



Practical Synthetic Human Trajectories Generation Based on Variational Point Processes

林宇菁

Trajectory Generation



- 轨迹生成任务

- 轨迹: $S = \{s_1, \dots, s_n\}, s_i = (l_i, t_i), t_{i+1} = t_i + k_i + \diamond_i$

- 移动轨迹生成: 真实世界的移动轨迹数据集, 生成新的轨迹, 且保留原始轨迹数据的**关键特征和重要效用**。

- 关键问题:

- 不同的出行模式 (理解为交通方式) 下轨迹分布特征不同

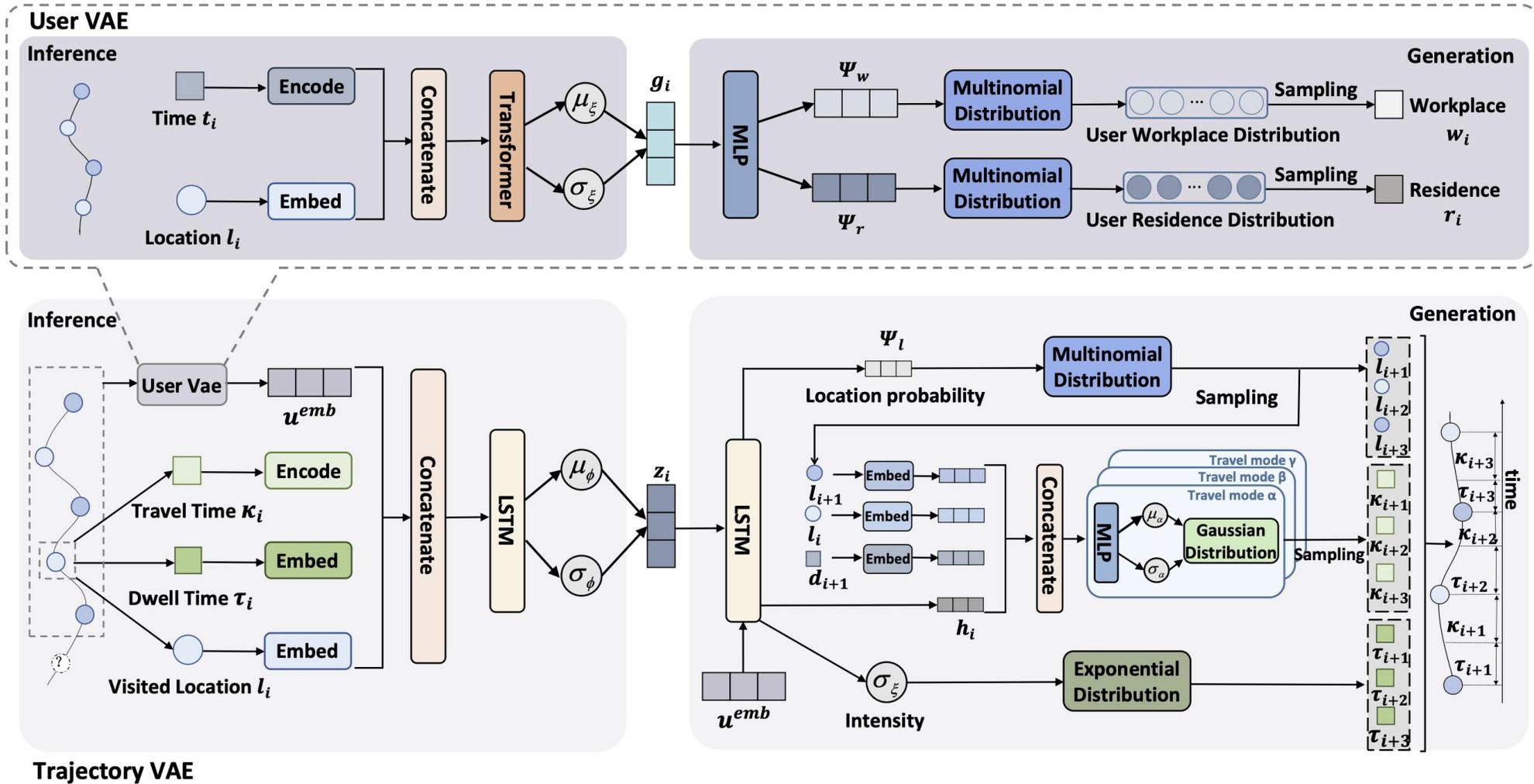
- 轨迹与居民密切相关, 居民本身分布特征不同 (居住地、工作地)

- VAE: 一种用于学习数据分布并生成相同分布数据的模型
- 基本思想:
 - 不同与一般编码器将一个样本对应到一个确定的表征向量, VAE 假设每个样本 x 对应的特征向量是从分布 $p_{z|x}=N(\mu_x, \sigma_x)$ 当中采样出来的 (隐变量模型)
 - 隐变量 z 服从标准正态分布 $p_z=N(0,1)$
- 求解思路: 使用神经网络估计样本的对应分布, 采样并重构该样本, 并对 z 施加正则化约束。
- 深层原理: VAE 实际上是一种广义 EM 估计算法

Model



Framework



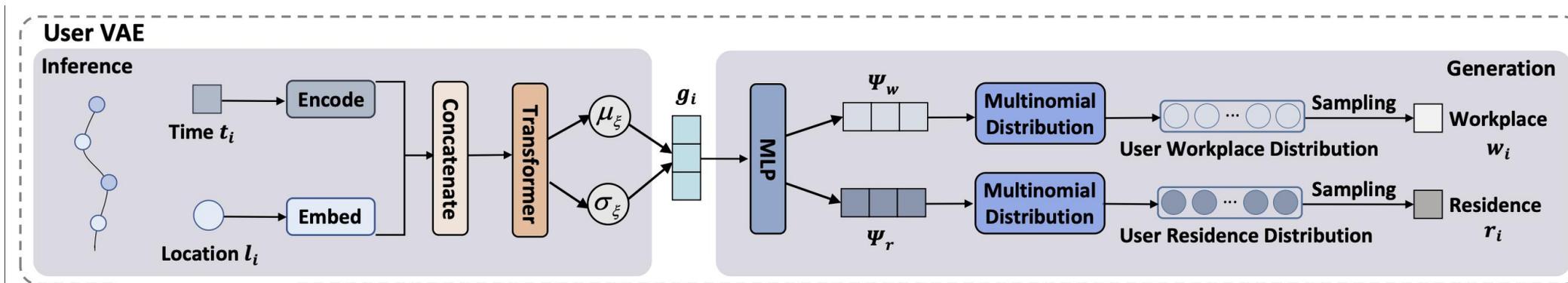
基本框架 = 下一跳模型 (User VAE + Trajectory VAE)

- User VAE: 建模不同的居民分布

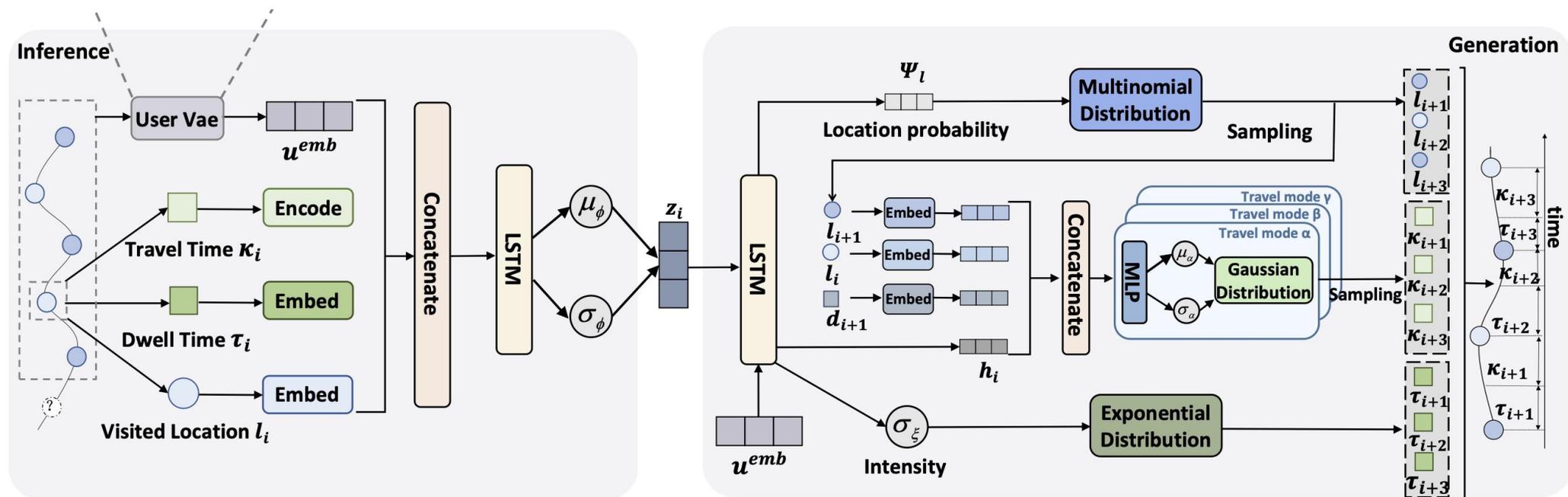
- 居民分布: 居住地分布与工作地分布

- 假设: 居民居住地和工作地服从多元正态分布

- 采样: 使用神经网络基于表征去预测多元正态分布的参数, 然后从这个正态分布采样。



- Traj VAE: 结合居民特征生成轨迹
 - 位置生成方式: 沿用下一跳框架 (基于历史状态推测下一跳位置)
 - 下一跳框架所需的表征: 使用VAE去提取

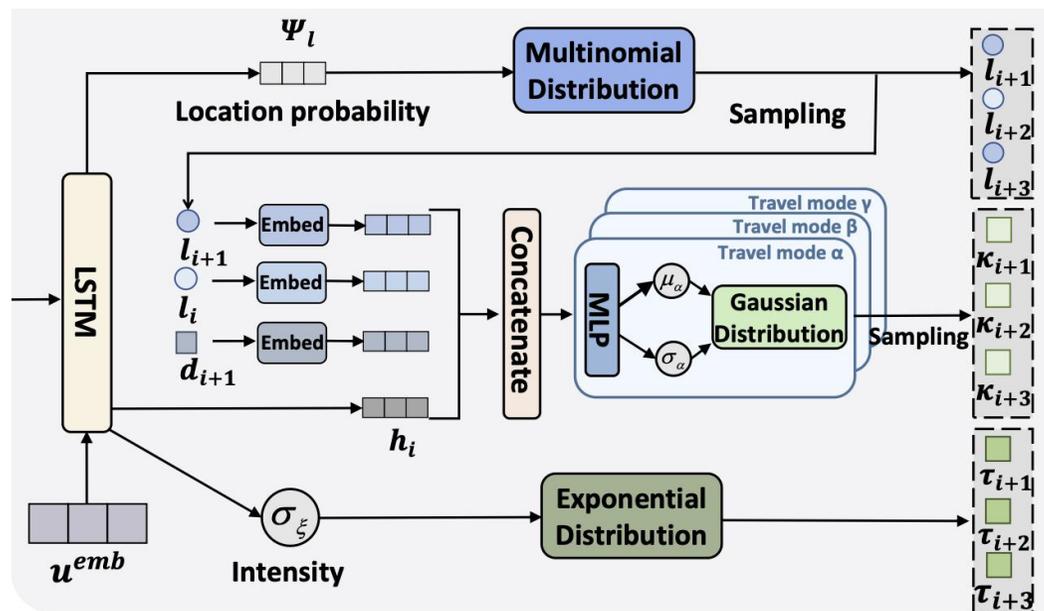


Trajectory VAE

- Traj VAE: 结合居民特征生成轨迹

- 下一跳到达时间: 结合得到的下一跳位置、当前位置预测从属不同出行模式的概率, 然后采样一个旅行时间

- 依据 Poisson 分布采样一个旅行时间和驻留时间



- 损失计算：原论文只提供了 VAE 部分的损失，即重构损失和正则化损失

$$\begin{aligned}\mathcal{L}_u &= \sum_u \mathcal{L}_{\lambda, \xi}(s_{1:N_u}^u) \\ &= \sum_u \sum_{i=1}^{N_u} E_{q_\xi(z_i^u | s_{1:i}^u)} [\log p_\lambda(s_i^u | z_{1:i}^u)] \\ &\quad - \text{KL}(q_\xi(z_i^u | s_{1:i}^u) || p_\lambda(z)).\end{aligned}$$

$$\begin{aligned}\mathcal{L}_t &= \sum_u \mathcal{L}_{\theta, \phi}(s_{1:N_u}^u) \\ &= \sum_u \sum_{i=1}^{N_u} E_{q_\phi(z_i^u | s_{1:i}^u)} [\log p_\theta(s_i^u | z_{1:i}^u)] \\ &\quad - \text{KL}(q_\phi(z_i^u | s_{1:i}^u) || p_\theta(z)).\end{aligned}$$

Experiment



Trajectory Generation

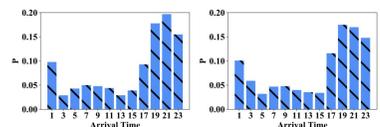


- 评估方式:

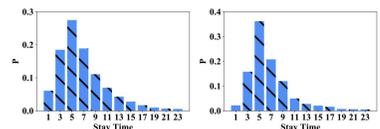
- Distance, G-rank, Duration, Move, and Stay 的分布 JSD 距离

	ISP					MME				
	Distance (JSD)	G-Rank (JSD)	Duration (JSD)	Move (JSD)	Stay (MSE×10 ²)	Distance (JSD)	G-Rank (JSD)	Duration (JSD)	Move (JSD)	Stay (MSE×10 ²)
Semi-Markov	0.016	0.197	0.026	0.062	<u>0.028</u>	0.012	0.213	0.018	0.053	<u>0.030</u>
TimeGEO	<u>0.013</u>	0.685	0.023	<u>0.055</u>	0.030	0.017	0.691	0.035	<u>0.049</u>	0.038
Hawkes	0.159	0.241	0.037	0.157	0.041	0.136	0.189	0.032	0.051	0.040
LSTM	0.125	0.269	<u>0.018</u>	0.059	0.031	<u>0.011</u>	0.245	<u>0.013</u>	0.048	0.032
MoveSim	0.028	0.238	0.312	0.121	0.056	0.042	0.314	0.298	0.107	0.031
VOLUNTEER	0.010	<u>0.221</u>	0.012	0.048	0.024	0.008	<u>0.217</u>	0.009	0.041	0.023
Improv.	23.1%	-	33.3%	12.7%	14.3%	27.2%	-	30.7%	16.3%	23.3%

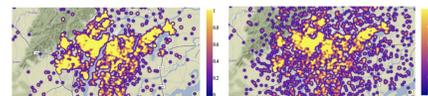
- 可视化比较



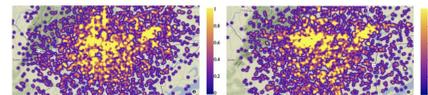
(a) Arrival time of real trajectories (b) Arrival time of generated trajectories



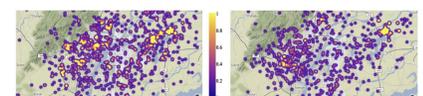
(c) Stay time of real trajectories (d) Stay time of generated trajectories



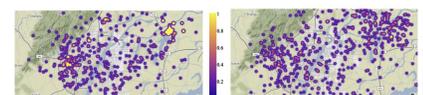
(a) Heatmap of real residence (b) Heatmap of generated residence based on our model



(c) Heatmap of generated residence based on Semi-Markov (d) Heatmap of generated residence based on MoveSim



(a) Heatmap of real workplace (b) Heatmap of generated workplace based on our model



(c) Heatmap of generated workplace based on Semi-Markov (d) Heatmap of generated workplace based on MoveSim

Thanks

