

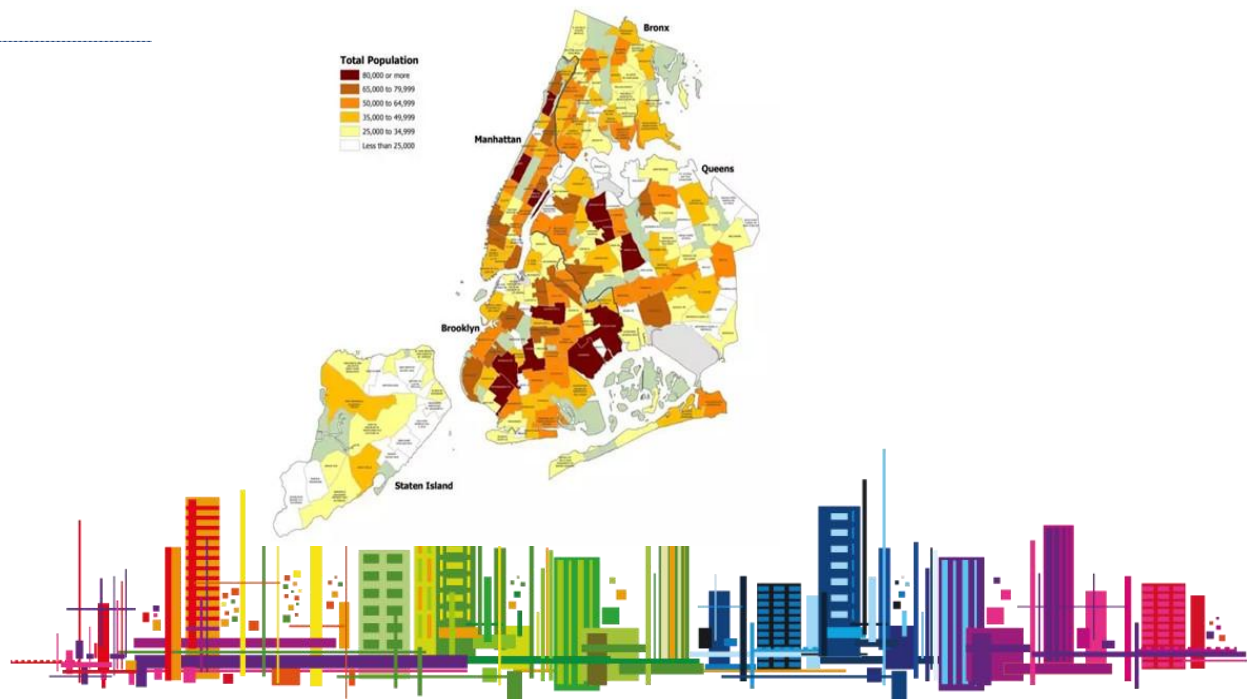
城市区域表示学习

主讲人：杨宜凡

19373750@buaa.edu.cn



- 统计数据显示，截至目前超过54%的世界人口居住在城市地区。随着城市人口的增加和面积的扩大，城市的管理会越来越复杂和困难。
- 随着城市化进程的加速，物联网传感器，GPS终端，智能手机所收集到的城市时空数据爆炸性增长。
- 有了数据的支撑，许多数据驱动的深度学习方法都有了快速的发展，它们为破解城市管理面临的问题提供了全新的解决思路。



- 城市由各种各样的地区组成，人们在那里工作、学习、娱乐和生活，区域是当代城市模型的基本二维空间单位，有不同的划分方式。
- 可以由政府行政划分产生，例如将一个居民社区或人口普查区当作一个区域。
- 可以由地理边界划分，例如由城市街道分割开来的一个个区域。
- 可以是模型自动生成的，例如《Automated identification and characterization of parcels with OpenStreetMap and points of interest》由OpenStreetMap道路网络生成若干区域边界，使用元胞自动机选择此区域是否属于城市，最后使用区域内POI综合得到区域的主要功能。

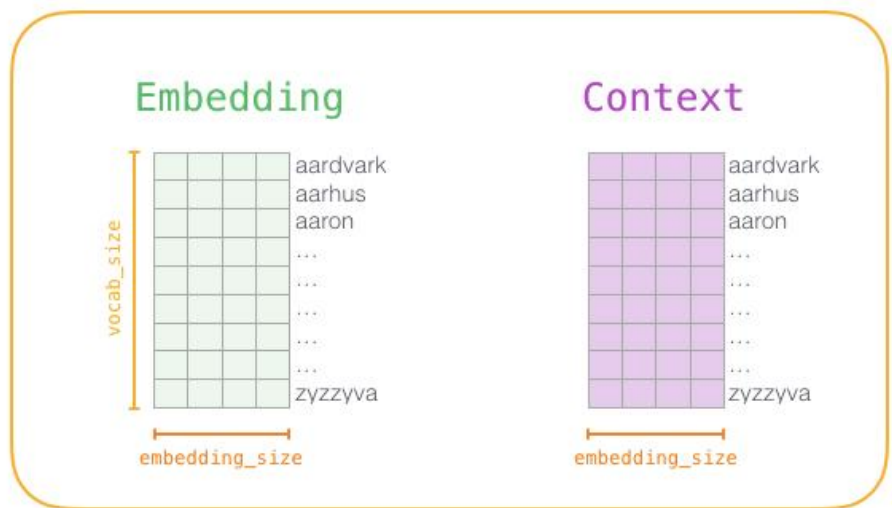


- 研究**城市区域的性质**可以为解决城市管理问题提供必要的信息，如房价预测，犯罪率预测等
- 但是区域本身作为非结构化的数据是不可计算的，因此需要获取高质量的**区域表征向量**作为下游任务的输入，城市区域表征学习正是为了解决这一问题。
- 城市区域可以利用的数据有：
 - 静态数据：
 - 区域内POI种类，分布
 - 区域地理位置数据
 - 动态数据：
 - 人口流动数据
- 主流方法：
 - Word2Vec
 - 自编码器
 - 图神经网络

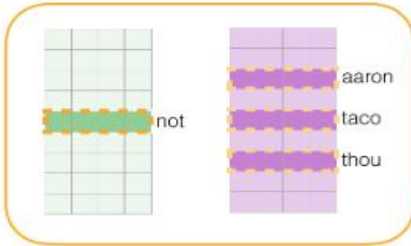


Word2Vec

thou shalt not make a machine in the ...



input word	output word	target	input • output	sigmoid()	Error
not	thou	1	0.2	0.55	0.45
not	aaron	0	-1.11	0.25	-0.25
not	taco	0	0.74	0.68	-0.68



Update Model Parameters



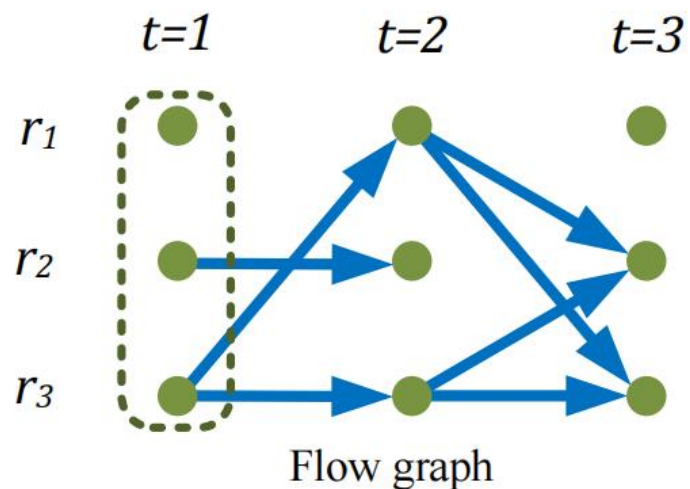
- **动机:**

将区域作为中心词，如何合理地构造上下文来建模区域之间的相关性。

- **构造人流量图**

两区域间人流量量越大，相关性越强。

此图是一个分层图，每一时间片上的每一个区域都作为节点，边权为本时间片上去到另一个节点的人流量。可以从图中抽取很多路径作为一条条轨迹。

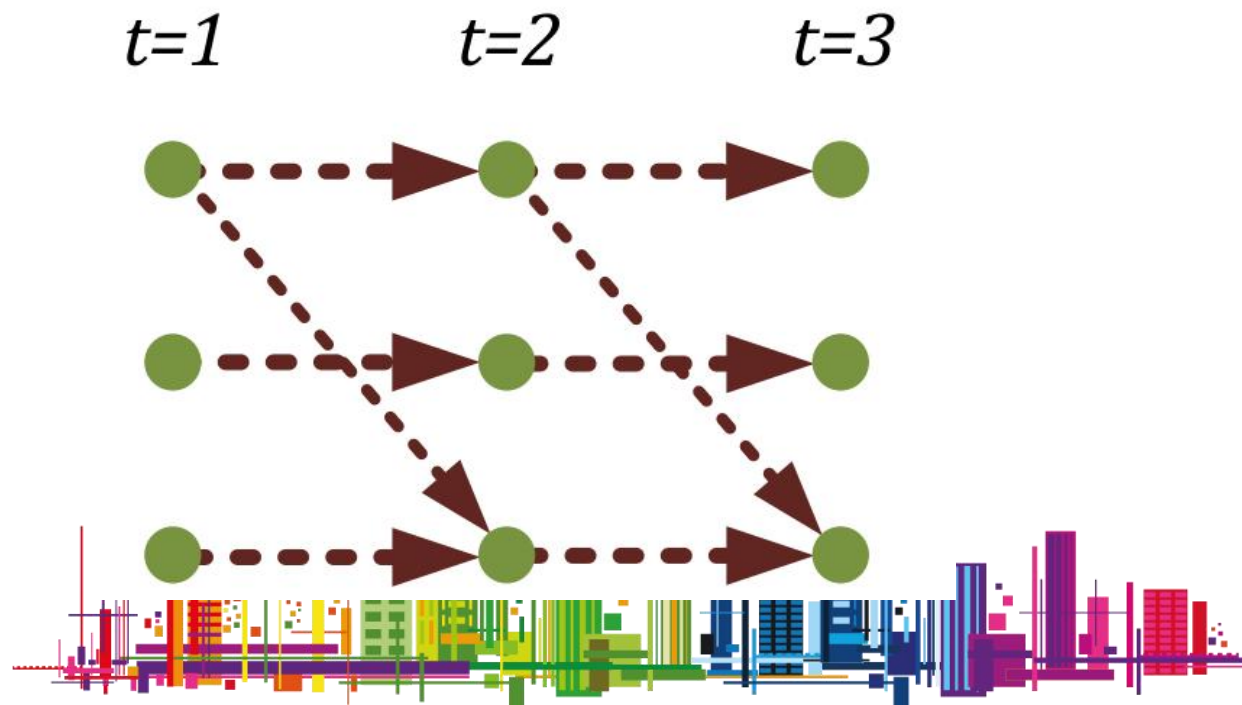


- **构造地理空间图，同时解决了轨迹稀疏带来的某些节点无法表征的问题。**

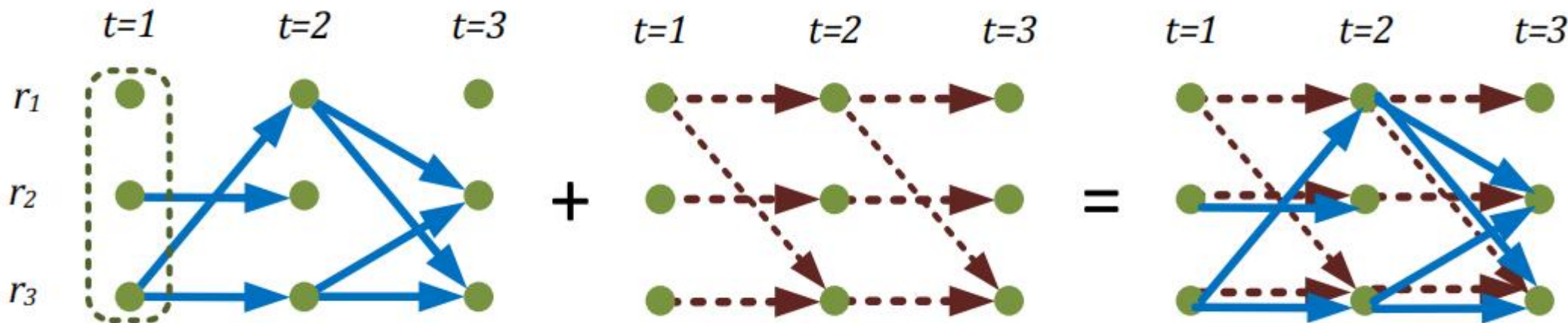
人类的流动性是受空间限制的，人们出现在不同地区的概率与他们需要旅行的距离成负相关。

节点集与流图一致，边权重与两个区域的质心距离负相关。

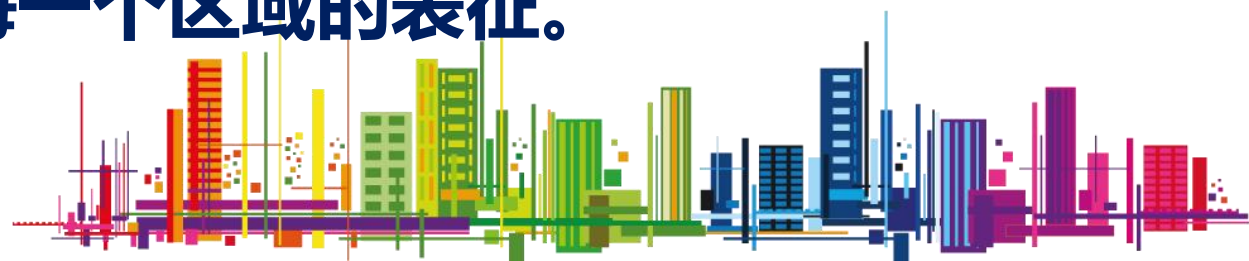
$$g_{ij} = \exp(-C \cdot d_{ij})$$



- 将两张图叠加，在上面做随机游走，生成的路径作为句子，每一个时间片上的区域作为中心词。



- 使用skip-gram网络来训练得到每一个区域的表征。



• 实验评估

下游采用了一个负二项回归模型, y_i 代表第*i*个区域的目标变量, 引入了该区域的辅助特征 (如人口统计数据 and 平均收入), 记为 X_i , 并引入了邻接区域的影响, $\text{sim}(i,j)=u_i^T u_j$

$$y_i = \exp(\alpha \cdot X_i + \beta \sum_{j \in N_i} \text{sim}(i,j) \cdot y_j + \gamma).$$

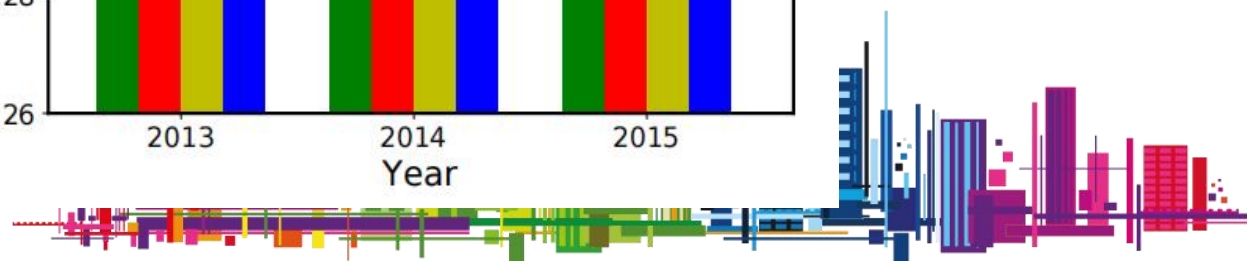
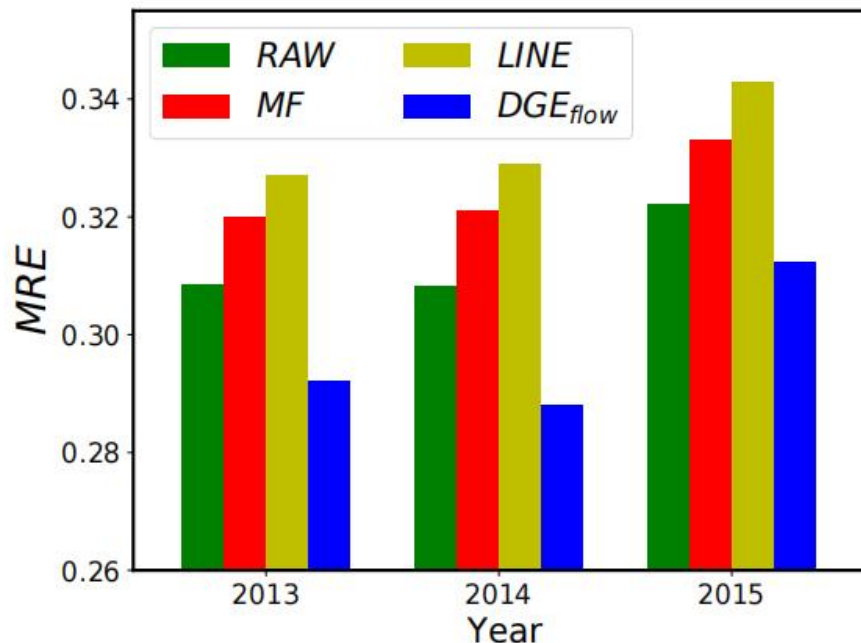
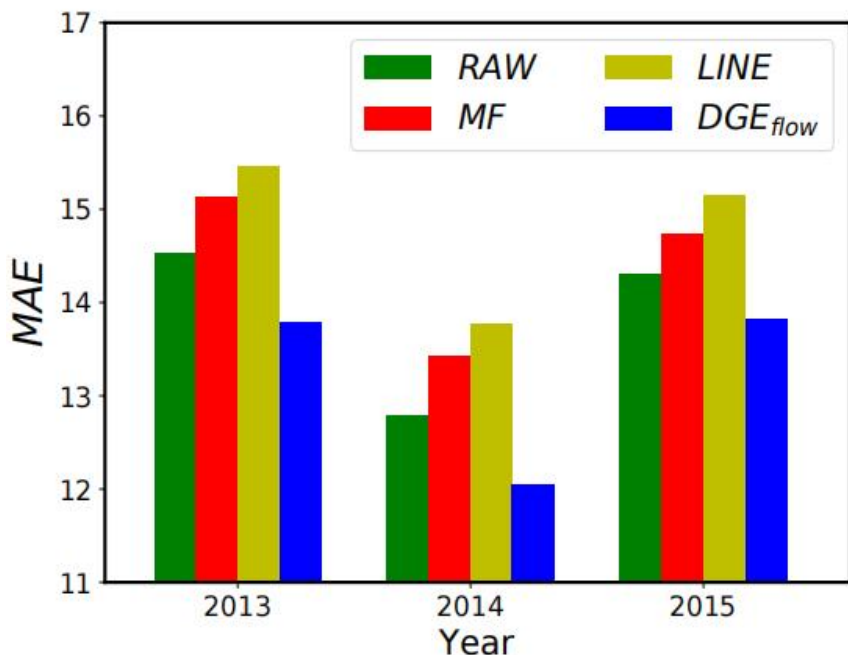
评估指标为MRE,MAE

$$MRE = \frac{1}{T} \sum_{t=1}^T \frac{\sum_i^n |y_{it} - \hat{y}_{it}|}{\sum_i^n y_{it}} \quad MAE = \frac{1}{T} \sum_{t=1}^T \sum_i^n |y_{it} - \hat{y}_{it}|$$



• 实验评估

下游任务1: 区域犯罪率回归, 辅助特征选择{人口特征, POI特征}, baseline选择LINE, MF (矩阵分解), RAW (不使用嵌入技术来计算相关性, 使用交通流量和空间距离做相关性度量)



• 实验评估

下游任务2：区域平均收入回归

Method	MAE	MRE
<i>RAW</i>	15127	0.250
<i>MF</i>	16674	0.2756
<i>LINE</i>	15534	0.2567
<i>DGE_{flow}</i>	-	-
<i>HDGE</i>	14740	0.2436

下游任务3：区域平均房价回归

Method	MAE	MRE
<i>RAW</i>	39.28	0.229
<i>MF</i>	39.83	0.233
<i>LINE</i>	40.438	0.236
<i>DGE_{flow}</i>	38.95	0.226
<i>HDGE</i>	-	-

HDGE考虑了区域表征的时间动态性；
并且在构建的流图考虑了轨迹的多跳转换，进而捕捉到了区域间的多跳相关性；
同时考虑地理位置和流量，建模的信息丰富，性能比较有优势。



自编码器

自编码器是一种无监督学习方法，是一种利用反向传播使得输出值等于输入值的神经网络，它先将输入压缩成潜在的低维空间表征，然后通过这种表征来重构输出。这样学到的潜在空间表征就代表了输入数据本身的一些性质。

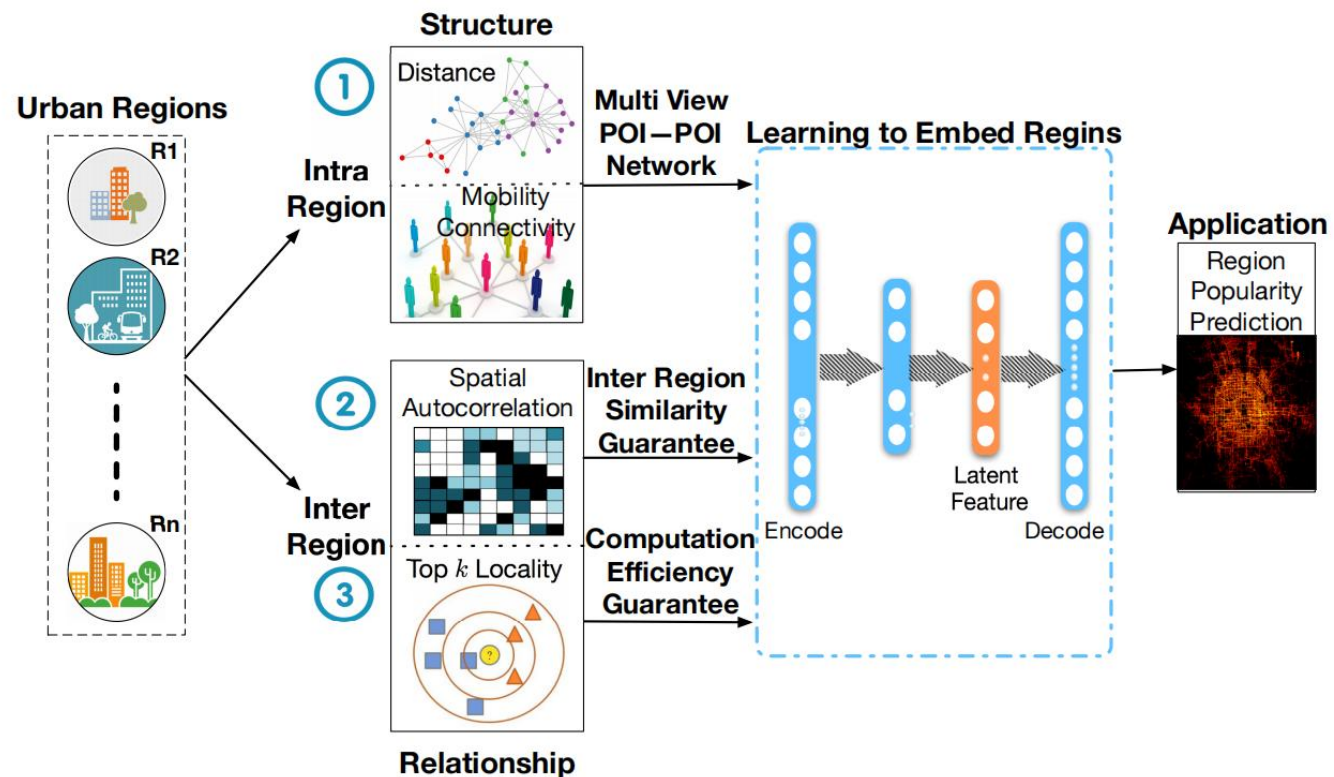


Efficient Region Embedding with Multi-View Spatial Networks: A Perspective of Locality-Constrained Spatial Autocorrelations Embedding

- **动机:**

如何同时考虑区域间相关性和区域内结构信息

- **模型框架:**

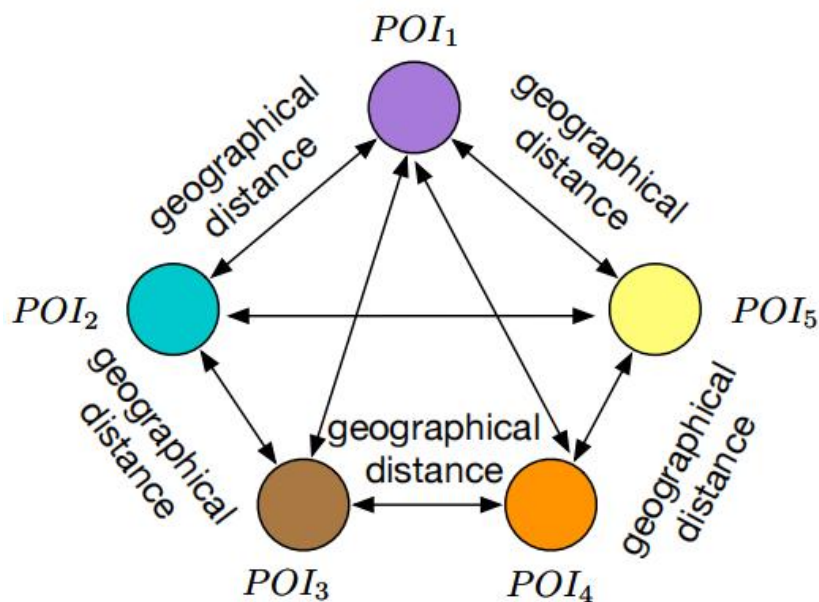


• 区域内结构：构建区域内多视图POI-POI网络

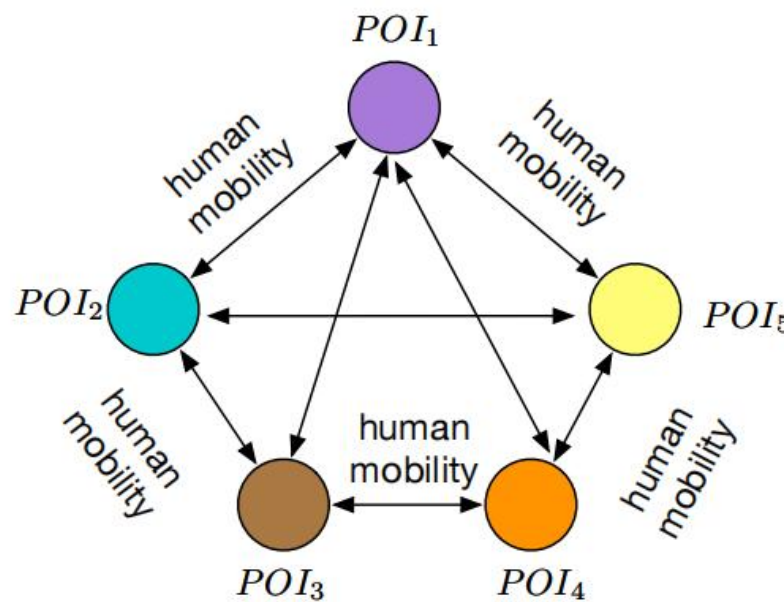
将每一个类别的POI作为节点，一个区域一张图。

基于平均距离的POI网络：边权重为两个类别POI的平均距离

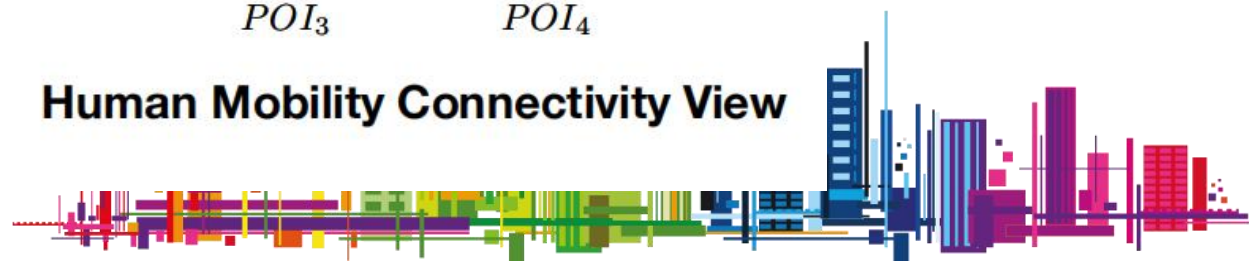
基于人口流动的POI网络：边权重为两个类别POI的人流量



Geographical Distance View



Human Mobility Connectivity View



• 自编码器模型

$$\begin{cases} (\mathbf{y}_i^\phi)^1 &= \sigma((\mathbf{W}^\phi)^1 \mathbf{x}_i^\phi + (\mathbf{b}^\phi)^1), \\ (\mathbf{y}_i^\phi)^k &= \sigma((\mathbf{W}^\phi)^k (\mathbf{y}_i^\phi)^{k-1} + (\mathbf{b}^\phi)^k), \forall k \in \{2, 3, \dots, o\}, \\ \mathbf{z}_i^\phi &= \sigma((\mathbf{W}^\phi)^{o+1} (\mathbf{y}_i^\phi)^o + (\mathbf{b}^\phi)^{o+1}). \end{cases}$$
$$\begin{cases} (\hat{\mathbf{y}}_i^\phi)^o &= \sigma((\hat{\mathbf{W}}^\phi)^{o+1} \mathbf{z}_i^\phi + (\hat{\mathbf{b}}^\phi)^{o+1}), \\ (\hat{\mathbf{y}}_i^\phi)^{k-1} &= \sigma((\hat{\mathbf{W}}^\phi)^k (\hat{\mathbf{y}}_i^\phi)^k + (\hat{\mathbf{b}}^\phi)^k), \forall k \in \{2, 3, \dots, o\}, \\ \hat{\mathbf{x}}_i^\phi &= \sigma((\hat{\mathbf{W}}^\phi)^1 (\hat{\mathbf{y}}_i^\phi)^1 + (\hat{\mathbf{b}}^\phi)^1). \end{cases}$$

$$\mathcal{H}(\mathcal{R}) = \frac{1}{2} \sum_{\phi \in \Phi} \sum_{r_i \in \mathcal{R}} \|(\mathbf{x}_i^\phi - \hat{\mathbf{x}}_i^\phi)\|_2^2$$



• 综合区域间相关性

使用一个矩阵 \mathbf{Q} 来参数化空间自相关，并将参数矩阵的学习纳入自动编码器。

$$(\mathbf{Q}_i^\phi)^* = \sum_{j \neq i, r_j \in \mathcal{R}} \mathbf{Q}_{i,j}^\phi \mathbf{z}_j^\phi.$$

$$\begin{cases} (\mathbf{y}_i^\phi)^1 & = \sigma((\mathbf{W}^\phi)^1 \mathbf{x}_i^\phi + (\mathbf{b}^\phi)^1), \\ (\mathbf{y}_i^\phi)^k & = \sigma((\mathbf{W}^\phi)^k (\mathbf{y}_i^\phi)^{k-1} + (\mathbf{b}^\phi)^k), \forall k \in \{2, 3, \dots, o\}, \\ (\mathbf{y}_i^\phi)^{o+1} & = \sigma((\mathbf{W}^\phi)^{o+1} (\mathbf{y}_i^\phi)^o + (\mathbf{b}^\phi)^{o+1}), \\ \mathbf{z}_i^\phi & = (\mathbf{y}_i^\phi)^{o+1} + (\mathbf{Q}_i^\phi)^*. \end{cases}$$



- top-k邻域性能优化

综合所有其他区域的表征需要的参数和计算代价都较高，因此只考虑综合相关性最高的k个区域。

度量相关性从距离和功能两个角度出发，距离即为区域质心距离，记为 $D_{i,j}$

功能相关性可以采用两个区域功能向量的余弦相似度来计算，记为 $F_{i,j}$

总相关性定义为

$$S_{i,j} = \frac{1}{D_{i,j}} \times F_{i,j}.$$

只考虑相关性最高的k各区域进行结果的综合

$$\mathbf{z}_i^\phi = (\mathbf{y}_i^\phi)^{o+1} + \sum_{j \neq i, r_j \in (\mathcal{R}_i)^{TopK}} \mathbf{Q}_{i,j}^\phi \mathbf{z}_j^\phi$$



• 实验评估

下游模型选择基本的线性回归模型

$$P_i = \mathbf{W}^T \mathbf{u}_i + \mathbf{b}_i$$

指标选择:

平方误差SE: $SE = \sum_i (y_i - f_i)^2$

相关系数: $R^2 = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2}$

肯德尔相关系数 $\text{Tau} = \frac{\#conc - \#disc}{\#conc + \#disc}$

#conc指的是预测出来的大小关系和实际相同的区域对数, #disc指的是预测出来的大小关系和实际不同的区域对数

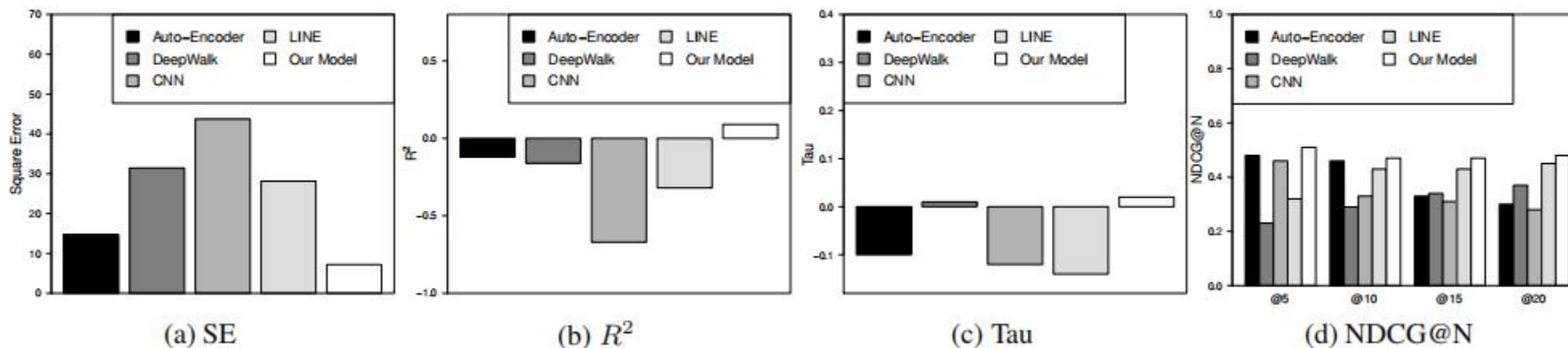
归一化折损累计增益NDCG:

$$\sum_{i'=1}^N \frac{y_{i'}}{\log_2(1+i')} / \sum_{i=1}^N \frac{y_i}{\log_2(1+i)}$$

i 代表实际标签的排序, i' 代表预测出来结果的排序



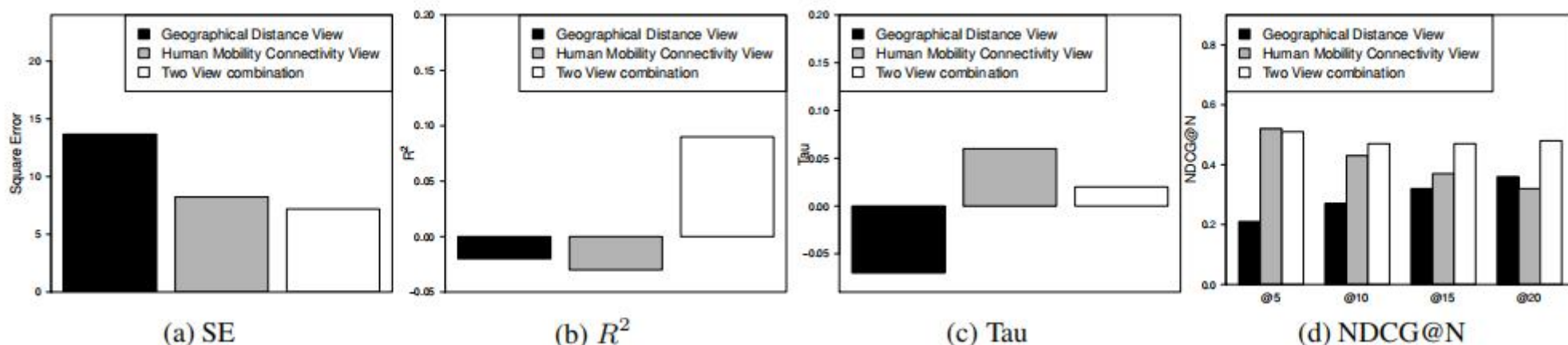
• 实验评估



baseline没有考虑区域间空间自相关性，因此证明了综合建模区域内的性质和空间自相关是有意义的。



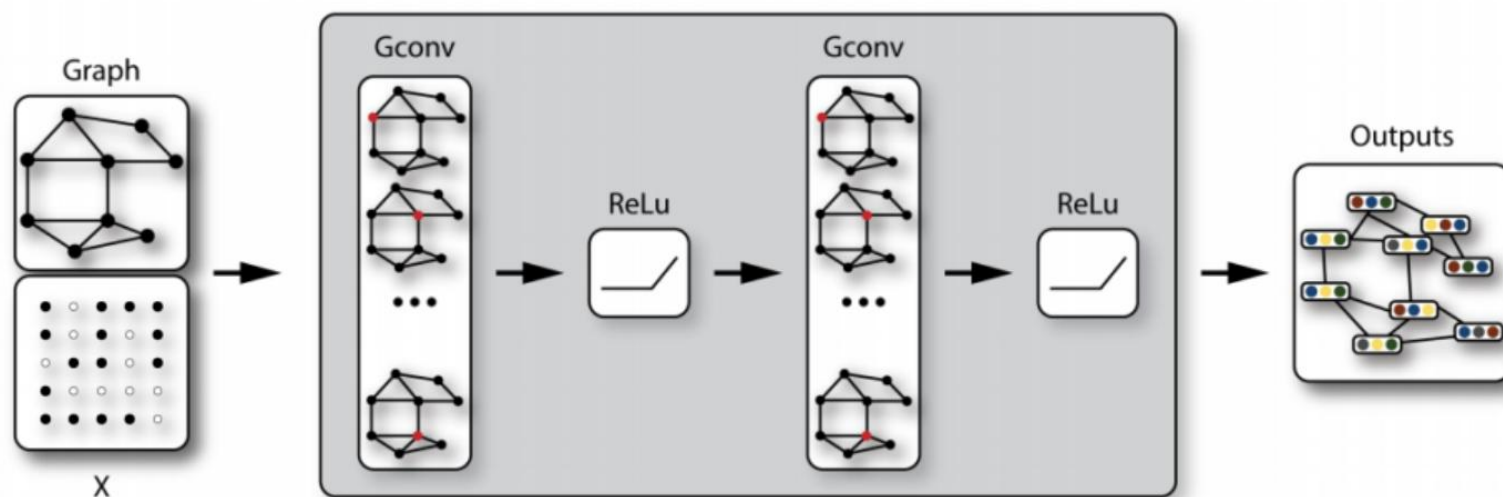
• 实验评估



测试了在建模区域内结构时不同视图的不同效果。单个视图的性能比这两个视图的组合更差。而且，人类移动性连接视图的性能略优于地理距离视图。一种可能的解释是，人类的移动性比静态的地理距离信息具有更丰富的语义信息。另一方面，单个视图不能捕获区域的完整模式。这两种观点的结合补充了每一种单一观点的不足。



图神经网络



70538854-01_2021.02.14

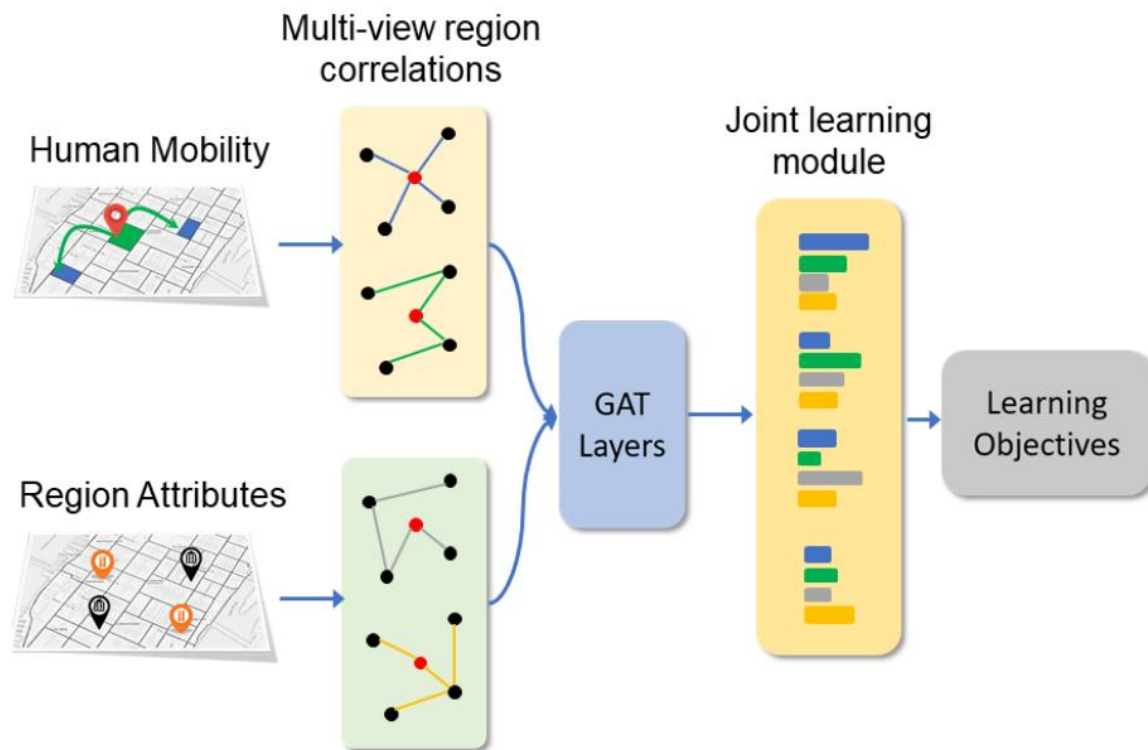


Multi-View Joint Graph Representation Learning for Urban Region Embedding

- **动机:**

如何建模多视图+如何将每一张图上学到的embedding有效地组合。

- **模型框架:**



• 如何建模多视图

• 基于人口流动数据来建模

接收同一出发点来的人流或者将人流发送到同一目的地的若干区域往往具有着相关性。

因此利用节点的入度分布和出度分布定义一个区域的上下文。

$$p_s(r|r_i) = \frac{w_{r_i}^r}{\sum_r w_{r_i}^r}, \quad p_d(r|r_i) = \frac{w_r^{r_i}}{\sum_r w_r^{r_i}}.$$

$$C_s^{ij} = \text{sim}(p_s(r|r_i), p_s(r|r_j)),$$

$$C_d^{ij} = \text{sim}(p_d(r|r_i), p_d(r|r_j)).$$



- 如何建模多视图
 - 基于区域属性来建模

$$C^{ij} = \text{sim}(\vec{a}_i, \vec{a}_j).$$

使用POI属性来建模：利用TF-IDF模型，将一个POI作为单词，将一个区域作为句子。计算得到每一种POI对此区域的重要性，组合成POI属性向量。

使用签到数据来建模：与上类似，采用TF-IDF模型计算得到每一种类型的签到数据作为签到属性的重要性。



• 学习单视图的表征

采用典型的Graph Attention Network来学习节点表征：聚合邻居的特征，权重通过注意力机制得到。

$$e_{ij} = \exp \left(\text{ReLU} \left(\vec{\mathbf{a}}^T \left[\mathbf{W} \vec{h}_i \parallel \mathbf{W} \vec{h}_j \right] \right) \right),$$

$$\alpha_{ij} = \text{softmax}_j (e_{ij}) = \frac{\exp(e_{ij})}{\sum_{k \in N_i} \exp(e_{ik})},$$

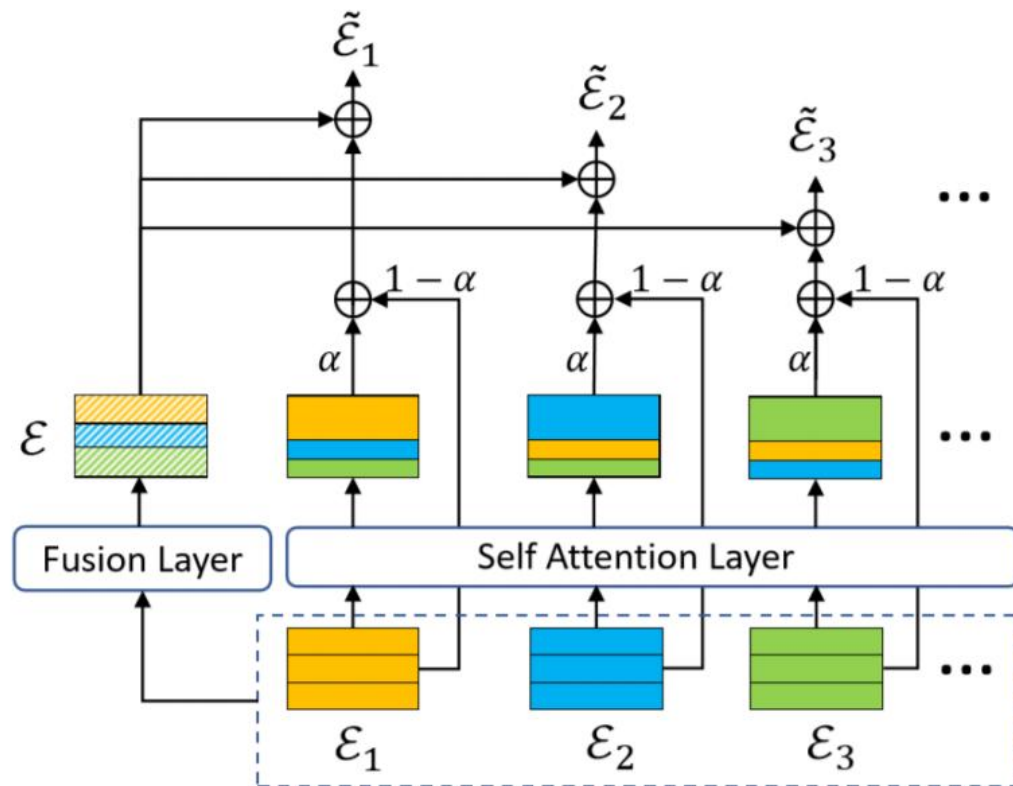
$$\vec{h}'_i = \sigma \left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W} \vec{h}_j \right),$$



• 联合学习模块

第一部分允许通过自注意力机制在所有视图之间共享信息。

第二部分是一个融合层，它通过学习自适应权值来结合多视图表示。



• 视图间信息共享

整合多视图的信息可以增强每一个单一视图的学习过程

信息共享采用基本的点积注意力机制来建模共享的权重

第*i*个视图学到的表征矩阵为 \mathcal{E}_i

$$K_i = \mathcal{E}_i W_k, \quad Q_i = \mathcal{E}_i W_q.$$

$$[A_i]_{i=1}^M = \text{softmax} \left(\left[\frac{Q_i K_i^T}{\sqrt{k}} \right]_{i=1}^M \right), \quad \hat{\mathcal{E}}_i = \sum_{i=1}^M A_i \mathcal{E}_i.$$

$$\mathcal{E}'_i = \alpha \hat{\mathcal{E}}_i + (1 - \alpha) \mathcal{E}_i, \quad 0 \leq \alpha \leq 1,$$



- 融合多个视图的结果

采用一组可学习的权重，采用一个全连接的神经网络，输入为表征，输出为权重。

$$\mathcal{E} = \sum_i^M w_i \mathcal{E}_i, \quad w_i = \sigma(\mathcal{E}_i W_f + b_f).$$

将上述两层的结果求平均得到最终的表征

$$\tilde{\mathcal{E}}_i = (\hat{\mathcal{E}}_i + \mathcal{E}) / 2.$$



- 优化目标
- 定义了两个任务：
 - OD预测：给定出发地根据表征预测目的地，或者相反。

$$\hat{p}_s(r_j|r_i) = \frac{\exp(e_s^i T e_d^j)}{\sum_j \exp(e_s^i T e_d^j)}, \quad \hat{p}_d(r_j|r_i) = \frac{\exp(e_d^i T e_s^j)}{\sum_j \exp(e_d^i T e_s^j)}$$

$$\mathcal{L}_{mob} = \sum_{(r_i, r_j) \in \mathcal{M}} -\log \hat{p}_s(r_j|r_i) - \log \hat{p}_d(r_i|r_j).$$

- 区域相关性重建：利用表征做点积重建区域间相似度视图，在POI属性和签到属性上各做一次

$$\mathcal{L}_{poi} = \sum_{i,j} (C_{poi}^{ij} - e_{poi}^i T e_{poi}^j)^2.$$

- 相加后得到总的优化目标

$$\mathcal{L} = \mathcal{L}_{mob} + \mathcal{L}_{poi} + \mathcal{L}_{chk}.$$

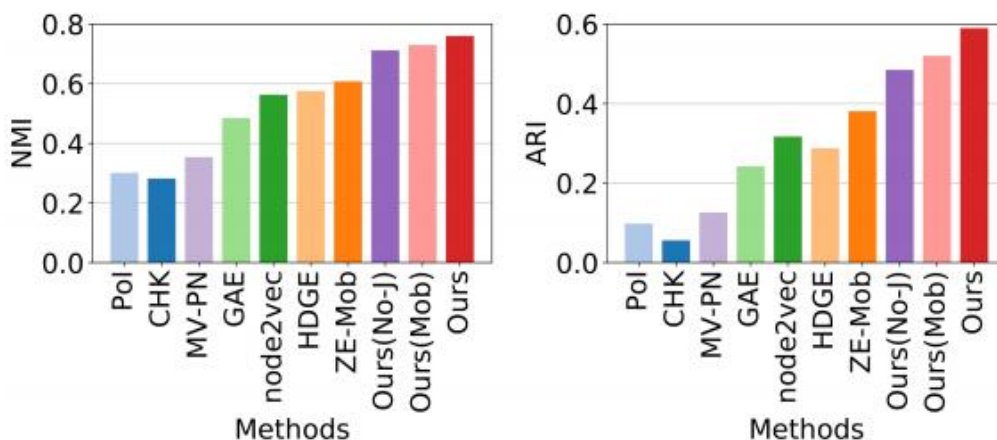


• 实验评估

下游任务1:

土地使用类型分类, 使用K-means聚类根据表征对区域进行聚类, 使用城市社区划分的社区功能作为标签。

评价指标为归一化互信息NMI和标准兰德系数ARI



考虑多视图相关的要比仅有单视图的一般效果要好(POI,CHK效果

最差, Mob单视图效果也要差于Mob+Poi+Chk)

多视图的简单组合并不能充分利用多视图的信息, 甚至有可能效果更差 (NO-J)

MVURE的多视图建模和融合机制是效果最好的



- **提升模型性能：**如何充分利用多种城市数据，同时考虑并综合区域间相关性和区域内性质，通常的思路是构造多张异构图分开学习并找到一个好的方法将异构图的结果综合起来。
- **学习随时间变化的表征：**随着时间的变化，城市区域的功能可能也会变化。大多数现有的工作只专注于学习静态区域嵌入。然而，这些并不能揭示不同时间背景下区域间的空间相互作用。
- **构建统一的区域表示学习框架：**截至目前的文章中仅有不足四分之一有相关代码的开源。而且普遍缺乏数据处理的代码和原始数据，研究者很难理解被处理后数据的含义和内容。为后人带来了复现困难，比较困难，创新困难的问题。



感谢聆听，请各位老师同学指正!

