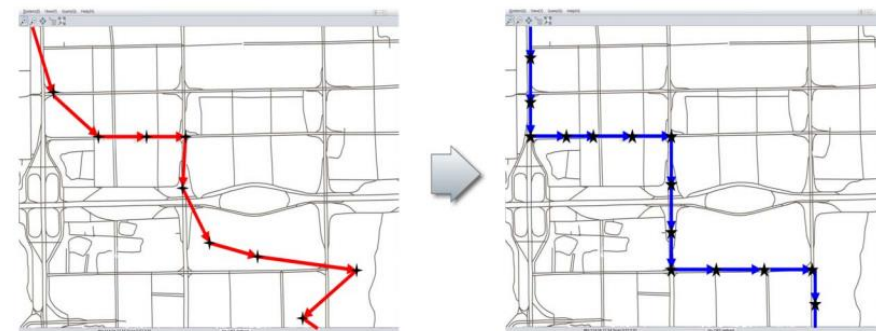


轨迹质量增强技术调研



问题定义

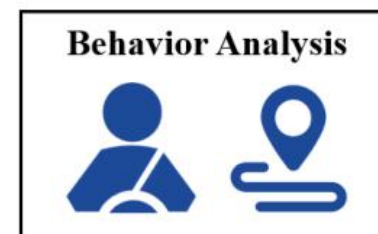
- 轨迹来源：GPS定位、手机信令数据定位等
- 轨迹质量增强：输入原始的低质量轨迹，输出更精确的高质量轨迹
- 轨迹质量的主要衡量标准：与真实位置的偏差(噪声)、轨迹分辨率



轨迹质量增强的意义

在实际应用中，很多场景需要提供精确度更高的轨迹数据

- 基于位置的服务(LBS)：出行导航、出行时间估计、驾驶员行为分析等
- 城市管理和规划：分析城市交通情况、人口分布等，指导交通管理和城市建设
- 学术研究：时空数据挖掘期望优质的数据集



- **基于轨迹上下文的增强技术**

- **神经网络和卡尔曼滤波混合模型——DHTR**
- **基于BERT的轨迹建模方法——TrajBERT**

- **基于路网的增强技术**

- **基于隐马尔可夫模型(HMM)的路网匹配**
- **深度学习路网匹配模型——DeepMM**
- **深度学习路网匹配模型——DMM**
- **端到端的轨迹恢复——MTrajRec**

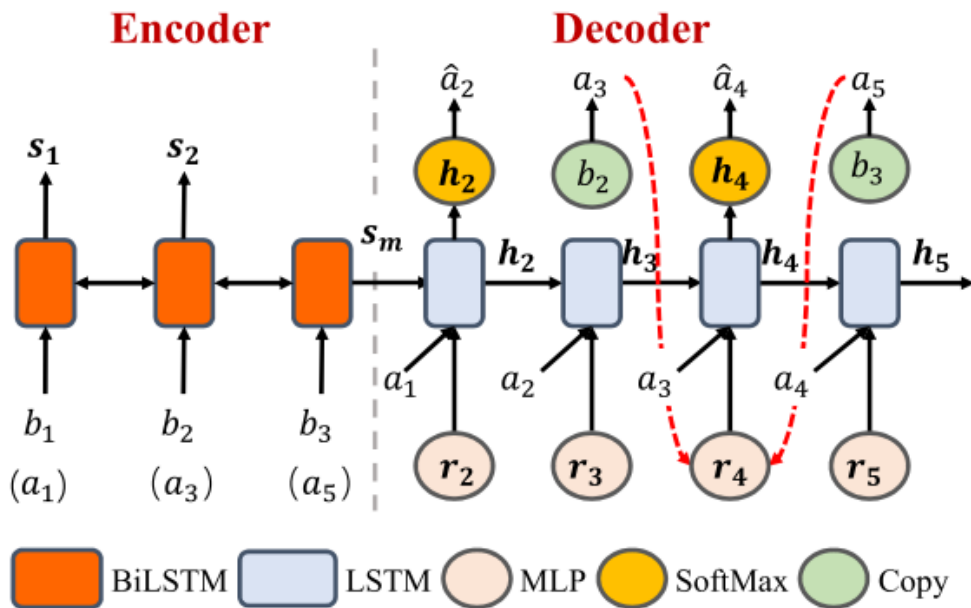
基于轨迹上下文的增强技术



神经网络和卡尔曼滤波混合模型——DHTR

输入：基于网格ID的轨迹，部分点有缺失

输出：补全后采样时间均匀的轨迹



神经网络部分——基于seq2seq模型

- Encoder: 双向LSTM
- Decoder: 单向LSTM建模序列依赖
- MLP: 通过引入前驱和后继限制插值点的范围
- ST-Attention: 时空注意力机制

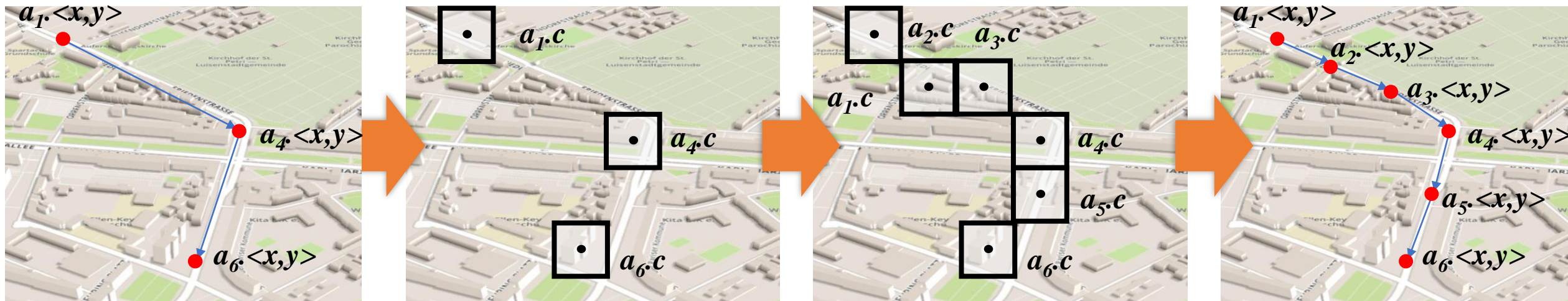
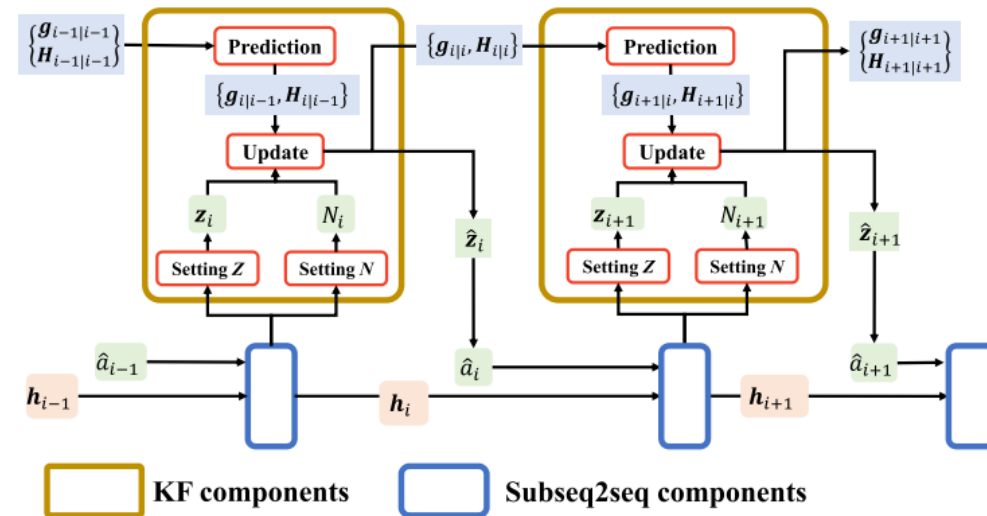
$$u_{i,k} = v^T \tanh \left(W^H h_i + W^S s_k + W^P p_{d_{i,k}^{(1)}} + W^Q q_{d_{i,k}^{(2)}} \right)$$

Wang J, Wu N, Lu X, et al. **Deep Trajectory Recovery with Fine-grained Calibration Using Kalman Filter**[J]. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(3): 921-934.

神经网络和卡尔曼滤波混合模型——DHTR

组合卡尔曼滤波器

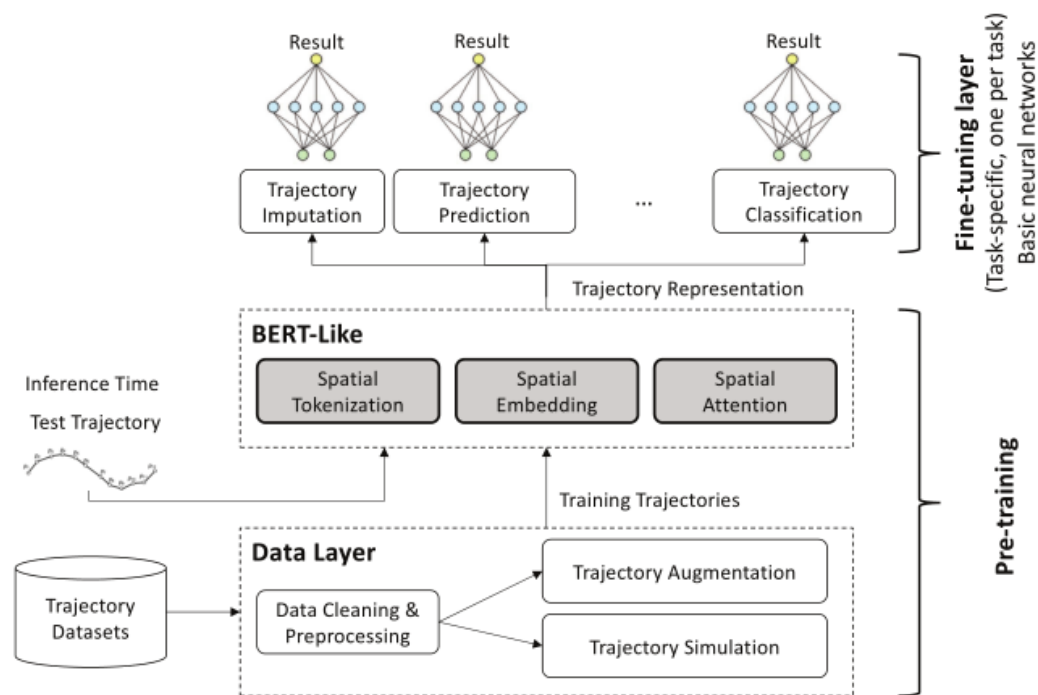
- 观测值由seq2seq模型提供(用网格中心坐标代替), 观测值协方差也通过模型预测概率动态计算
- 协方差状态转移协方差矩阵可训练
- 神经网络和卡尔曼滤波器分开训练



基于BERT的轨迹建模方法——TrajBERT

基于BERT模型的通用轨迹处理策略

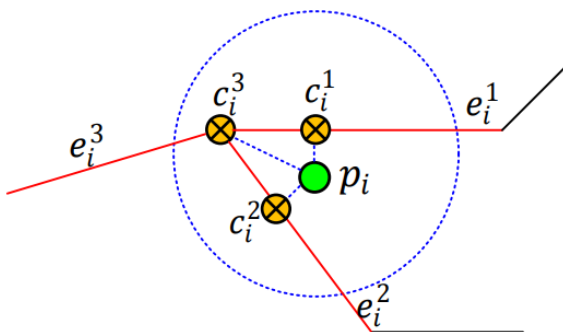
- 离散化：用方格、六边形、基站ID、POI等离散的地理要素来代表轨迹点的位置
- 嵌入：通过Embedding层将离散的轨迹点转换成定长的向量，可以引入地理空间距离辅助训练(区别于NLP)
- 注意力机制：建模轨迹序列中不同点之间的相互影响，引入时间和空间上的限制



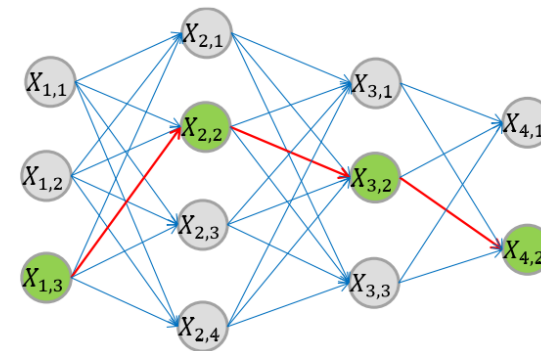
基于路网的轨迹增强技术



基于隐马尔可夫模型(HMM)的路网匹配



$$p(i_t|o_t) = \frac{1}{\sqrt{2\pi}\sigma} e^{-0.5\left(\frac{dist}{\sigma}\right)^2}$$
$$p(d_t) = \frac{1}{\beta} e^{-\frac{d_t}{\beta}}$$



候选路段搜索

给定搜索半径r和候选路段阈值k, 以轨迹点为中心, 查询半径r范围内有交集的路段(使用Rtree加速查询), 返回最优的k个候选路段。

时空关系建模

设置规则计算观测概率和状态转移概率, 构建状态转移图。

最优路径推断

使用维特比算法推断最优路径

基于隐马尔可夫模型(HMM)的路网匹配

- [2009] Hidden Markov Map Matching Through Noise and Sparseness

使用正态分布计算观测概率，使用指数分布计算转移概率

$$p(z_t|r_i) = \frac{1}{\sqrt{2\pi}\sigma_z} e^{-0.5\left(\frac{\|z_t-x_{t,i}\|_{great\ circle}}{\sigma_z}\right)^2} \quad p(d_t) = \frac{1}{\beta} e^{-d_t/\beta}$$

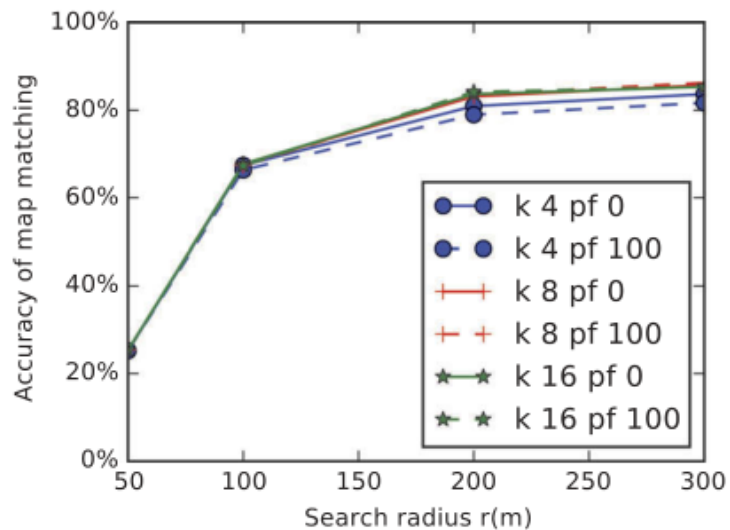
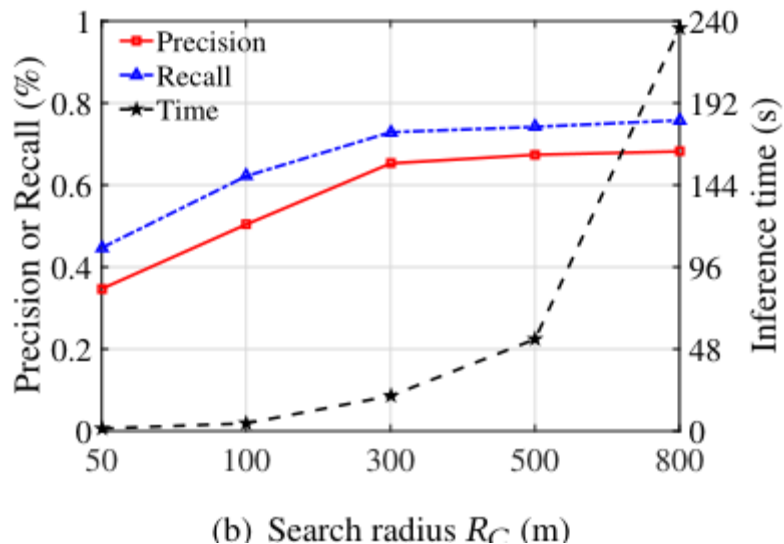
- [2009] Map-Matching for Low-Sampling-Rate GPS Trajectories

提出了全局地图匹配算法ST-Matching，在时空分析的基础上引入了轨迹的时间和速度约束

- [2018] Fast map matching, an algorithm integrating hidden Markov model with precomputation

通过预计算并存储结点之间的最短路径来加速路网匹配，并提供了完善的C++代码库，具有较高的实用价值

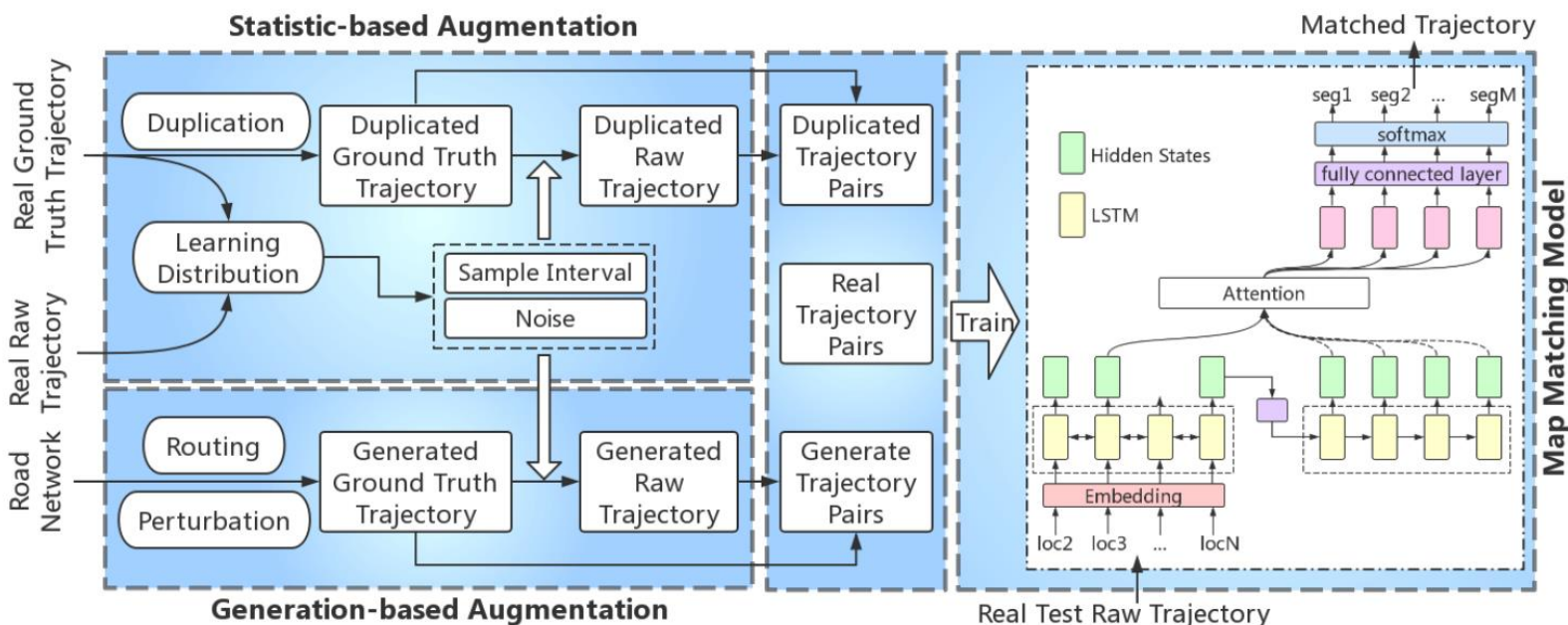
基于隐马尔可夫模型(HMM)的路网匹配



参数调整的影响：

随着搜索半径的增加，运行时间成指数爆炸式增长，但准确率和召回率有边际效应递减的现象

深度学习路网匹配模型——DeepMM



- 数据增强方法
 - 统计式方法
 - 生成式方法
- 神经网络: seq2seq和Attention结合

Zhao K, Feng J, Xu Z, et al. **DeepMM: Deep learning based map matching with data augmentation**[C]//Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. 2019: 452-455.

深度学习路网匹配模型——DMM

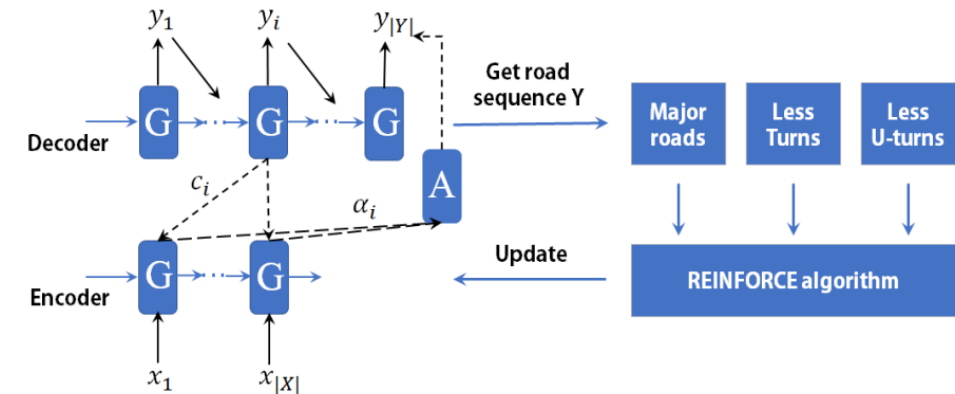
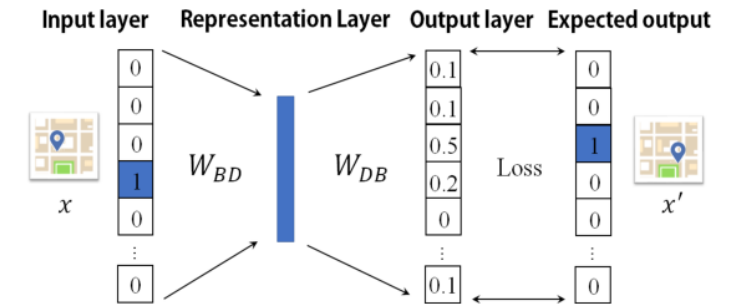
DMM是针对手机定位数据的路网匹配方法，轨迹中用户位置用所接入的基站位置代替

- 位置表征：使用Auto-Encoder学习表征，最大化邻近基站预测概率

最大化邻近基站预测概率 $\text{maximize} \sum_{x' \in C_x} \log P(x'|x)$

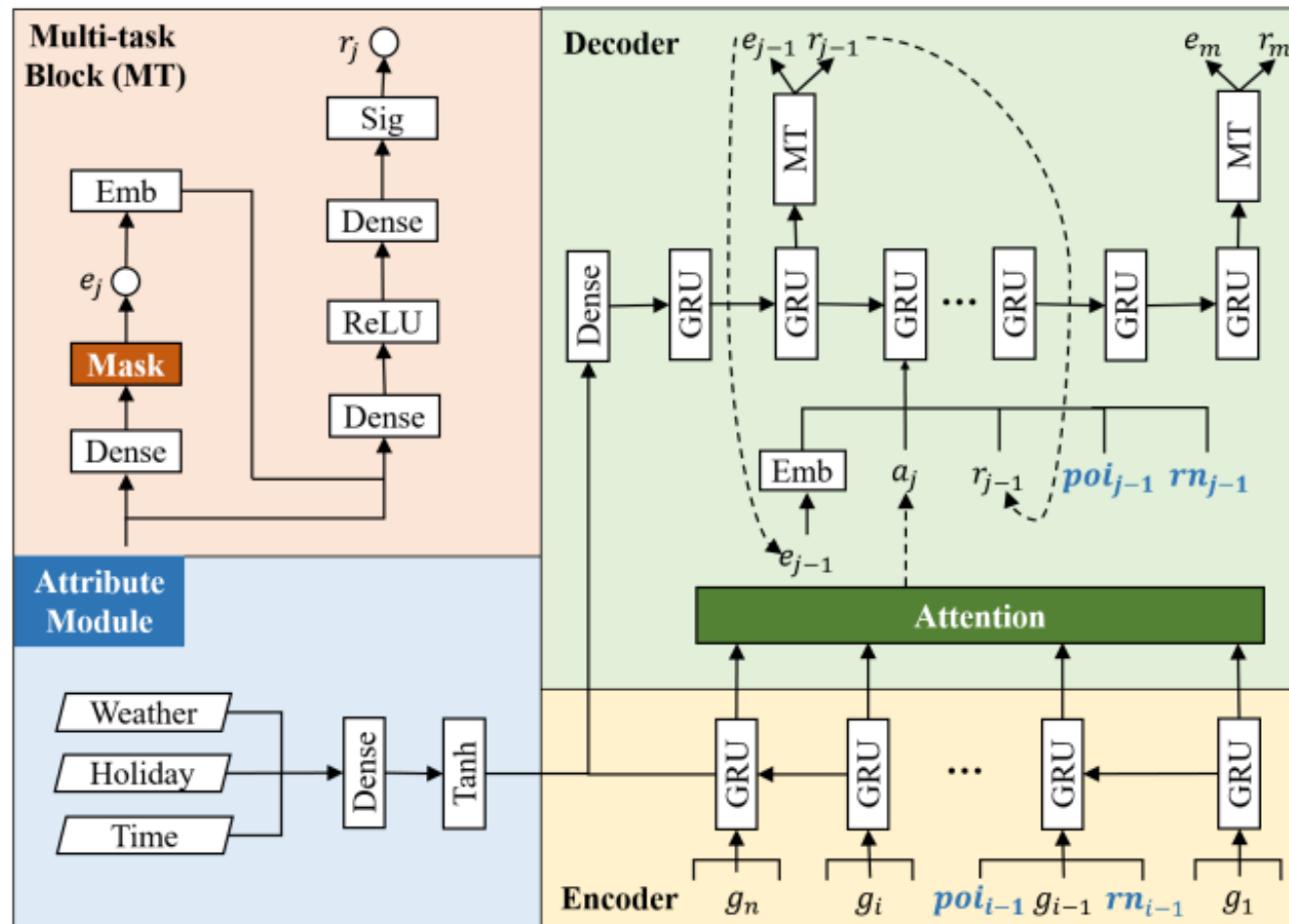
- 深度神经网络路网匹配：基于Encoder-Decoder的RNN架构和Alignment模块，将基站序列转换为路段序列

- 强化学习优化：奖励函数 $r(Y) = \lambda_P \cdot r_P + \lambda_T \cdot r_T + \lambda_U \cdot r_U$
考虑三方面因素：匹配路段和原始基站的加权空间距离、拐弯数量、U形路径数量



端到端的轨迹恢复——MTrajRec

- Encoder、Decoder结合Attention架构
- 引入天气、节假日、时间这些外部因素的影响
- 多任务预测模块：输出路段以及匹配到的点在路段上的比例



Ren H, Ruan S, Li Y, et al. **Mtrajrec: Map-constrained trajectory recovery via seq2seq multi-task learning**[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. 2021: 1410-1419.

端到端的轨迹恢复——MTrajRec

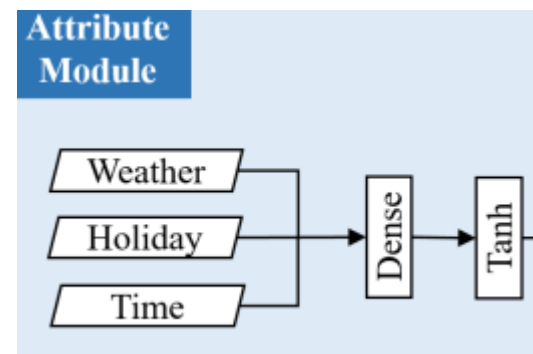
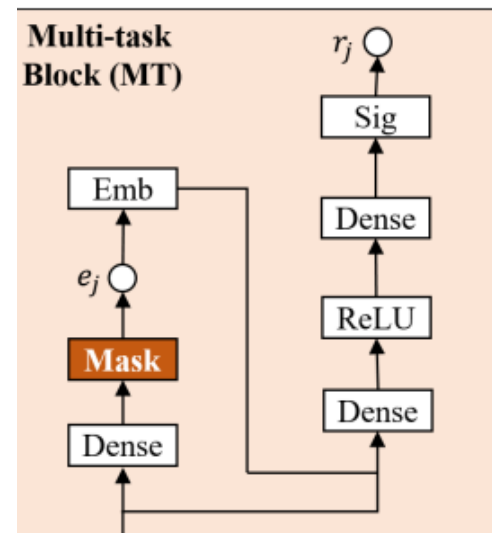
Mask层：将 softmax 预测概率限制在邻近的路段上

$$c_{j,\ell} = \begin{cases} f(d_{j,\ell}), & \text{if } j \in \{g_i.tid | i \in [1, n]\} \text{ and } d_{j,\ell} < 50 \\ 0, & \text{if } j \in \{g_i.tid | i \in [1, n]\} \text{ and } d_{i,\ell} \geq 50, \\ 1, & \text{otherwise} \end{cases}$$

$$P(\hat{e}_j | \mathbf{h}_j) = \frac{\exp(\mathbf{h}_j^\top \cdot \mathbf{w}_c) \odot \mathbf{c}_j}{\sum_{c' \in C} \exp(\mathbf{h}_j^\top \cdot \mathbf{w}_{c'}) \odot \mathbf{c}_j}$$

Attribute 模块，引入外部影响因素的特征信息

- 环境特征：天气等，影响速度
- 空间特征：道路等级、交叉口数量等



轨迹质量增强技术总结



分类	算法/模型	特性	共性-优点	共性-不足
基于轨迹上下文的轨迹质量增强技术	DHTR	巧妙将卡尔曼滤波引入神经网络训练，能 同时实现降噪和补全 。	不依赖于轨迹点以外的其他地理空间数据，即使在郊区/野外场景中也能使用；适用的轨迹类型多样化，既能处理 显式轨迹 (带GPS坐标)，也能处理 隐式轨迹 (仅有地点的ID表示)。	需要大量的历史数据进行训练；并且训练的数据对空间中各个位置的 覆盖程度要高 ，如果有些位置没有训练数据覆盖，则很可能无法恢复这些位置的轨迹。
	TrajBERT	框架 通用性较高 ，可提取多种类型的轨迹特征，适用于多种下游任务。		
基于路网的轨迹质量增强技术	HMM	算法比较成熟， 稳定性好 ；不依赖于大数据训练。	很适合处理城市中的轨迹质量增强，能充分利用路网结构，最终获得匹配到路网上的轨迹， 精确度较高 ；基于神经网络的方法能够充分挖掘路网空间中的特征，方法的上限比较高。	除了HMM以外，其他方法都需要大量的数据进行训练；相比于HMM，深度学习方法获取的轨迹点和道路ID映射关系可解释性较低；除HMM以外，在 城市之间迁移需要重新训练 。
	DeepMM	充分利用神经网络的优势；提供了训练数据增强的思想。		
	DMM	适用于基站轨迹序列的匹配；采用了较多的优化手段。		
	MTrajRec	以 端到端 的方式同时完成轨迹插值和路网匹配，可以认为是以上这些方法中最先进的。		



感谢各位同学
敬请批评指正