



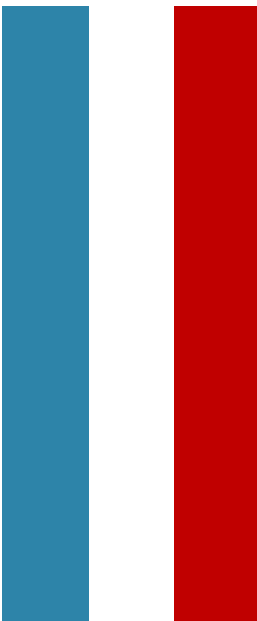
ST2Vec: Spatio-Temporal Trajectory Similarity Learning in Road Networks

Ziquan Fang, Yuntao Du, et al

Machine Learning 2021

汇报人：庞媛媛

2023/1/7





Ziquan Fang
浙江大学

研究领域

- Spatio-temporal Data Mining
- Distributed Processing

文章

标题	引用次数	年份
Real-time distributed co-movement pattern detection on streaming trajectories L Chen, Y Gao, Z Fang, X Miao, CS Jensen, C Guo Proceedings of the VLDB Endowment 12 (10), 1208-1220	49	2019
Metakg: Meta-learning on knowledge graph for cold-start recommendation Y Du, X Zhu, L Chen, Z Fang, Y Gao IEEE Transactions on Knowledge and Data Engineering	32	2022
MDTP: A multi-source deep traffic prediction framework over spatio-temporal trajectory data Z Fang, L Pan, L Chen, Y Du, Y Gao Proceedings of the VLDB Endowment 14 (8), 1289-1297	26	2021
Self-guided learning to denoise for robust recommendation Y Gao, Y Du, Y Hu, L Chen, X Zhu, Z Fang, B Zheng Proceedings of the 45th International ACM SIGIR Conference on Research and ...	24	2022
Spatio-Temporal Trajectory Similarity Measures: A Comprehensive Survey and Quantitative Study D Hu, L Chen, H Fang, Z Fang, T Li, Y Gao arXiv preprint arXiv:2303.05012	2	2023
A Lightweight Framework for Fast Trajectory Simplification Z Fang, C He, L Chen, D Hu, Q Sun, L Li, Y Gao 2023 IEEE 39th International Conference on Data Engineering (ICDE), 2386-2399		2023

文章



Yuntao Du
BIGAI

研究领域

- Transfer learning
- Domain adaptation

标题

引用次数

年份

[Adarnn: Adaptive learning and forecasting of time series](#)

124

2021

Y Du, J Wang, W Feng, S Pan, T Qin, R Xu, C Wang

Proceedings of the 30th ACM international conference on information ...

[Generation, augmentation, and alignment: A pseudo-source domain based method for source-free domain adaptation](#)

20

2023

Y Du, H Yang, M Chen, H Luo, J Jiang, Y Xin, C Wang

Machine Learning, 1-21

[Semi-supervised learning with multi-head co-training](#)

18

2022

M Chen, Y Du, Y Zhang, S Qian, C Wang

Proceedings of the AAAI Conference on Artificial Intelligence 36 (6), 6278-6286

[Adversarial separation network for cross-network node classification](#)

16

2021

X Zhang, Y Du, R Xie, C Wang

Proceedings of the 30th ACM International Conference on Information ...

[Spatial-temporal graph convolutional gated recurrent network for traffic forecasting](#)

3

2022

L Zhao, M Chen, Y Du, H Yang, C Wang

arXiv preprint arXiv:2210.02737

主要内容

Main Contents

1

背景

2

内容

3

方法

4

实验

5

总结

1 背景

■ 轨迹相似性计算

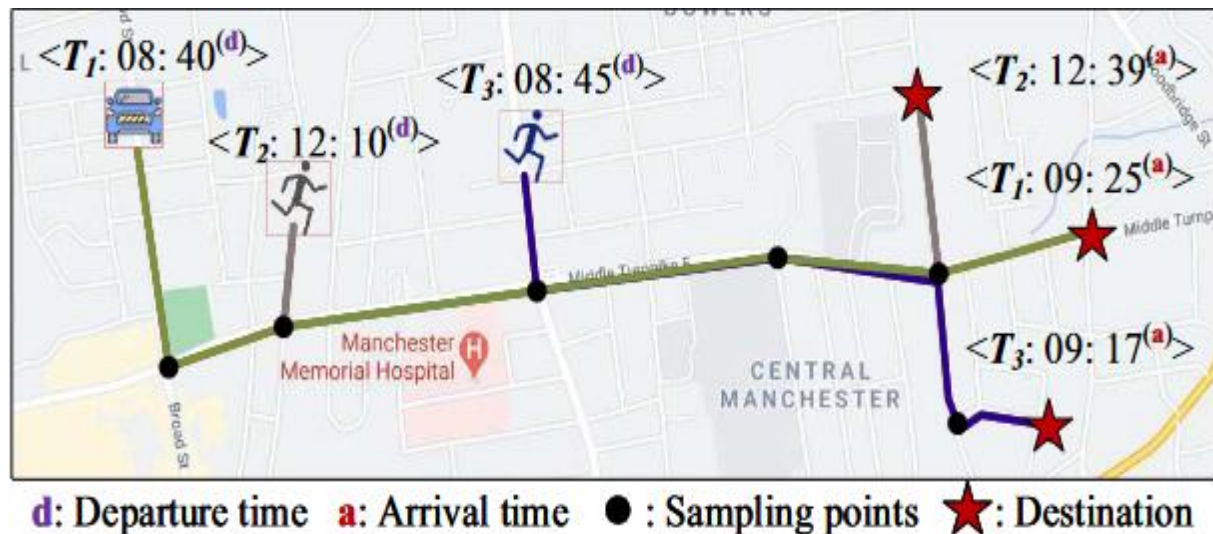
■ 现有方法及缺点

手工距离测量

计算成本高且复杂度高

神经网络方法

忽略时间维度



时空轨迹相似性的一个例子

■ 为什么考虑时空相似性？

情境：T₁表示一名司机假话的行程，T₂和T₃是两个寻找打车机会的人。空间角度推荐T₂，时间角度推荐T₃。

■ 应用

为司机分配潜在的共享乘车伙伴；

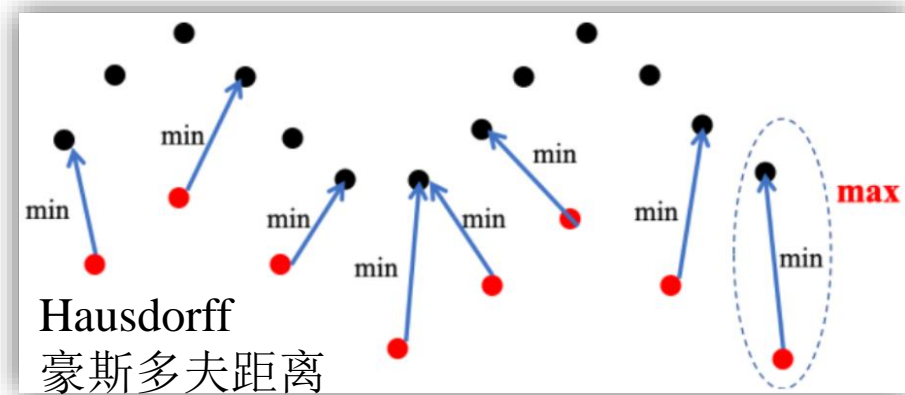
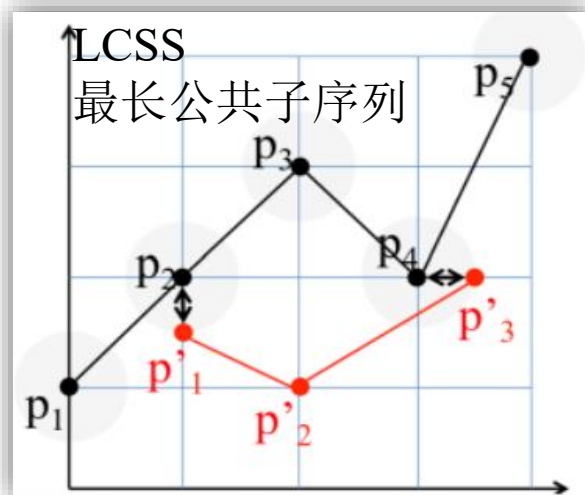
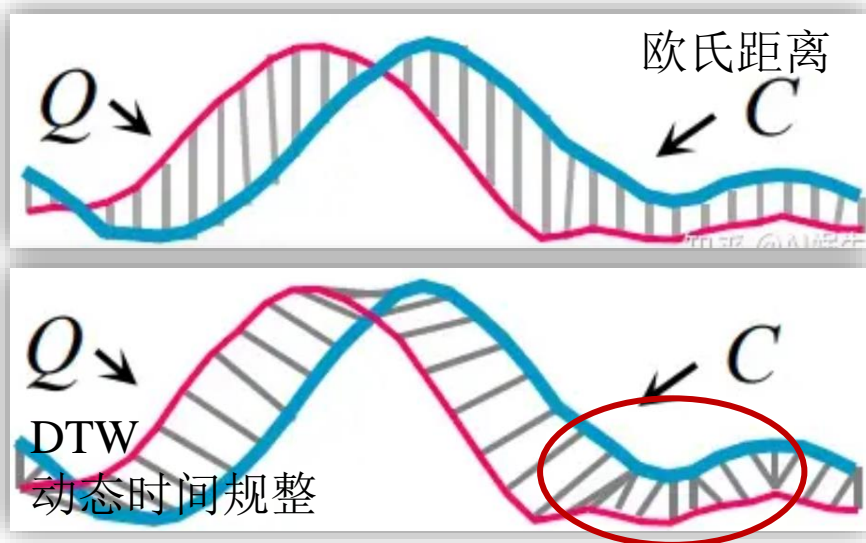
交通管理部门通过汇总相似轨迹和计算道路的出行频率来预测交通拥堵情况；

社交应用程序识别具有相似生活轨迹的用户来进行好友推荐。

补充学习：相似度计算

■ 手工距离测量

① 基于自由空间的测量方法



② 基于路网的测量方法：

TP、DITA、LCRS、NetERP

原始轨迹映射到路网路径上再基于经典的距离度量定义相似度

■ 神经网络方法（轨迹嵌入相似性反映GPS轨迹相似性）

2 内容

■ 如何捕捉轨迹之间的时间相关性

时间有很强的连续性和周期性

时间建模模块TMM

■ 如何融合时空轨迹嵌入，实现统一的时空相似性学习

不同用户赋予空间和时间相似性不同权重以适应不同应用

时空协同融合模块STCF

■ 如何优化模型以提高有效性和效率

目标的有效性—相似性查询的质量
效率—模型收敛速度

新的三元组采样策略
课程学习
两种融合方法

课程学习 (Curriculum Learning, CL) 是一种训练策略，模仿人类的学习过程，主张让模型从容易的样本开始学习，并逐渐进阶到复杂的样本和知识。

2 内容

- 道路网络: $G = (L, E)$, 顶点 $l_i = (x_i, y_i)$ 表示道路交叉口或道路重点
- GPS轨迹: $T = \langle (g_1, t_1), (g_2, t_2), \dots, (g_n, t_n) \rangle$, g 表示由经纬度组成的观测地理位置
- 路网中的轨迹: 地图匹配程序将轨迹与OpenStreetMap中路网匹配映射。
空间轨迹 $T^{(s)} = \langle l_1, l_2, \dots, l_m \rangle$ 和时间轨迹 $T^{(t)} = \langle t_1, t_2, \dots, t_m \rangle$

- 相似性函数: 空间和时间相似性的加权线性组合

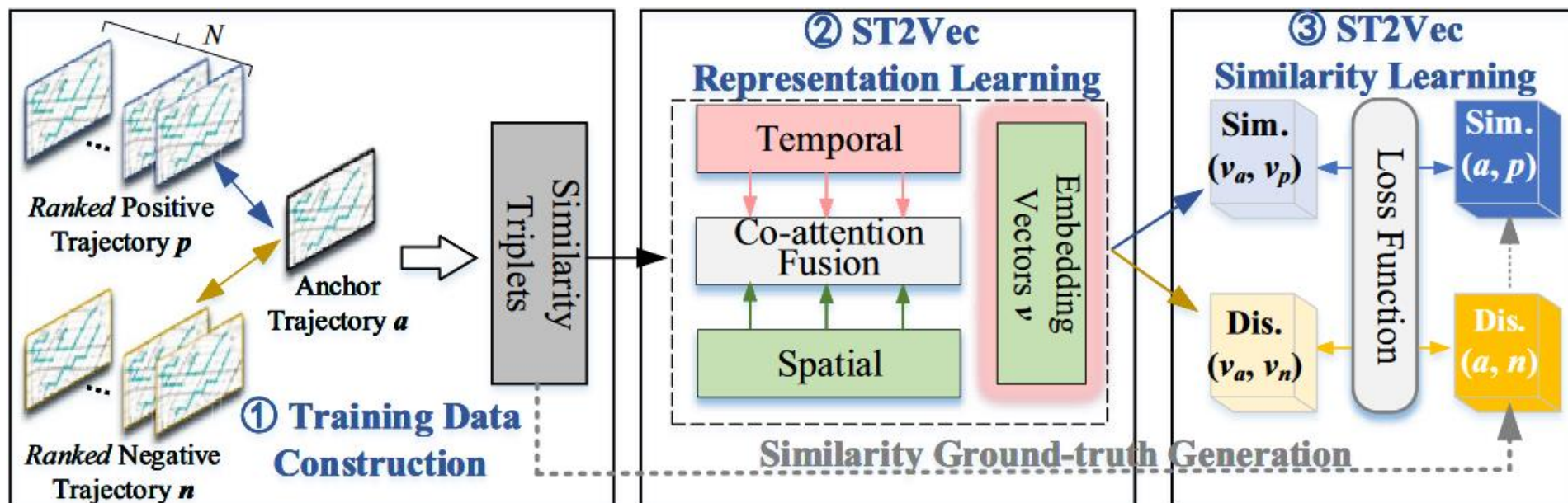
$$\mathcal{D}(T_i, T_j) = \lambda \cdot \mathcal{D}_S(T_i^{(s)}, T_j^{(s)}) + (1 - \lambda) \cdot \mathcal{D}_T(T_i^{(t)}, T_j^{(t)}) \quad \lambda \in [0, 1]$$

- 学习目标: 神经网络驱动函数 $\mathcal{G}(\cdot, \cdot)$ 最大程度地接近 $\mathcal{D}(T_i, T_j)$

$$\arg \min_{\mathcal{M}} |\mathcal{G}(v_{T_i}, v_{T_j}) - \mathcal{D}(T_i, T_j)|$$

\mathcal{M} 是神经网络的模型参数, \mathcal{D} 是时空轨迹相似性函数, v_{T_i}, v_{T_j} 是 T_i, T_j 的时空嵌入。

2 内容



- ❑ 利用相似和不相似轨迹对构建相似性三元组;
- ❑ 同时考虑空间和时间纬度学习嵌入轨迹; 嵌入的本质是压缩
- ❑ 学习轨迹嵌入直至嵌入向量的轨迹相似性接近地面实况相似性。

3 方法: ST2Vec方法

■ 时间建模模块 (Temporal Modeling Module, TMM)

□ 时间嵌入

$$T^{(t)} = \langle t_1, t_2, \dots, t_m \rangle$$

$$t'[i] = \begin{cases} \omega_i t + \varphi_i, & \text{if } i = 0 \\ \cos(\omega_i t + \varphi_i), & \text{if } 1 \leq i \leq q \end{cases}$$

$$\langle t_1, t_2, \dots, t_m \rangle \rightarrow \langle t'_1, t'_2, \dots, t'_m \rangle$$

□ 时序嵌入

$$h_i = \text{LSTM}(t'_i, h_{i-1}, i_i, f_i, o_i, m_i)$$

□ 去耦合注意力

轨迹中不同时间点的重要性不同

$$\tilde{h}_i^{(p)} = \sum_{k=1}^i \text{att}(h_i^{(p)}, h_k^{(p)}) \cdot h_k^{(p)}$$

计算同一轨迹中时间点之间的注意力得分，捕捉轨迹点之间相关性。

$$\text{att}(h_i^{(p)}, h_k^{(p)}) = \frac{\alpha_{i,k}}{\sum_{k'=1}^i \exp(\alpha_{i,k'})}$$

$$\alpha_{i,k} = w_1^\top \cdot \tanh(W_1 \cdot h_k^{(p)} + W_2 \cdot h_i^{(p)})$$

普通的序列模型RNN、LSTM无法处理时间的周期性和非周期性时间模式。
∴时间嵌入+时序嵌入+注意力结合

cos周期激活函数，捕捉周期性行为
线性项表示时间的推移，捕捉非周期性行为

i_i, f_i, o_i , and s_i 表示输入门、遗忘门、输出门和存储单元。建立时间依赖性模型。最后一层的隐藏状态视为深度时间轨迹表示。

3 方法：ST2Vec方法

将空间轨迹嵌入低维空间向量，
捕捉受道路网络约束的空间信息。

■ 空间建模模块（Spatial Modeling Module, SMM）

□ 位置嵌入

$$T^{(s)} = \langle l_1, l_2, \dots, l_m \rangle$$

Node2Vec捕捉道路网络中相邻位置的共生关系，得到原始位置→嵌入位置的映射，即 $l_i \rightarrow n_i$

嵌入位置通过GNN获得局部光滑的位置嵌入 $l'_i = \text{GCN}(n_i) = \sigma \left(\left(\sum_{j \in N_i} c_{ij} W_s n_j \right) \parallel n_i \right)$

得到空间轨迹的细粒度表示 $\langle l_1, l_2, \dots, l_m \rangle \rightarrow \langle l'_1, l'_2, \dots, l'_m \rangle$

□ 位置序列嵌入和空间注意力：同时间建模模块

补充学习: Node2Vec

综合考虑DFS邻域和BFS邻域的图嵌入方法

■ **Skip-Gram模型**: 给出目标单词（中心词）的情况下预测上下文单词

Step1 滑窗法对句子采样，中心词和相邻词组的单词对

Step2 中心词转换为one-hot编码作为神经网络的输入，输出为相邻词

Step3 训练后，得到预测的相邻词及出现概率

■ 有偏的随机游走模型

随机游走获取的节点序列无法反映**同质性**和**同构性**（由DFS和BFS体现），有偏的随机游走采用**搜索参数**控制采样过程满足上述特性。

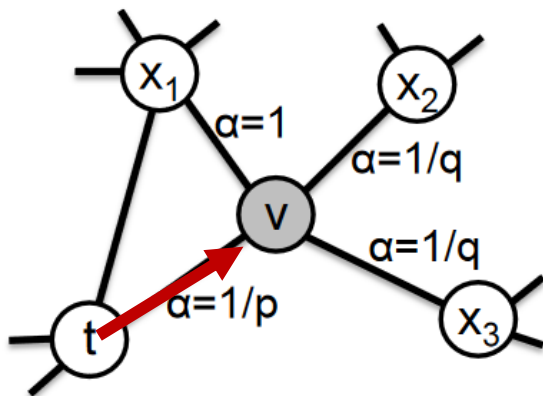
p: 控制重复访问刚刚访问过的顶点的概率

q: 控制游走向外还是向内

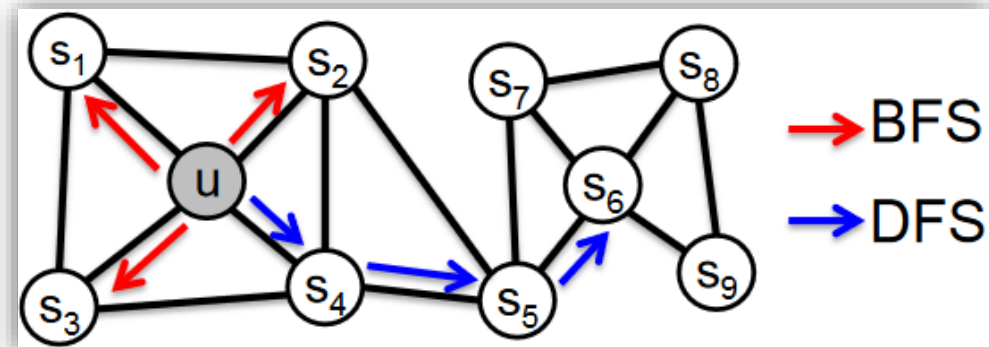
p较小，采样在t附近，表现BFS特性

q较小，采样远离t，表现DFS特性

$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ 1 & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases}$$



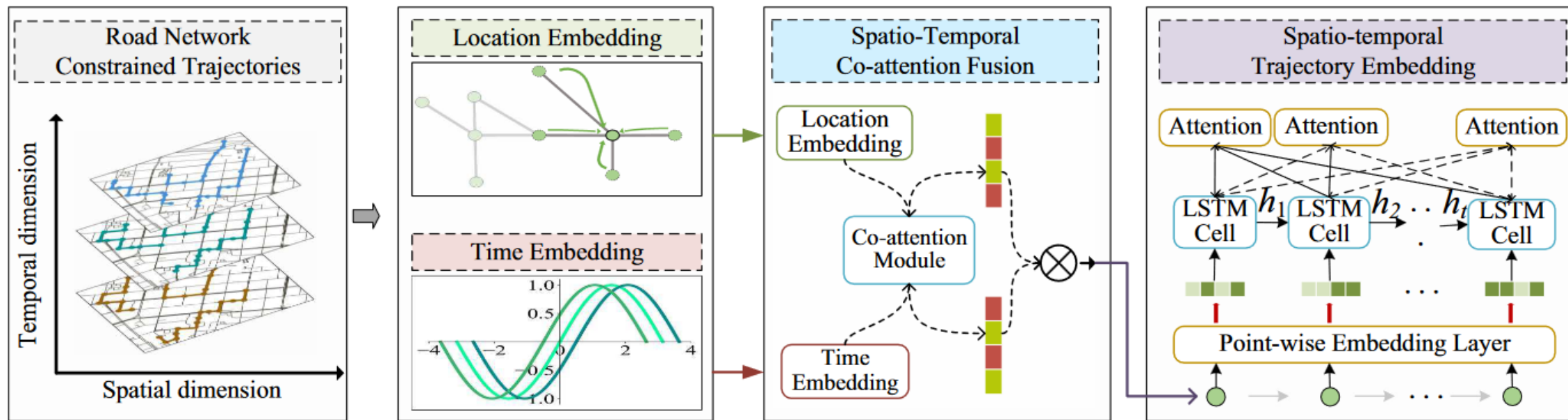
Source Text	Training Samples
The quick brown fox jumps over the lazy dog. →	(the, quick) (the, brown)
The quick brown fox jumps over the lazy dog. →	(quick, the) (quick, brown) (quick, fox)
The quick brown fox jumps over the lazy dog. →	(brown, the) (brown, quick) (brown, fox) (brown, jumps)
The quick brown fox jumps over the lazy dog. →	(fox, quick) (fox, brown) (fox, jumps) (fox, over)



■ **Node2Vec**: 有偏的随机游走采样序列+skip-Gram获得相邻位置嵌入的信息

3 方法: ST2Vec方法

■ 时空协同融合模块 (Spatio-Temporal Co-attention Fusion, STCF)



两种融合策略

□ 分离融合

$$v_T = LSTM_t(t'_1, t'_2, \dots, t'_m) + LSTM_s(l'_1, l'_2, \dots, l'_m)$$

参数量大

□ 统一融合

输入初始时间序列嵌入 $\tau^{(t)} = \langle t'_1, t'_2, \dots, t'_m \rangle$ 和空间嵌入序列 $\tau^{(s)} = \langle l'_1, l'_2, \dots, l'_m \rangle$ 到协同融合模块

3 方法: ST2Vec方法

■ 时空协同融合模块 (Spatio-Temporal Co-attention Fusion, STCF)

□ 统一融合

初始时间序列嵌入 $\tau^{(t)} = \langle t'_1, t'_2, \dots, t'_m \rangle$

空间嵌入序列 $\tau^{(s)} = \langle l'_1, l'_2, \dots, l'_m \rangle$

Step1 对时间和空间特征转换 $z_\tau^1 = W_F \tau^{(t)}, \quad z_\tau^2 = W_F \tau^{(s)}$

$$\beta_{i,j} = \frac{\exp(W'_Q z_\tau^i \cdot W'_K z_\tau^{jT})}{\sum_{j' \in \{1,2\}} \exp(W'_Q z_\tau^i \cdot W'_K z_\tau^{j'T})},$$

Step2 计算两个特征之间的相互作用并得到增强表示

$$\begin{aligned} \tau^{(\hat{t})} &= \text{Norm} \left(FFN' (\beta_{1,1} z_\tau^1 + \beta_{1,2} z_\tau^2) + \tau^{(t)} \right), \\ \tau^{(\hat{s})} &= \text{Norm} \left(FFN' (\beta_{2,1} z_\tau^1 + \beta_{2,2} z_\tau^2) + \tau^{(s)} \right), \end{aligned}$$

Step3 增强表示通过LSTM实现统一的时空轨迹嵌入

$$v_T = LSTM(\tau^{(\hat{t})}, \tau^{(\hat{s})})$$

3 方法：训练和模型优化

■ 数据选择

相似性三元组 (T_a, T_p, T_n) （锚点轨迹、正向轨迹、负向轨迹）

正负：相似程度高低

■ 训练过程

$$\mathcal{G}(\cdot, \cdot) \leftarrow f_{\theta}^{(t,s)} : (T_a^{(t,s)}, T_p^{(t,s)}, T_n^{(t,s)}) \rightarrow (v_a^{(t,s)}, v_p^{(t,s)}, v_n^{(t,s)})$$

时空相似度：基于 $\|v_a - v_p\|_2$ and $\|v_a - v_n\|_2$ 的 L_2 准则

归一化： $\mathcal{D}'_{a,p} = \exp(-\alpha \cdot \mathcal{D}_{a,p}) \in [0, 1]$

■ 损失函数

$$\mathcal{L} = \mathcal{D}'_{a,p} (\mathcal{D}'_{a,p} - \exp(-\|v_a - v_p\|_2))^2 + \mathcal{D}'_{a,n} (\mathcal{D}'_{a,n} - \exp(-\|v_a - v_n\|_2))^2$$

■ 课程学习法

由易到难训练样本：最不相似轨迹 → 最相似轨迹

4 实验

- 数据集：T-Drive、Rome、Xi'an
- 预处理：GPS轨迹映射匹配到路网、删除少量数据
- 验证方法：top-k相似性，使用传统的基于非倾斜的距离度量（TP、DITA、LCRS、NetERP）
- 评估指标：HR@10、HR@50、R10@50（越接近1模型越有效）

HR@k: top-k命中率，捕捉top-k结果与相应地面实况结果之间的重叠程度。

R k@t: top-k地面实况的top-t召回率，捕捉top-k地面实况在相应top-t结果中的比例。

- 基准：对NEUTRAJ、Traj2SimVec、T3S、GTS四种基线在时间控制方面扩展

窗口引导基线（*_w）：将轨迹分布在离散的时隙中，执行top-k相似性查询

LSTM引导基线（*_l）：将时间轨迹直接输入LSTM模型

TMM引导基线（*_t）：将时间建模模块集成到baseline

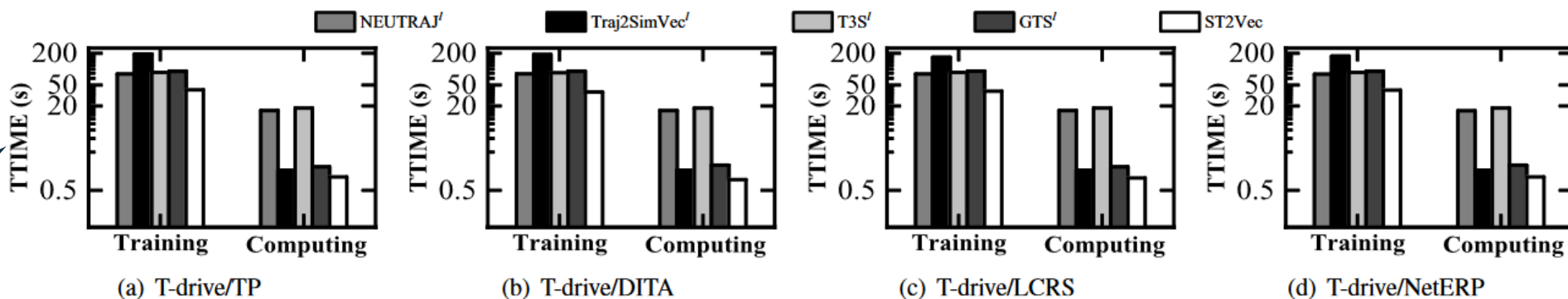
- 评估方向：有效性、高效性、可扩展性、敏感性、消融实验、案例研究等

4 实验

- 有效性研究
- 高效性研究（离线模型训练、在线计算）

THE COMPARISON OF SIMILARITY LEARNING ON TP, DITA, LCRS, AND NetERP DISTANCES USING T-DRIVE DATASET

Category	Methods	TP [22]			DITA [26]			LCRS [44]			NetERP [14]		
		HR@10	HR@50	R10@50	HR@10	HR@50	R10@50	HR@10	HR@50	R10@50	HR@10	HR@50	R10@50
Window Guided Baselines	NEUTRAJ ^w	0.0978	0.1373	0.1582	0.0805	0.1243	0.1442	0.0357	0.0419	0.0861	0.0054	0.0173	0.0198
	Traj2SimVec ^w	0.0827	0.1261	0.1397	0.053	0.0682	0.1151	0.016	0.098	0.1861	0.0209	0.0986	0.1010
	T3S ^w	0.1295	0.1733	0.2045	0.0838	0.1266	0.1489	0.0435	0.0678	0.1187	0.01253	0.0292	0.0388
	GTS ^w	0.3034	0.3980	0.6975	0.1178	0.2223	0.3991	0.0188	0.0538	0.0652	0.0252	0.0408	0.0505
LSTM Guided Baselines	NEUTRAJ ^l	0.1765	0.2221	0.2703	0.0767	0.1103	0.1340	0.0533	0.1126	0.1694	0.0259	0.0502	0.0736
	Traj2SimVec ^l	0.1446	0.1902	0.2263	0.05261	0.0642	0.1071	0.0329	0.1397	0.2257	0.0328	0.1050	0.1244
	T3S ^l	0.1535	0.1984	0.2382	0.0806	0.1191	0.1422	0.0486	0.0904	0.1445	0.0193	0.0398	0.0563
	GTS ^l	0.3709	0.4756	0.7965	0.1277	0.2321	0.4143	0.0360	0.1074	0.1342	0.03984	0.0655	0.0894
Our TMM Guided Baselines	NEUTRAJ ^l	0.3371	0.4091	0.7001	0.1412	0.2719	0.4892	0.0924	0.2848	0.3632	0.1086	0.1832	0.2841
	Traj2SimVec ^l	0.3987	0.5364	0.6593	0.1321	0.3072	0.3643	0.0968	0.2826	0.3741	0.2128	0.3212	0.5553
	T3S ^l	0.3944	0.5011	0.7917	0.1284	0.2288	0.4073	0.1442	0.4331	0.5672	0.1464	0.2767	0.4077
	GTS ^l	0.4243	0.5640	0.8026	0.3244	0.4370	0.6381	0.1643	0.4427	0.6242	0.2154	0.3477	0.5343
Our Methods	ST2Vec	0.4624	0.5868	0.8361	0.3773	0.5037	0.7031	0.1806	0.5469	0.7293	0.2386	0.3493	0.6133



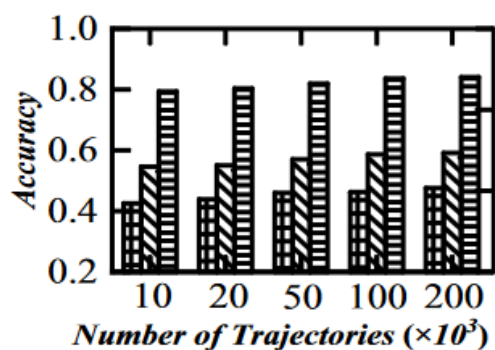
训练 s/epoch
计算 s/4千条轨迹

4 实验

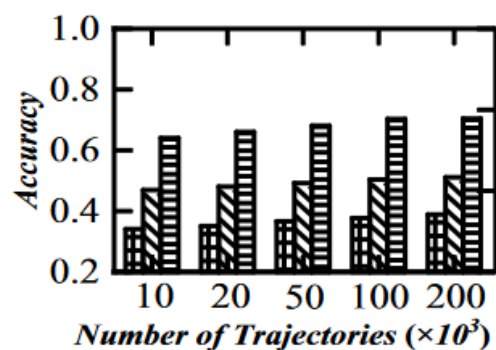
- 可扩展性研究（轨迹数量10k→200k）
- 参数敏感性研究（训练数据大小、为每个轨迹构建的三元组数量N、时空权重 λ ）

MODEL SCALABILITY EVALUATION WITH VARYING NUMBER OF TRAJECTORIES TO PERFORM TOP- k SIMILARITY COMPUTATION

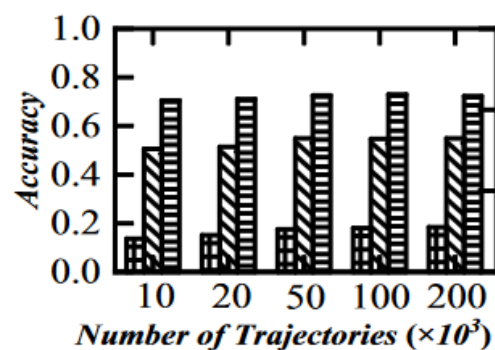
Datasets	Methods	TP Distance				DITA Distance				LCRS Distance				NetERP Distance			
		10k	50k	100k	200k	10k	50k	100k	200k	10k	50k	100k	200k	10k	50k	100k	200k
T-Drive	NEUTRAJ ^l	27.81	131.39	261.16	534.24	23.75	135.18	258.04	537.96	31.10	127.03	261.04	529.72	25.13	127.22	257.48	538.23
	Traj2SimVec ^l	93.66	458.90	927.87	1862.11	92.10	454.28	926.25	1865.86	98.55	461.15	924.49	1866.50	93.34	456.74	928.39	1858.86
	T3S ^l	30.52	146.94	275.91	542.53	33.80	142.58	276.96	541.31	34.99	147.42	279.50	542.94	29.77	148.35	279.40	540.53
	GTS ^l	34.65	159.52	299.19	602.52	37.67	156.25	297.24	597.64	35.63	158.60	300.50	607.09	37.68	159.61	296.45	606.36
	ST2Vec	30.32	145.94	293.35	597.71	25.81	146.65	293.03	596.56	29.38	143.89	297.92	593.77	28.17	147.26	290.46	598.88



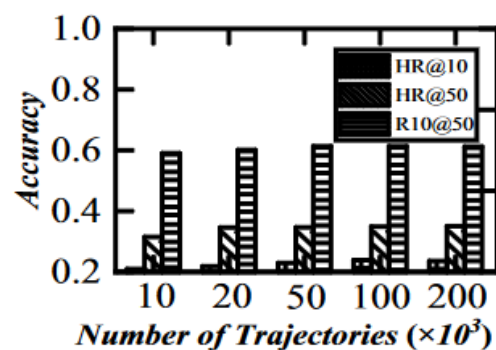
(a) T-drive/TP



(b) T-drive/DITA



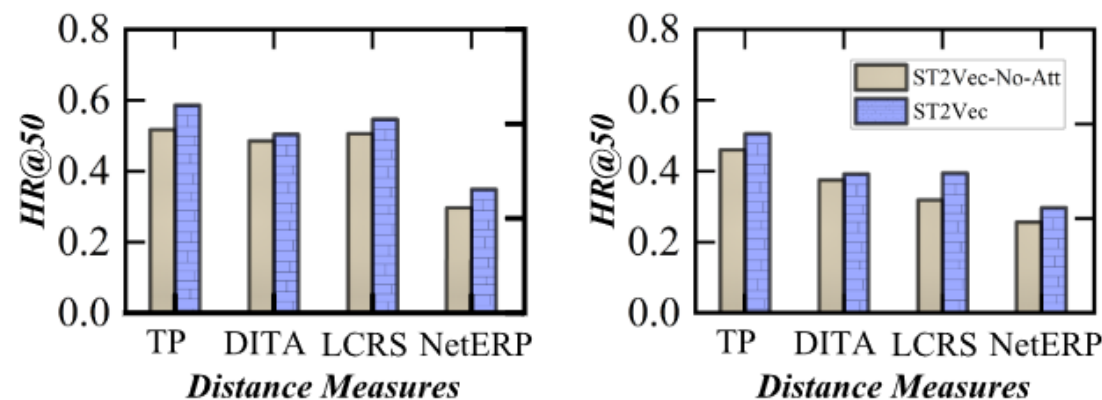
(c) T-drive/LCRS



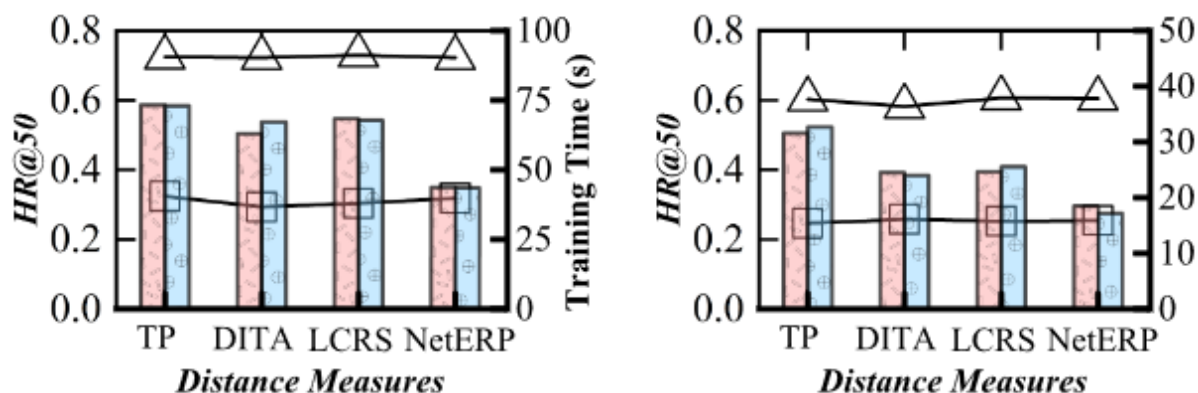
(d) T-drive/NetERP

4 实验—消融实验

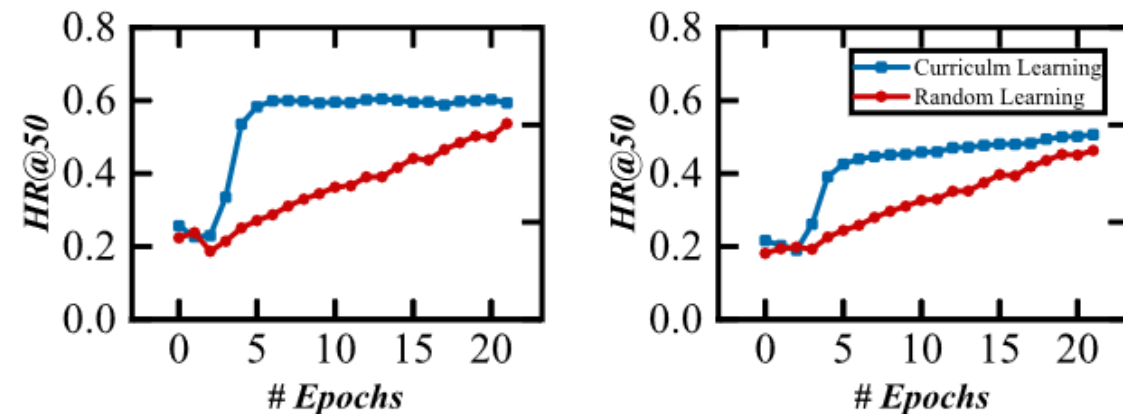
- 注意力的有无
- 时空融合策略的选择：分离融合和统一融合
- 优化方法：课程学习和随机学习



(a) T-Drive
(b) Rome
Fig. 8. ST2Vec Performance vs. with/without Attention



(a) T-Drive
(b) Rome
Fig. 9. ST2Vec Performance vs. Fusion Manners



(a) T-Drive
(b) Rome
Fig. 10. The Convergence Curve of ST2Vec with respect to 20 epochs

4 实验

- 效率加速研究
- 案例研究（轨迹top-k查询和聚类，直观检验ST2Vec能力）

TABLE V

TIME COST OF ONLINE SIMILARITY SEARCH ON T-DRIVE

Measures	Methods	1k	5k	10k	200k
TP	Non-learning	1.492s	3.127s	5.893s	117.832s
	ST2Vec	0.004s	0.014s	0.028s	0.521s
DITA	Non-learning	0.921s	3.301s	6.291s	125.826s
	ST2Vec	0.004s	0.015s	0.028s	0.522s
LCRS	Non-learning	1.292s	4.614s	8.784s	175.824s
	ST2Vec	0.004s	0.014s	0.028s	0.525s
NetERP	Non-learning	1.535s	6.246s	12.674s	253.481s
	ST2Vec	0.004s	0.015s	0.028s	0.522s

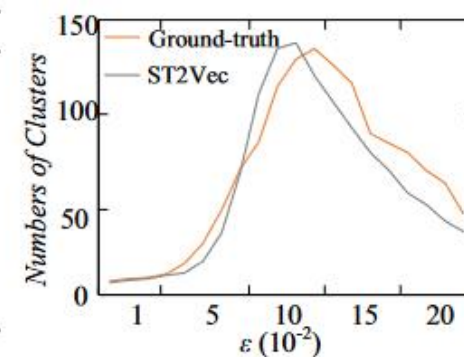


Fig. 11. Case Studies: Top-k Querying and Clustering

5 总结

□ 优点

时空融合：捕捉轨迹之间的复杂相似性；

表示学习：嵌入表示将轨迹映射到低维空间，高效性；

模型优化：课程学习策略，收敛速度和准确性提升；

实验验证：有效性、高效性、可扩展性、参数敏感性；

应用案例：通过top-k相似查询和聚类展示ST2Vec下游分析的潜力。

□ 缺点

模型复杂度：ST2Vec包含多个模块，训练难度增加；

参数敏感性：对其他超参数仍然敏感；

泛化能力：仅限道路网络。

□ idea

模型简化

参数优化：自动化的超参数优化技术；

泛化到其他领域：社交网络轨迹、移动应用数据等。

动态权重调整：使模型能够根据具体应用场景自动调整对空间和时间信息的重视程度。

谢谢！

