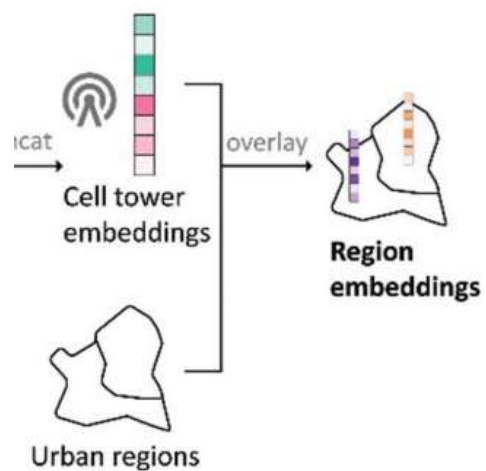
$$\begin{aligned} \mathcal{D}(\mathbf{c}_{s,i}^o, \mathbf{g}_s^d) = & \log \left(f_{\varphi}(\mathbf{c}_{s,i}^o, \mathbf{g}_s^d) \right) \\ & - \log \left(1 - f_{\varphi}(\tilde{\mathbf{c}}_{s,i}^o, \mathbf{g}_s^d) \right) \end{aligned}$$

通过最小化目标函数，将其在原始空间图和扩散空间图中对应节点嵌入相加；得到信号塔的空间嵌入；时间嵌入得到的过程相同

3.3 将信号塔嵌入映射到区域嵌入

使用Voronoi图划定每个信号塔的服务区域(Zhang et.al 2021)

区域嵌入定义为Voronoi多边形在空间上与该区域相交的信号塔嵌入的面积加权求和

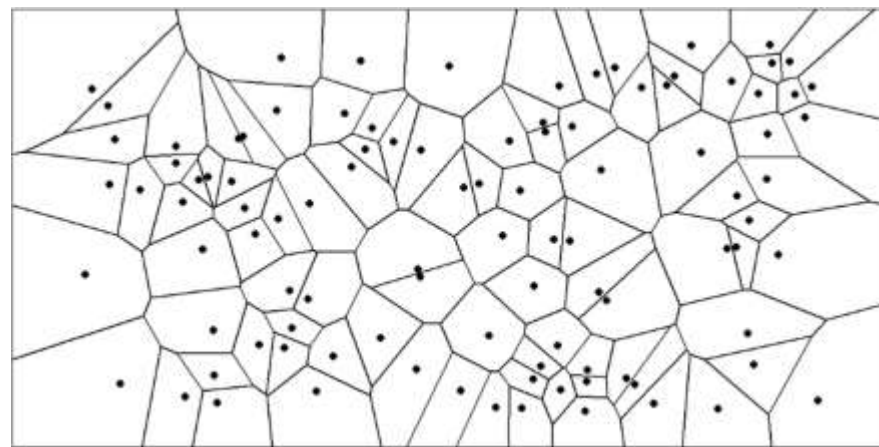


$$\mathbf{r}_i = \frac{1}{a_{r_i}} \sum_j a_{r_i, c_j} \mathbf{c}_j$$

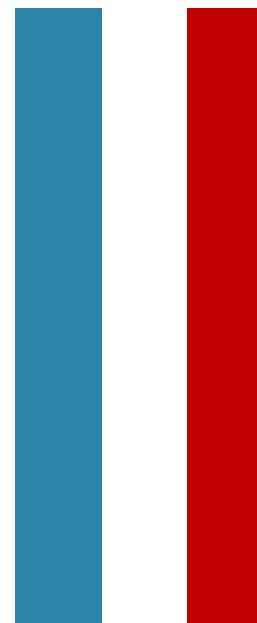
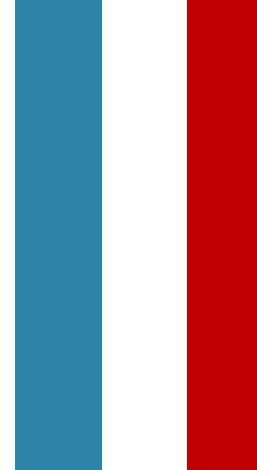
a_{r_i} : 区域的面积

a_{r_i, c_j} : 区域与信号塔所在Voronoi多边形的面积

\mathbf{c}_j : 信号塔嵌入



4 Experiment



4.1 基线模型

➤ **Traj2Vec(Zhang et al. 2021):**

- 本质上用于学习区域嵌入的Skip-Gram模型
- 相当于只是用MTE迁移视图生成区域嵌入

➤ **MNE(Zhang et al. 2018):**

- 将三个视图视为多路图，即由三种类型的边连接的节点
- 迁移视图中的边是基于信号塔之间的迁移频率构建

➤ **HDGE(Wang and Li. 2017):**

- 构建信号塔的流程和空间图
- 不考虑时间图

➤ **HIER(Shimizu, Yabe, and Tsubouchi 2020):**

- 使用LSTM和下一个位置预测训练来学习区域嵌入
- 使用信号塔直接访问次数作为输入特征来学习时间模式

➤ **SK-3views:**

- 使用Skip-Gram分别学习三种视图的嵌入
- 最后将三个视图的嵌入连接起来形成区域嵌入

➤ **Graph-3views:**

- 使用图表示学习模型MVGRL代替MTE中的Skip-Gram

4.2 探究模型有效性

➤ 通过测量区域嵌入的余弦相似性对相似信号塔进行相似度搜索：



● 对粉色点(工业园区)搜索相似位置

1. 只使用空间图嵌入(黄色点): 尽管空间位置上接近, 得到的点(住宅区)功能并不相似
2. 只使用时间图嵌入(橙色点): 功能相似, 但相距很远
3. 结合时空图嵌入(蓝色点), 它们在空间上相邻, 功能上相似, 说明模型平衡了空间接近性和功能相似性

4.3 下游任务实验

■ 土地利用分类

- 数据：深圳市2014年土地利用数据，涵盖5487个区域
- 细粒度土地利用类型: (1)自然和开放空间（nat.），(2)商业（com.），(3)住宅（res.），(4)工业（ind.），(5)公共服务（pub.）和(6)运输和物流（trans.）
- 表现：

Table 1. Performance of land use classification using MTE and several baseline methods, with the evaluation measures of accuracy (ACC), weighted precision (WP), and weighted F1 score (WF1).

Model	ACC	WP	WF1
Traj2Vec	0.527±0.010	0.519±0.011	0.508±0.011
MNE	0.543±0.010	0.535±0.012	0.525±0.012
HDGE	0.520±0.009	0.511±0.009	0.505±0.009
HIER	0.537±0.011	0.532±0.011	0.513±0.011
Sk-3views	0.528±0.010	0.525±0.012	0.502±0.010
Graph-3views	0.521±0.010	0.516±0.010	0.506±0.010
MTE-temporal	0.530±0.010	0.519±0.011	0.509±0.010
MTE-spatial	0.520±0.010	0.508±0.010	0.508±0.010
MTE	0.554±0.011	0.546±0.012	0.538±0.012

➤ 可视化MTE、MTE-spatial、MTE-temporal的土地利用混淆矩阵：

- 住宅区和工业区的准确性高于其他土地利用类型
- 对于住宅区，时间相邻性带来更大的提升，说明住宅区的时间规律性比空间相邻性更能说明问题
- 对于自然和开放空间、交通和物流等土地利用类型，空间相邻性得分更高，表明土地利用类型的空间聚类效应比其时间规律更突出

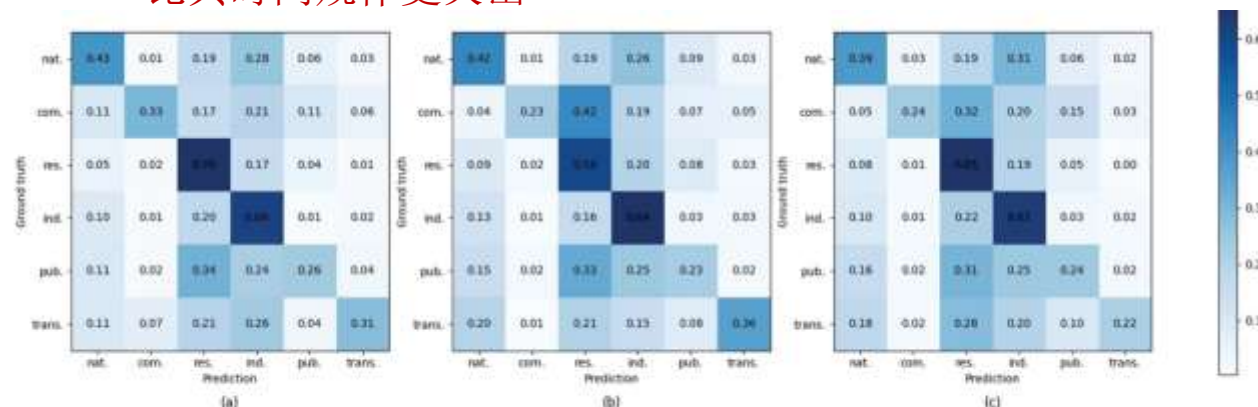


Figure 6. The confusion matrixes for (a) MTE, (b) MTE-spatial, and (c) MTE-temporal.

4.3 下游任务实验

■ 人口密度估计

- 数据：来自WorldPop1 2013年人口密度数据，6789个真值数据
- 表现：

Table 2. Performance of population density estimation. Units for RMSE and MAE are number of people/km².

Model	$R^2 \uparrow$	RMSE \downarrow	MAE \downarrow	MAPE \downarrow
Traj2Vec	0.669 \pm 0.023	5355.76 \pm 322.30	2996.36 \pm 75.13	0.793 \pm 0.066
MNE	0.691 \pm 0.020	5165.66 \pm 259.55	2912.74 \pm 77.24	0.746 \pm 0.06
HDGE	0.705 \pm 0.021	5080.28 \pm 292.38	2802.99 \pm 78.36	0.711 \pm 0.062
HIER	0.506 \pm 0.025	6557.95 \pm 329.47	3740.22 \pm 88.79	1.134 \pm 0.083
Sk-3views	0.644 \pm 0.025	5513.84 \pm 341.16	3118.17 \pm 86.53	0.846 \pm 0.062
Graph-3views	0.680 \pm 0.026	5273.55 \pm 334.89	2858.80 \pm 78.01	0.673 \pm 0.049
MTE-temporal	0.313 \pm 0.023	7692.49 \pm 292.24	4751.09 \pm 94.05	1.574 \pm 0.114
MTE-spatial	0.827\pm0.016	3863.49\pm244.08	2132.52\pm71.66	0.467\pm0.036
MTE	0.816 \pm 0.015	3987.62 \pm 243.51	2226.17 \pm 65.64	0.493 \pm 0.037

- MTE表现远高于其他模型 $R^2 > 0.8$
- MTE-spatial产生的效果略好于MTE；表明空间邻接性比时间相似性更能说明人口密度的估计；加入时间视图只能降低性能
 - 人口密集或稀少的地区通常聚集在一起；时间规律有助于区分土地利用类型，却不能反映不同人口密度水平的居住区
 - 例如，一个地区由高层公寓(人口密集)，另一个地区有独立住宅(人口稀疏)，尽管时间规律相似，但人口密度差异很大

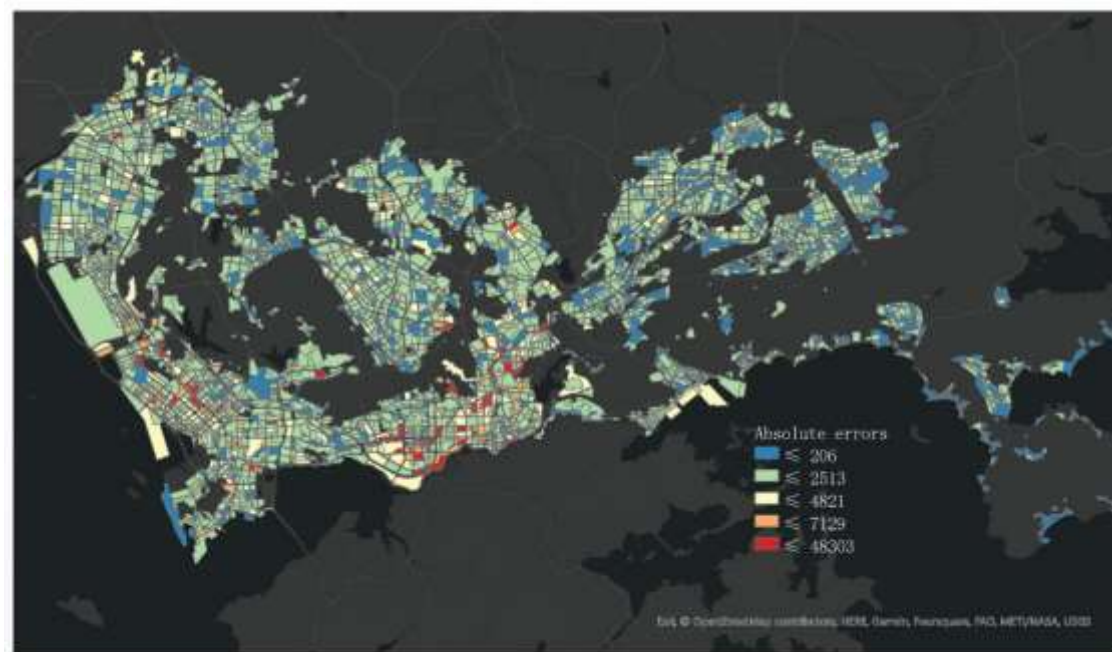


Figure 7. Visualization of absolute errors of population density estimation on a map. Unit is number of population/km².

4.3 下游任务实验

■ 房价预测：流动性数据已被证明对房价建模是有效的(Kang et.al 2021)

- 数据：来自链家2020年房价数据作为真值数据(970个)
- 表现：

Table 3. Performance of house price prediction. Units for RMSE and MAE are CNY/m².

Model	R ² ↑	RMSE↓	MAE↓	MAPE↓
Traj2Vec	0.532±0.043	16678.13±1310.83	11591.79±528.89	0.203±0.012
MNE	0.561±0.040	16184.94±1533.87	11240.92±572.29	0.195±0.012
HDGE	0.552±0.045	16454.15±1556.21	11090.67±539.56	0.188±0.010
HIER	0.460±0.038	18062.00±1470.42	12735.77±557.24	0.229±0.015
Sk-3views	0.506±0.041	17292.47±1278.74	12022.85±543.42	0.212±0.013
Graph-3views	0.430±0.053	18578.60±1663.10	12969.44±615.66	0.229±0.014
MTE-temporal	0.073±0.041	23704.17±1366.91	17862.36±663.14	0.324±0.018
MTE-spatial	0.576±0.047	15958.75±1498.07	10523.48±510.15	0.178±0.011
MTE	0.579±0.044	15734.94±1380.20	10502.69±501.10	0.179±0.011

- 从迁移视图和空间视图获得的信息比时间视图更有效
- 但时间视图也有一定的有效性，因为MTE具有比MTE-spatial更佳的性能
- 时间视图能够带来边际性能增益

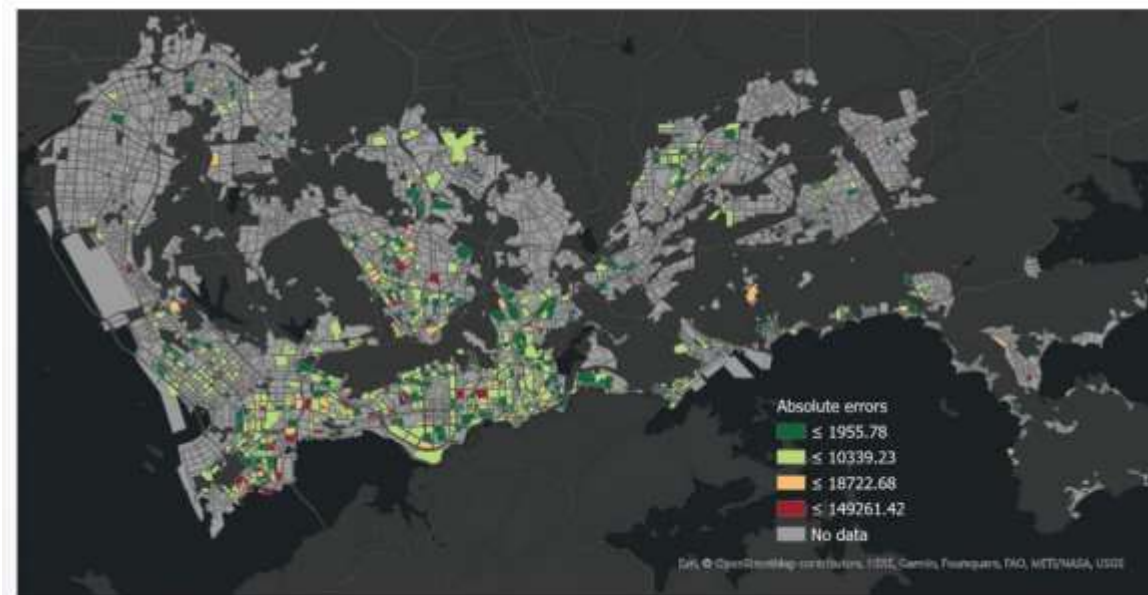


Figure 8. Visualization of absolute errors of house price prediction on a map. Unit is CNY/m².

4.3 下游任务实验

■ 超参数调优

- 迁移视图中调整窗口大小 w
- 空间视图的 k_S 与时间视图中的 k_T

Table 4. Parameter analysis in terms of the window size w in the transition view. LU-F1 denotes the weighted F1 score in the land use classification task, PD-MAE denotes MAE in the population density estimation task, and HP-MAE denotes MAE in the house price prediction task. These three columns are in the form of metric (rank).

Window size w	LU-F1 \uparrow	PD-MAE \downarrow	HP-MAE \downarrow	Average rank \downarrow
3	0.505(5)	2951.830(1)	11885.103(3)	2.33
5	0.506(4)	3018.193(4)	11875.263(2)	3.67
7	0.507(3)	3048.656(5)	11918.788(4)	4.33
9	0.513(1)	2982.641(2)	11591.789(1)	1.67
11	0.509(2)	3001.700(3)	12342.406(5)	3

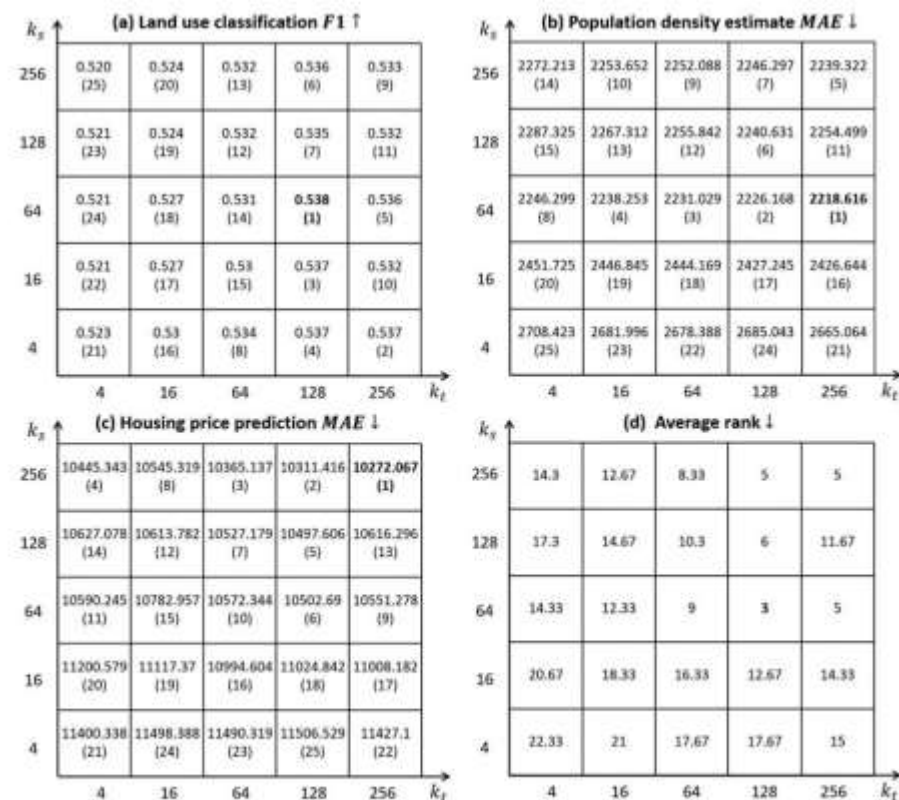


Figure 9. Parameter sensitivity analyses in terms of k_s (for the spatial graph) and k_t (for the temporal graph), i.e. the k number in constructing KNN graphs.

- 在三个下游任务中对参数调优进行加权，选择最佳组合
- $w=9$, $k_S=64$, $k_T=128$ 在三个下游任务之间产生平衡的最佳性能
- 在迁移视图中存在长期依赖关系，而在空间和时间视图中存在中等的短期依赖关系

5 讨论

■ 经过实验结果，以下三种假设成立

1. 人类轨迹携带的三种信息视图在不同的下游任务中各有优势；结合起来会产生最有利的结果
2. 建模远程依赖关系有利于建模迁移视图
3. 在时间和空间试图中对近邻依赖关系进行建模是可取的

□ 不同视角应当对特定的下游任务进行加权：

1. 迁移视图在所有任务中都很重要，是学习区域嵌入不可缺少的部分
2. 空间视图在土地利用分类任务中效果不如时间视图
3. 时间视图对土地利用分类有利，对房价预测有利，但不适用于估计人口密度
4. 为我们使用简单的连接来组合时间和空间图嵌入的方式提供解释；而不是使用复杂的融合手段(对比学习等)

□ 比较不同类型的数据在不同的下游任务中的有效性：

在不同类型的地理空间数据用于城市感知任务时，应该考虑不同类型地理空间数据的适应度

- 如何自适应合理选取合适的模式，或者让每种数据模式在不同任务中的权重不同

□ 预训练方法值得研究：

诸多城市交通研究(e.g. Li, Xia, et al. 2024; Zhang, Gong, Zhang, et al. 2023). 表明预训练方法在预测任务中是有效的

6 结论

■ 提出一种MTE方法用完全无监督的方式学习具有人类轨迹的有效区域表示

- 远程依赖关系在迁移视图中是有意义的：因此本文对迁移视图使用Skip-Gram学习远程依赖
- 时空视图的短期依赖关系更可取：对因此利用无监督图表示学习来学习两个视图的嵌入
- 下游任务的广泛实验(土地利用分类、人口密度估计、房价预测)：本文观察到不同视图的有效性在特定下游任务中存在差异
- 提出讨论：本研究提供了各种深度学习技术对地理空间数据所携带的不同信息视图建模的适应性的见解

谢谢！

