Uncertainty Quantification of Sparse Travel Demand Prediction with Spatial-Temporal Graph Neural Networks

Aug 16, 2022

Dingyi Zhuang
dingyi@mit.edu

Shenhao Wang*
shenhao@mit.edu
shenhaowang@ufl.edu

Haris N. Koutsopoulos
h.koutsopoulos@northeastern.edu

Jinhua Zhao
jinhua@mit.edu

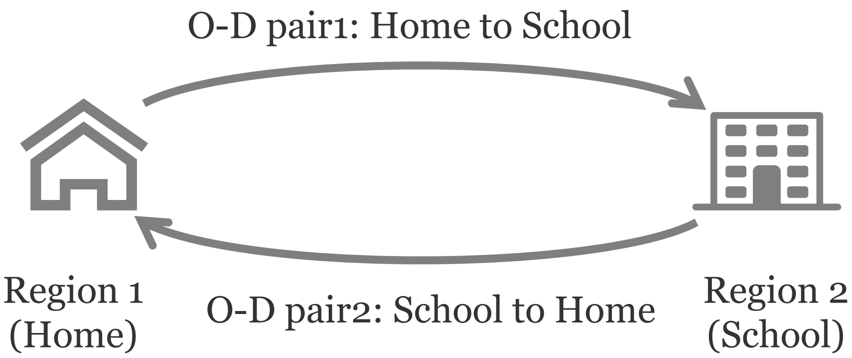
Dingyi Zhuang: About Me

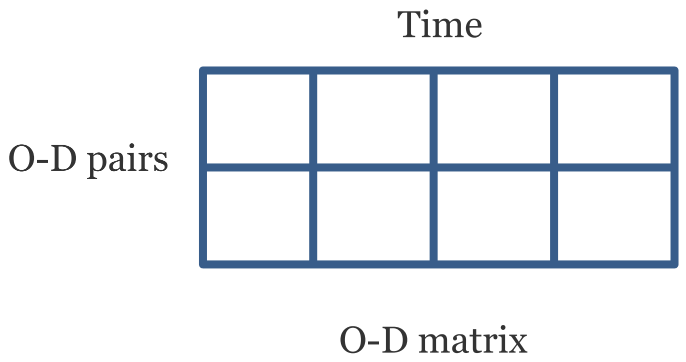
- 2nd year Ph.D. of Transportation Engineering @MIT JTL Urban Mobility Lab
- Main research interests: urban computing, spatial-temporal data mining, and graph neural networks
- Feel free to reach out dingyi@mit.edu or visit my website https://zhuangdingyi.github.io/

Contents

Travel demand prediction

Origin-Destination (O-D) travel demand prediction which aims to predict the number of passengers' travel demands from one region to another.





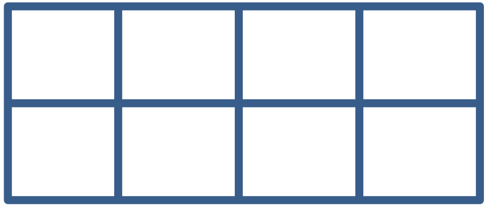
Important applications: intelligent transportation system, on-demand mobility service, public safety, and so on.

Sparsity in mobility service

- Existing travel demand prediction use high-volume and low spatial-temporal resolution mobility service data, e.g. subway stations. The O-D matrices are dense.
- Ride-hailing or bike-sharing service has more granular resolutions. The O-D matrices are sparse.

Every 15min

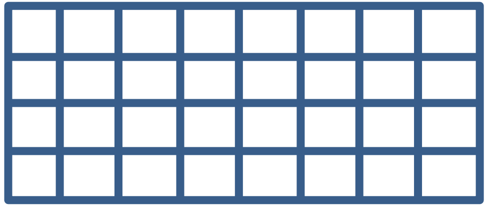
Rail stations



Subway O-D matrix
(daily 300K in Chicago)

Every 5min

Citywide locations



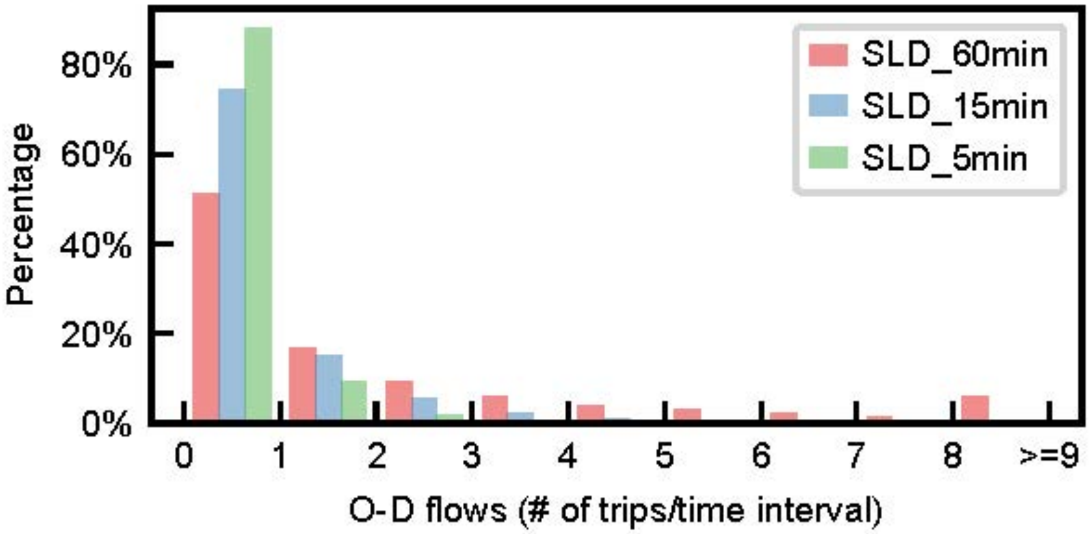
Bike sharing O-D matrix
(daily 10K in Chicago)

Sparsity is a problem

- Existing deep learning models for travel demand prediction target dense O-D matrices and apply Gaussian assumptions.
- A large number of zeros in sparse O-D matrices deviates from Gaussian distribution.
- We need discrete distributions like the negative binomial distribution but hope it can capture the zeros

Sparsity is ubiquitous

- Sparsity is ubiquitous in prediction if the resolutions go up



- Zero demand is important in transportation management. It means no trips instead of missing data. We not only need to predict the non-zero entries, but also predict zeros.

Contributions

Our main contributions include:

- We propose the STZINB-GNN to quantify the spatial-temporal uncertainty of O-D travel demand using a parameter π to learn data sparsity

- The parameters of the probabilistic GNNs successfully quantify the sparse and discrete uncertainty, particularly in high-resolution data sets

- STZINB-GNN outperforms other models by using two real-world travel demand datasets with various spatial-temporal resolutions

Contents

Problem description

- Consider m origins and u destinations with travel demand in the window of length T min.
- Consider all possible O-D pairs as node sets V, with $|V| = m \times u$. Construct O-D graph $\mathcal{G} = (V, E, A)$ with edge set E and $A \in \mathbb{R}^{|V| \times |V|}$ represent O-D pairs' relationship.
- Denote $X_t \in \mathbb{N}^{|V| \times T}$ the demand for all O-D pairs in the t-th time window.
- Our goal is to leverage historical records $X_{1:t}$ as data inputs to predict the distribution of future $X_{t:t+k}$ (k time windows ahead).

Adjacency matrix for O-D pairs

We use O-D pairs instead of regions as vertices V. The adjacency matrix should model spatial correlations of O-D pairs. Consider O-D pairs i and j:

$$A_{i,j}^{O} = haversine\left(lng_i^O, lat_i^O, lng_j^O, lat_j^O\right)^{-1}, \forall i, j \in V$$

$$A_{i,j}^{D} = haversine\left(lng_i^D, lat_i^D, lng_j^D, lat_j^D\right)^{-1}, \forall i, j \in V$$

$$A_{i,j} = \sqrt{\frac{1}{2}\left(\left(A_{i,j}^{O}\right)^2 + \left(A_{i,j}^{D}\right)^2\right)}$$

The basic idea is to leverage the O-D pairs' geographical adjacency. Future studies can assign different weights to the origins or destinations or even combine with demographic graphs.

ZINB distribution

A random variable that follows NB distribution has a probability mass function:

$$f_{NB}(x_k; n, p) = \binom{x_k + n - 1}{n - 1}(1-p)^{x_k} p^n$$

where n and p are the shape parameters that determine the number of successes and the probability of a single success.

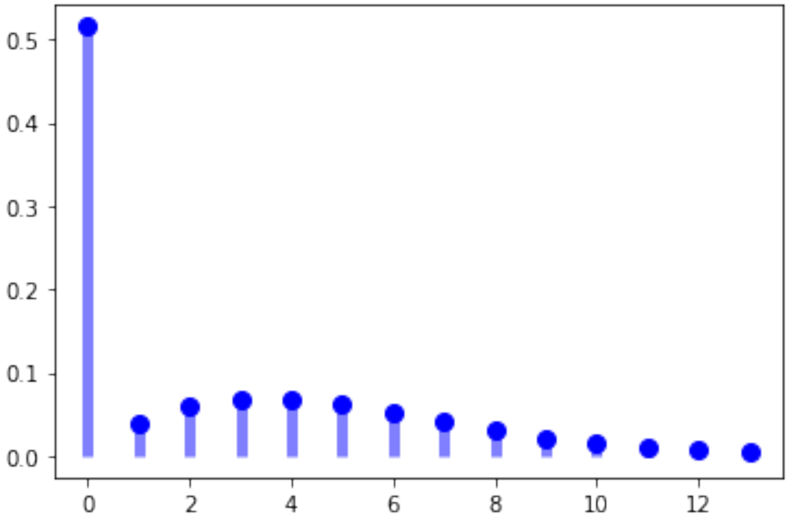
In ZINB distribution, a new parameter π is introduced to learn the inflation of zeros.

$$f_{ZINB}(x_k; n, p, \pi) = \begin{cases} \pi + (1-\pi) f_{NB}(0; n, p), if\ x_k = 0 \\ (1-\pi) f_{NB}(0; n, p), if\ x_k > 0 \end{cases}$$

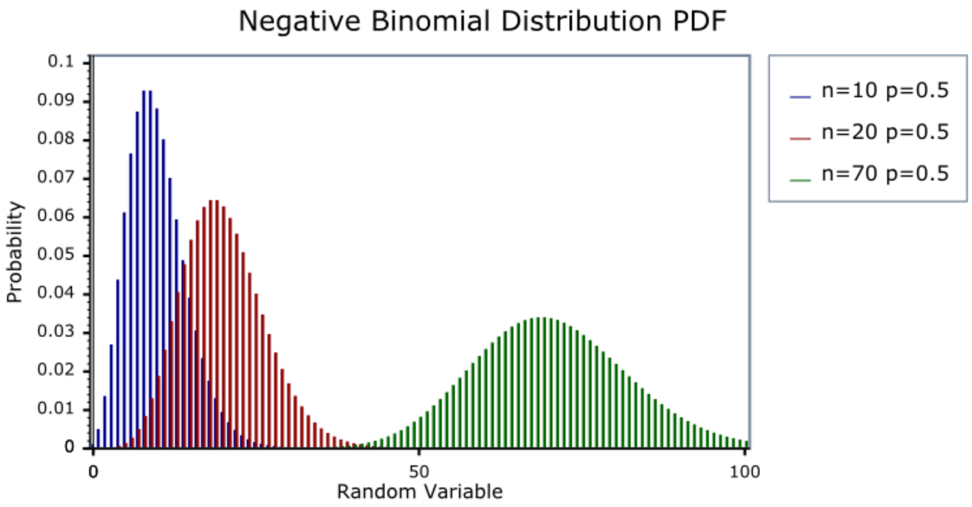
It is either zeros with probability π or NB distribution with probability $1 - \pi$.

NB and ZINB

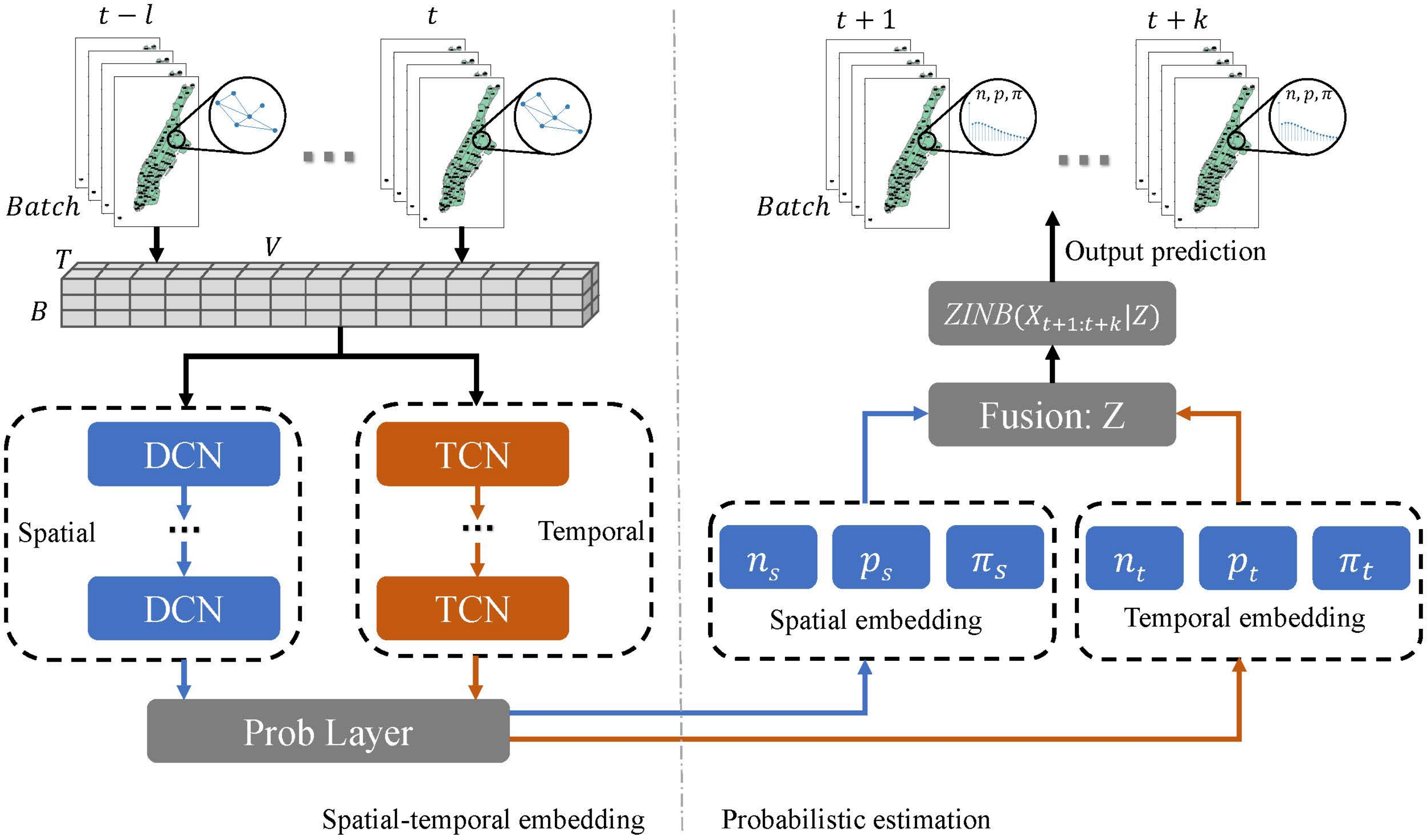
ZINB PDF



NB PDF



STZINB-GNN framework



STZINB-GNN probabilistic loss

Log likelihood of ZINB distribution:

$$LL_y = \begin{cases} \log \pi + \log (1-\pi)p^n \text{ , } when\ y = 0 \\ \log 1 \ -\pi + \log \Gamma(n+y) - \log \Gamma(y+1) - \log \Gamma(n) + n\log p + y\log (1-\pi), when\ y > 0 \end{cases}$$

where π, n, p are also selected according to the index of $y = 0$ or $y > 0$. The final negative log likelihood loss is:

$$NLL_{STZINB} = -LL_{y=0} - LL_{y>0}$$

Code available: https://github.com/ZhuangDingyi/STZINB

Contents

Dataset

Chicago Data Portal (**CDP**[1])

- Trip pick-up and drop-off records of ride-sharing companies in the Chicago area
- The city of Chicago is divided into 77 zones (with 77×77 O-D pairs). Temporal resolution is fixed 15min. We vary the spatial resolutions.

Smart Location Database (**SLD**[2])

- For-hire vehicle pick-up and drop-off trips in the Manhattan area
- The city is divided into 67 zones. We vary the temporal resolutions.

Both datasets use 4 months in total. 60%, 10%, and 30% of data are used for training, validation, and testing.

1. https://data.cityofchicago.org/Transportation/Transportation-Network-Providers-Trips/m6dm-c72p
2. https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

Sparsity in datasets

Name	Resolution	# of O-D pairs	Data size	Zero rate
CDP_SAMP10	15 min	10×10	(100,11521)	81%
SLD_SAMP10	15 min	10×10	(100,11520)	54%
SLD_5min	5 min	67×67	(4489,34560)	88%
SLD_15min	15 min	67×67	(4489,11520)	70%
SLD_60min	60 min	67×67	(4489,2880)	60%

The zero rate increases from 50% to **88%** as the temporal resolution for SLD dataset increases from 60 minutes to 5 minutes.

Evaluation metric

Numeric estimation metrics:

- Mean absolute error: $MAE = \frac{1}{k|V|}\sum_{i=1}^{k|V|} |x_i - \widehat{x_i}\ |$
- True zero rate: $\frac{\sum 1_{\widehat{x_i}=0, x_i=0}}{\sum 1_{, x_i=0}}$
- F1 score: $\frac{precision \times recall}{precision + recall}$

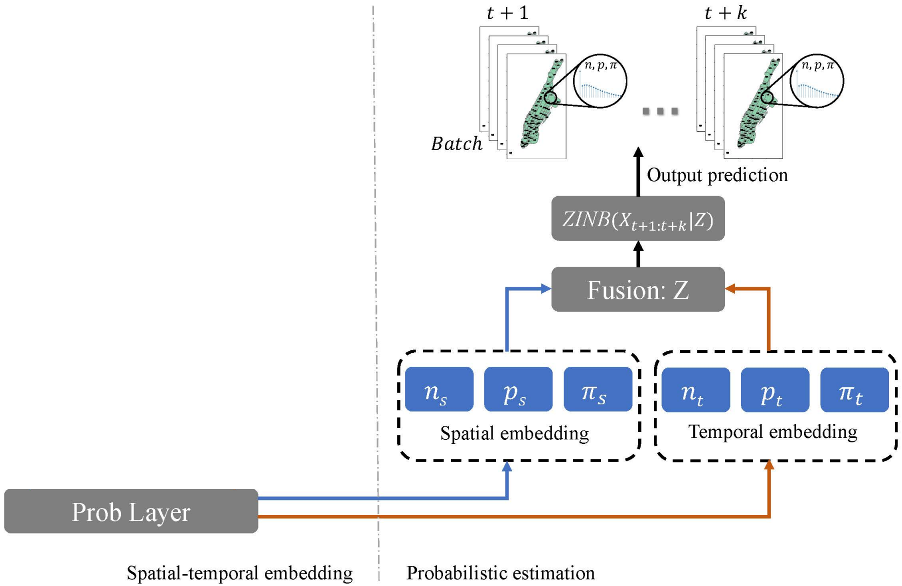
Probabilistic metrics:

- Mean prediction interval width[1]: $\mathrm{MPIW} = \frac{1}{k|V|}\sum_{i=1}^{k|V|} |U_i - L_i\ |$
- KL-divergence: $\frac{1}{k|V|}\sum_{i=1}^{k|V|}(\hat{x}_i \log \frac{\hat{x}_i + \epsilon}{x_i + \epsilon})$

[1]Khosravi, Abbas, et al. "Lower upper bound estimation method for construction of neural network-based prediction intervals." *IEEE transactions on neural networks* 22.3 (2010): 337-346.

Baseline models

- Historical average (HA)
- Spatial-Temporal Graph Convolution Networks (STGCN)
- Models with different probabilistic assumptions:
  - STNB-GNN
  - STG-GNN
  - STTN-GNN

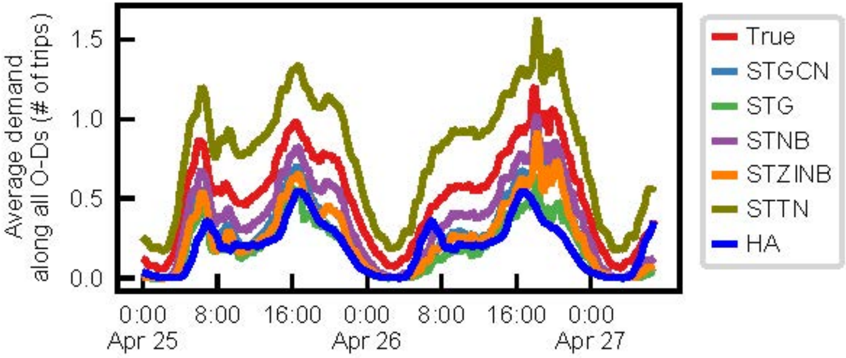


Numerical comparison

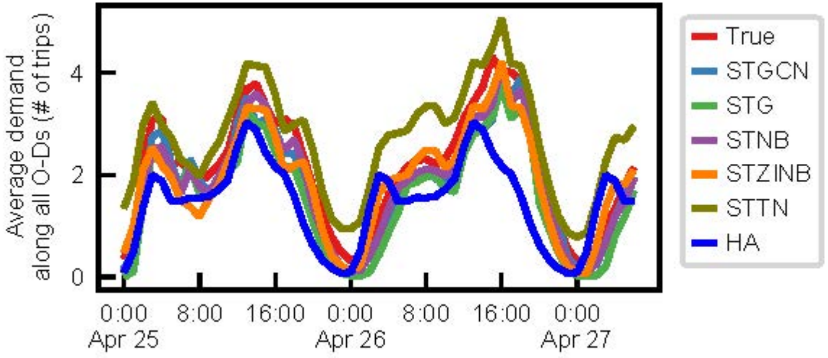
Table 2: Model comparison under different metrics. X/Y values correspond to the mean/median values of the distribution.

Data scenario	Metrics	STZINB-GNN	STNB-GNN	STG-GNN	STTN-GNN	HA	STGCN
CDP_SAMP10	MAE	0.368/**0.366**	0.382/0.379	0.409/0.409	0.432/0.606	0.522	0.395
	MPIW	**1.018**	1.020	2.407	2.089	/	/
	KL-Divergence	**0.291**/0.424	0.342/0.478	0.435/0.435	1.058/0.928	1.377	0.897
	True-zero rate	0.796/0.788	0.796/0.788	0.790/0.790	0.758/0.764	0.759	**0.800**
	F1-Score	**0.848**/0.846	**0.848**/0.841	0.818/0.818	0.842/0.846	0.809	0.840
SLD_SAMP10	MAE	0.663/0.666	0.627/**0.616**	0.630/0.630	0.695/0.665	0.697	0.630
	MPIW	**1.310**	3.628	2.604	1.931	/	/
	KL-Divergence	0.518/**0.507**	0.980/1.662	1.022/1.022	3.578/3.052	0.978	0.768
	True-zero rate	0.499/**0.502**	0.465/0.418	0.461/0.461	0.308/0.336	0.364	0.478
	F1-Score	**0.567**/0.566	0.556/0.552	0.555/0.555	0.477/0.500	0.456	0.563
SLD_5min	MAE	0.149/0.150	0.147/**0.144**	0.155/0.155	0.155/0.155	0.149	0.159
	MPIW	**0.094**	1.249	0.922	0.741	/	/
	KL-Divergence	0.015/0.014	0.042/0.145	**0.001/0.001**	0.001/0.001	0.060	0.056
	True-zero rate	**0.879/0.879**	0.875/0.866	0.877/0.877	0.877/0.877	0.874	0.874
	F1-Score	**0.882/0.882**	0.880/0.878	0.879/0.879	0.879/0.879	0.876	0.879
SLD_15min	MAE	0.370/0.372	0.351/**0.342**	0.356/0.356	0.365/0.356	0.418	0.373
	MPIW	**0.603**	2.283	1.353	1.215	/	/
	KL-Divergence	0.167/**0.156**	0.357/0.704	0.353/0.353	1.445/1.211	0.445	0.395
	True-zero rate	0.725/**0.727**	0.710/0.684	0.709/0.709	0.632/0.648	0.703	0.708
	F1-Score	**0.751**/0.750	0.746/0.745	0.750/0.750	0.716/0.726	0.744	0.750
SLD_60min	MAE	1.040/1.067	0.958/**0.947**	1.199/1.199	1.275/1.254	1.014	0.997
	MPIW	3.277	5.753	2.282	**1.592**	/	/
	KL-Divergence	0.982/1.270	**0.926**/0.963	2.176/2.176	4.120/3.734	2.421	1.114
	True-zero rate	0.458/**0.476**	0.443/0.425	0.390/0.390	0.288/0.308	0.447	0.438
	F1-Score	0.536/0.537	**0.538**/0.534	0.479/0.479	0.407/0.423	0.490	**0.538**

Numerical comparison



SLD_15min



SLD_60min

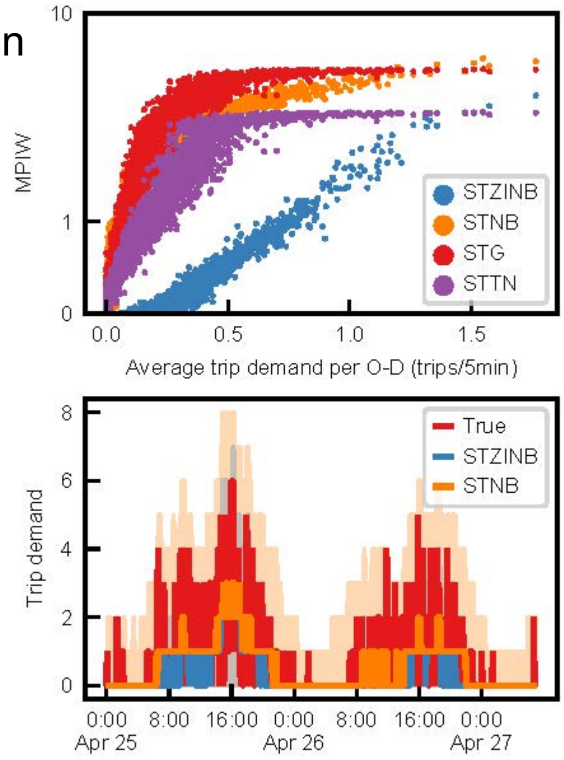
- With finer resolutions, the STNB-GNN and STZINB-GNN models are more likely to accurately predict the average travel demand.
- When the resolution decreases, like in the 60min case, all the deep learning models deliver similar performance.
- 60min resolution is commonly used in the majority of the deep learning studies, but our results demonstrate the importance of discrete probabilistic assumptions when the temporal unit is shorter than 60 minutes

Uncertainty quantification

- We want our model to predict precisely, with smaller confidence intervals
- Model output distributions for future data points
- We want to quantify uncertainty in three resolution cases

Uncertainty quantification

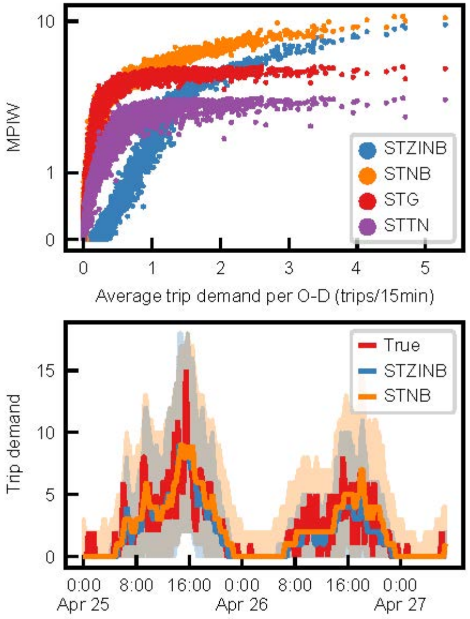
SLD_5min



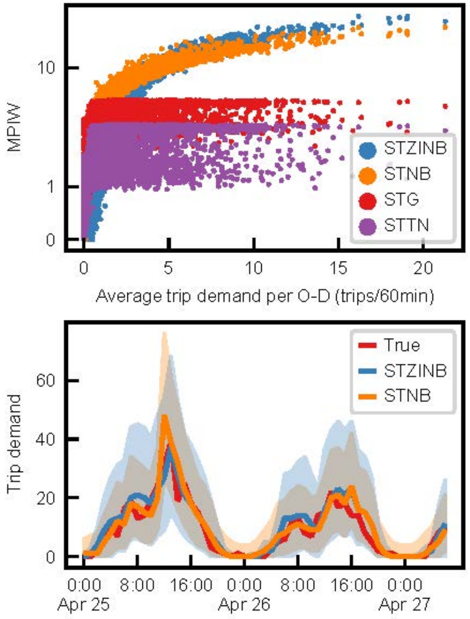
- STZINB-GNN leads to significantly smaller MPIW than the other models in the granular 5min resolution case.
- Average travel demand in the SLD_5min case are mostly zeros. The sparsity parameter π captures the zeros.
- The STG-GNN and STTN-GNN are not able to capture the skewness of the data distribution, leading to large MPIWs.

Uncertainty quantification

SLD_15min

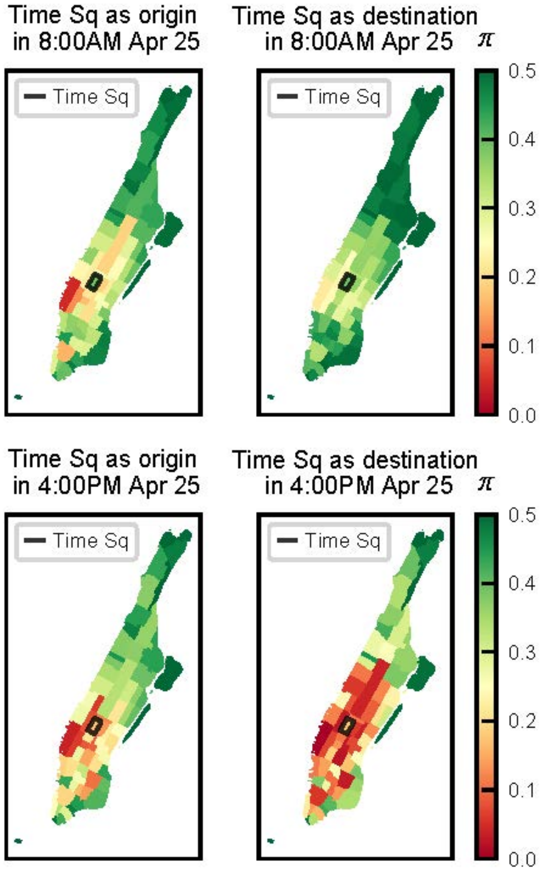


SLD_60min



SLD_15min and SLD_60min cases have larger average travel demand per 15 and 60 minutes time interval. In these two cases, the STZINB-GNN has smaller MPIW only when the average travel demand is small.

Physical interpretation



- Sparsity parameters π represent the inflow/outflow activity of the region
- Spatial locality exists, where communities are more likely to commute to their neighbors
- The variant temporal patterns are also obvious in AM and PM peaks

Contents

Conclusions and takeaways

We introduce STZINB-GNN framework:

- with a sparsity parameter π for sparse travel demand prediction.
- embeds the spatial and temporal representation of distribution parameters for each spatial-temporal data point.
- outperforms the baseline models when the data are represented in high resolutions but performs worse when the resolution becomes coarser.
- has tighter prediction-intervals, compared to other baseline models
- has physical meanings, which could help transportation decision makers assign mobility services to zero or non-zero demand areas
- can be extended to other prediction tasks that use highly sparse data points, such as anomaly detection and accident prediction

Potential cooperation

- Spatial-temporal kriging
- Deep hybrid choice model with urban road networks
- Autonomous vehicles and control (with Prof. Cathy Wu)
- Long-horizon demand prediction (virtual nodes or graph normalization techniques)
- Uncertainty quantification in deep learning

Other topics working on:

- Implementing RL based control on Chicago to relieve bus bunching
- Deep hybrid choice model + different unstructured data type
- Mobility pattern evolving before, during, and after COVID periods
- Computational fairness

Thanks! Questions?