POI-based Traffic Generation via Supervised Contrastive Learning on Reconstructed Graph

Ziyang Su, Zhanyu Liu, Jianrong Ding and Guanjie Zheng

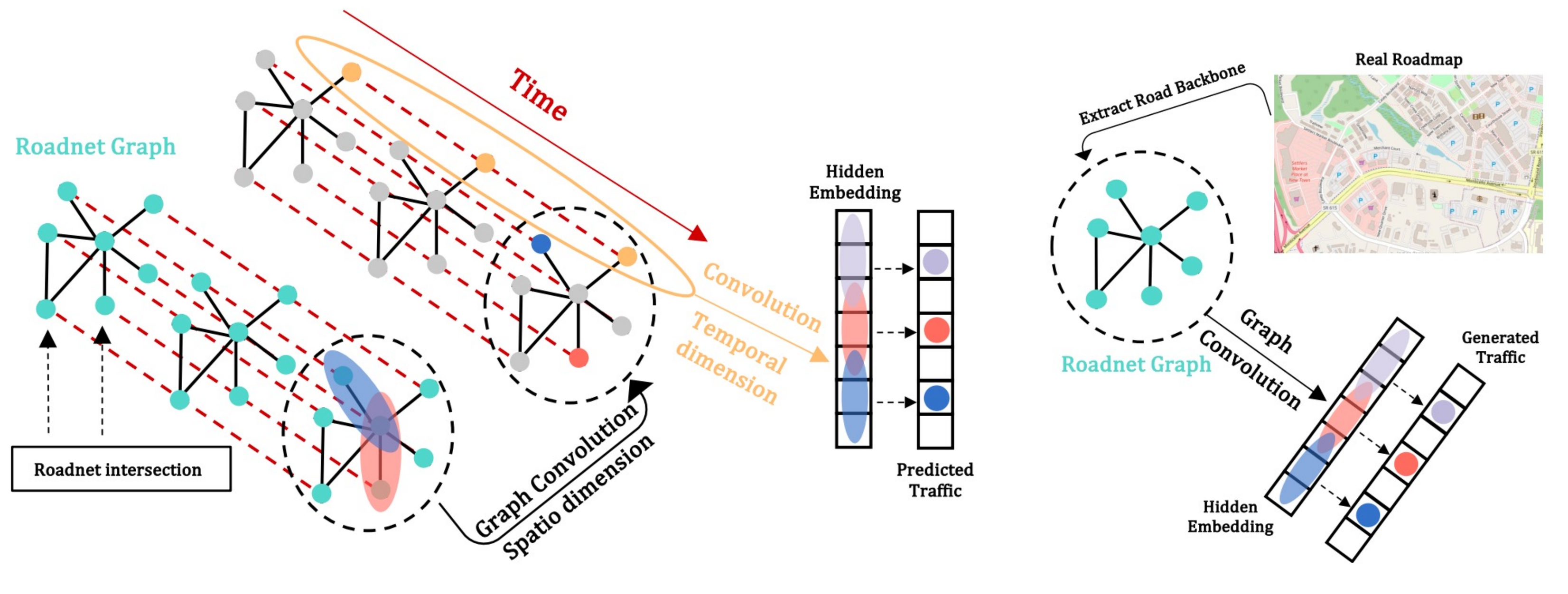
Shanghai Jiao Tong University, Shanghai, China
{suziyang,zhyliu00,rafaelding,gjzheng}@sjtu.edu.cn

DASFAA 2024 Gifu Japan

SHANGHAI JIAO TONG UNIVERSITY

Motivation of Traffic Generation Problem

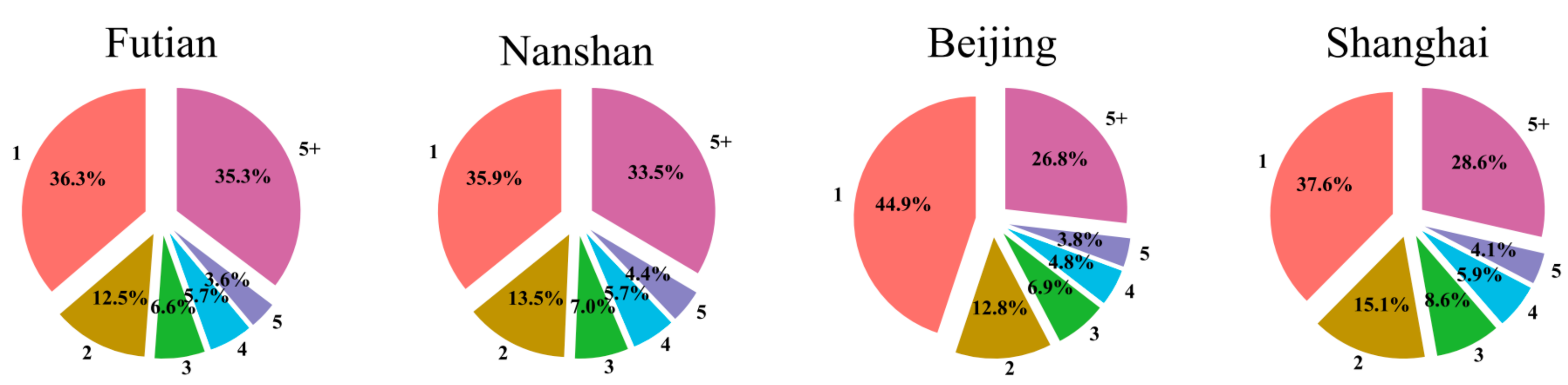
Conventional traffic prediction models are *contingent upon an expansive reservoir of historical traffic data* to enhance their predictive performance. However, *the acquisition of real-world traffic data*, particularly in smaller cities, is a big challenge.



Common frameworks of Traffic Prediction and Traffic Generation problems

Challenges

- Real-world road network graphs often *deviate from* the elemental principle that *"neighbor nodes exhibit similar features"*. Hop number of the shortest path between node pairs with nearest POI feature:



- Most previous models suffer from *the perceived inadequacy of end-to-end loss.*

Experiment Results

TG-SCR achieves the *best performance* on four real-world dataset.

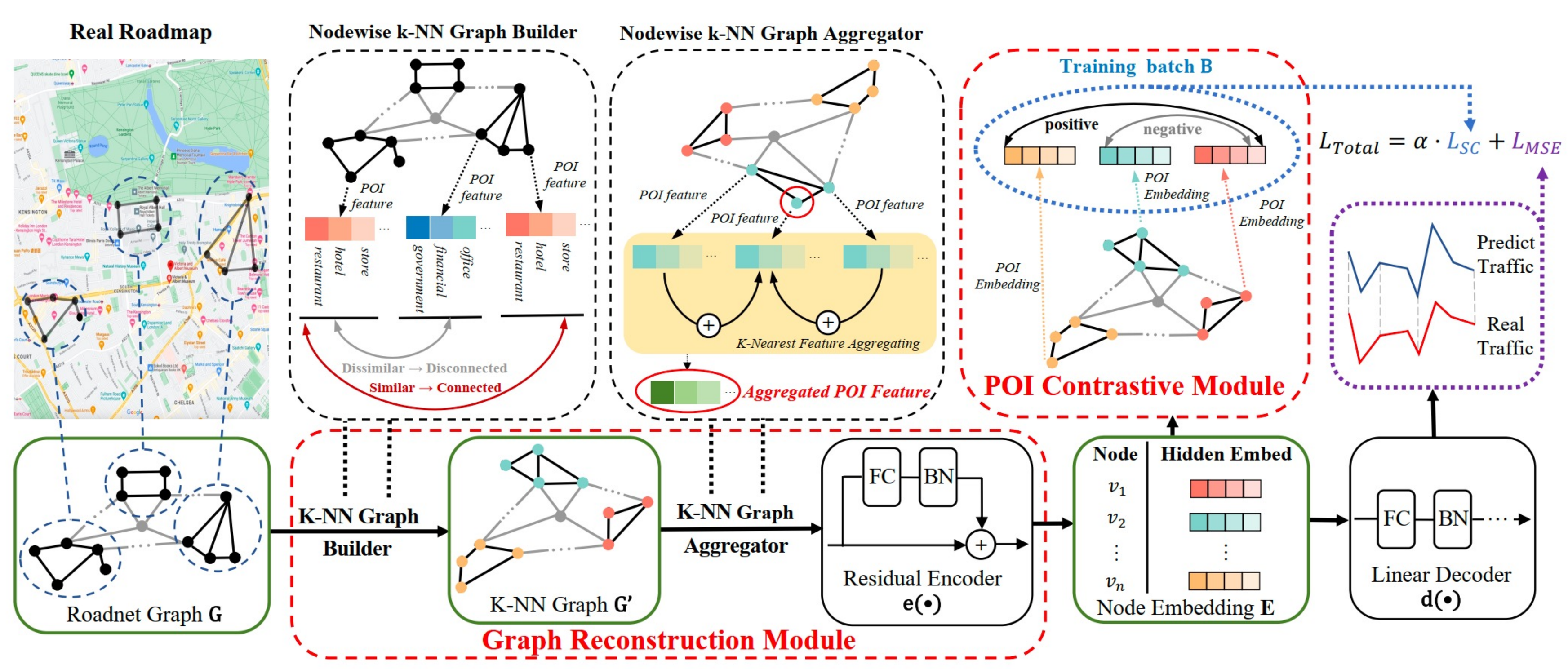
Method	Futian			Nanshan		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
RF	64.28	52.83	35.13%	74.72	30.66	24.88%
XGBoost	76.13	62.01	36.37%	67.13	36.81	26.76%
SVR	262.91	157.56	64.06%	100.18	53.59	51.84%
GCN	173.57 ± 4.17	102.46 ± 2.61	$49.76\% \pm 3.54\%$	92.57 ± 0.93	66.50 ± 0.43	$25.38\% \pm 1.39\%$
GraphSAGE	102.64 ± 3.04	74.62 ± 2.56	$42.94\% \pm 2.59\%$	90.18 ± 1.00	63.59 ± 0.46	$24.73\% \pm 0.91\%$
GAT	92.47 ± 1.57	68.47 ± 0.47	$39.83\% \pm 0.82\%$	71.48 ± 1.31	31.09 ± 0.55	$21.41\% \pm 1.57\%$
GIN	79.33 ± 1.11	58.92 ± 0.54	$40.97\% \pm 0.60\%$	65.25 ± 0.62	29.59 ± 0.28	$21.08\% \pm 0.79\%$
DFG	82.39 ± 0.46	63.64 ± 0.37	$38.61\% \pm 0.65\%$	72.46 ± 0.57	36.47 ± 0.30	$23.35\% \pm 0.59\%$
DeepCrowd	92.46 ± 1.38	67.54 ± 0.70	$38.45\% \pm 0.86\%$	74.36 ± 0.37	36.46 ± 0.23	$23.72\% \pm 0.34\%$
TG-SCR	$\mathbf{53.19 \pm 0.40}$	$\mathbf{41.03 \pm 0.21}$	$\mathbf{30.06\% \pm 0.69\%}$	$\mathbf{51.56 \pm 0.42}$	$\mathbf{27.43 \pm 0.24}$	$\mathbf{19.05\% \pm 0.29\%}$

Method	Beijing			Shanghai		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
RF	9.03	4.56	17.62%	6.62	3.02	13.04%
XGBoost	14.52	7.33	29.24%	7.81	4.50	18.89%
SVR	19.12	10.16	36.34%	25.05	13.57	46.84%
GCN	20.46 ± 1.46	13.49 ± 0.84	$56.20\% \pm 1.95\%$	18.34 ± 0.87	11.52 ± 0.33	$42.71\% \pm 3.73\%$
GraphSAGE	22.56 ± 1.36	14.22 ± 0.83	$53.37\% \pm 1.86\%$	14.36 ± 1.31	9.64 ± 0.83	$41.69\% \pm 5.31\%$
GAT	18.92 ± 1.30	12.38 ± 0.56	$33.27\% \pm 2.02\%$	9.44 ± 1.03	6.82 ± 0.95	$28.59\% \pm 3.51\%$
GIN	10.29 ± 0.99	6.22 ± 0.40	$24.68\% \pm 2.42\%$	6.57 ± 0.17	3.79 ± 0.06	$21.06\% \pm 1.42\%$
DFG	3.48 ± 0.16	2.29 ± 0.06	$9.55\% \pm 0.21\%$	2.39 ± 0.04	1.74 ± 0.02	$9.54\% \pm 0.60\%$
DeepCrowd	10.29 ± 1.04	6.22 ± 0.46	$24.86\% \pm 2.89\%$	6.57 ± 0.09	4.25 ± 0.03	$16.63\% \pm 0.92\%$
TG-SCR	$\mathbf{1.47 \pm 0.07}$	$\mathbf{1.10 \pm 0.05}$	$\mathbf{5.12\% \pm 0.14\%}$	$\mathbf{1.67 \pm 0.03}$	$\mathbf{1.26 \pm 0.02}$	$\mathbf{7.31\% \pm 0.83\%}$

Method	Futian			Nanshan		
	RMSE	MAE	MAPE	RMSE	MAE	MAPE
TG-SCR	$\mathbf{53.19 \pm 0.40}$	$\mathbf{41.03 \pm 0.21}$	$\mathbf{30.06\% \pm 0.69\%}$	$\mathbf{51.56 \pm 0.42}$	27.43 ± 0.24	$\mathbf{19.05\% \pm 0.29\%}$
TG-SCR w/o GRM	55.67 ± 0.42	42.63 ± 0.21	$31.28\% \pm 1.66\%$	52.72 ± 0.61	$\mathbf{27.19 \pm 0.23}$	$19.28\% \pm 1.08\%$
TG-SCR w/o GRM & SCM	92.46 ± 1.38	67.54 ± 0.70	$38.40\% \pm 0.86\%$	72.46 ± 0.57	36.47 ± 0.42	$23.58\% \pm 8.36\%$

Pipeline of Our Proposed Model: TG-SCR

- **Graph Reconstruction Module.** This module includes a *k-NN Graph Builder* and a *k-NN Graph Aggregator* to reform the original roadnet graph and eliminate biases.
- **POI Contrastive Module.** This module adds a *supervised contrastive loss* to help TG-SCR explicitly hold the former similarity relationships.



- Training algorithm of **TG-SCR**:

Algorithm 1 Training procedure of TG-SCR

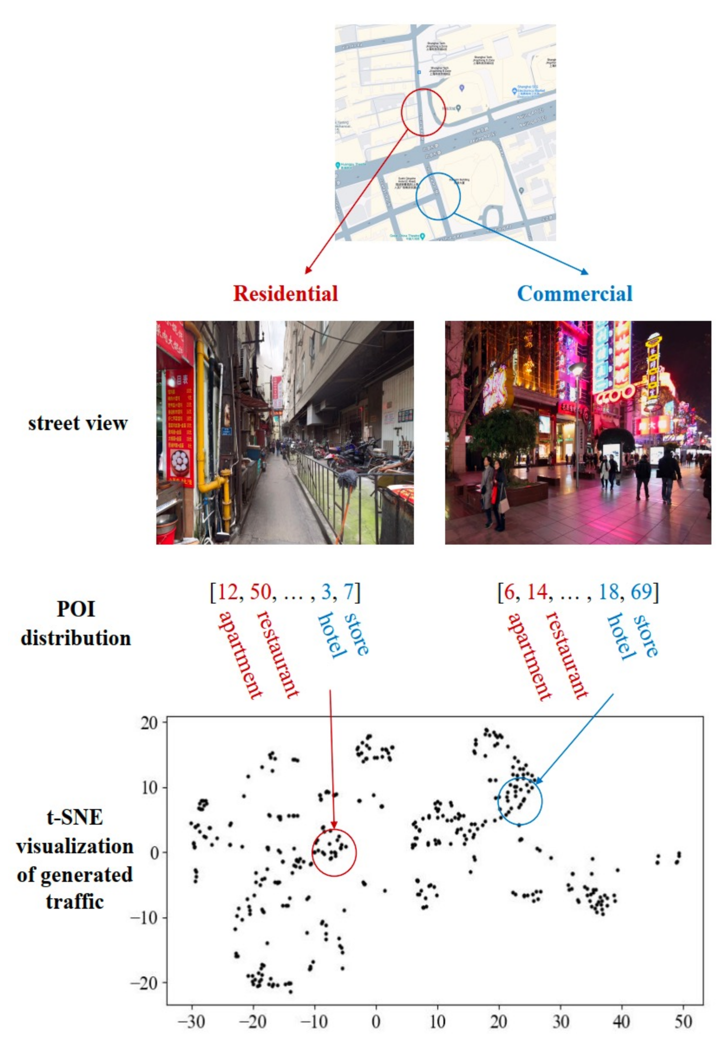
Input: Roadnet graph G with map-matched POI feature X, encoder $e(\cdot)$ and decoder $d(\cdot)$ parameterized with θ

1: Build k-NN graph G' from G based on Eq (3)
2: Get Aggregated POI feature X' from X based on Eq (4)
3: Compute POI embeddings: $E' = e(X')$
4: **repeat**
5: Sample a training batch B including N nodes
6: Define positive set $\mathbb{P}$ and negative set $\mathbb{N}$ based on Eq (5) and Eq (6)
7: Compute L_{Total} based on Eq (9)
8: Back Propagation
9: Update θ based on Eq (7)
10: **until** $L_{Total} < \epsilon$

Output: Predicted traffic $\hat{Y}$

Real-World Case in Shanghai

- Despite close in graph, the generated traffic flows of nodes in these two locations *have a significant difference.*
- Through t-SNE visualization, TG-SCR keenly captures the POI distribution relationship.



Acknowledgements

This work was sponsored by National Key Research and Development Program of China under Grant No.2022YFB3904204, National Natural Science Foundation of China under Grant No. 62102246, 62272301, and Provincial Key Research and Development Program of Zhejiang under Grant No. 2021C01034. Part of the work was done when the students were doing internships at Yunqi Academy of Engineering.