Urban Language Models

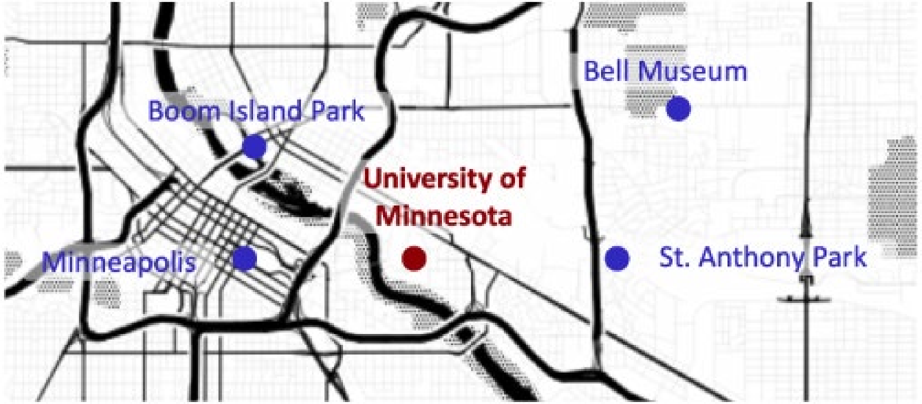
Shaohuai Liu

Content

- SpaBERT
  - Encoder-only model, learn spatial representations of geo-entities for down stream tasks
- GeoLM
  - Contrastive learning between natural language and SpaBERT
- UrbanGPT
  - Spatio-temporal model, prediction task only.

SpaBERT

- Context helps understanding the central token
  - Linguistic context:
    - *The scientist's explanation was so convoluted that even the students with the best grades struggled to understand it.*
  - Surrounding geo-entities also help



SpaBERT

- Problem setting
  - Generate a contextualized representation for each entity g_i
    - Set of geo-entities $S = \{g_1, \cdots, g_l\}$, $g_i = (name, loc)$
    - Spatial context of entity g_p: $SC(g_p) = \{g_{n_1}, \cdots, g_{n_k}\}$, $dist(g_p, g_{n_i}) < T$
    - Didn't use graph encoder
  - Use pretrained entity representation in downstream tasks

SpaBERT

- Linearizing neighboring geo-entity names as pseudo sentences

```
[CLS] University of Minnesota [SEP]
Minneapolis [SEP] St.  Anthony Park [SEP]
Bloom Island Park [SEP] Bell Museum [SEP]
```

- Encoding spatial relations

Token Embed.	$[CLS]$	T_1^p	T_2^p	T_3^p	$[SEP]$	$T_1^{n_1}$	$T_2^{n_1}$	$[SEP]$	$T_1^{n_2}$	$T_2^{n_2}$	$T_3^{n_2}$	$[SEP]$
Sequence Pos. Embed.	POS_0	POS_1	POS_2	POS_3	POS_4	POS_5	POS_6	POS_7	POS_8	POS_9	POS_{10}	POS_{11}
Spatial-Coord Embed.	$DSEP$	0	0	0	$DSEP$	$dist_{x,y}^{n_1}$	$dist_{x,y}^{n_1}$	$DSEP$	$dist_{x,y}^{n_2}$	$dist_{x,y}^{n_2}$	$dist_{x,y}^{n_2}$	$DSEP$

$$dist_x^{n_k} = (g_{n_k}^{locx} - g_p^{locx})/Z$$
$$dist_y^{n_k} = (g_{n_k}^{locy} - g_p^{locy})/Z$$

SpaBERT

- Pretraining tasks
  - Masked Language Modeling (MLM)
    - Re-complete randomly masked partial entity names given spatial coordinates

```
[CLS] ### of Minnesota [SEP] Minneapolis
[SEP] St.  ### Park [SEP] ### Island Park
### Bell Museum [SEP]
```

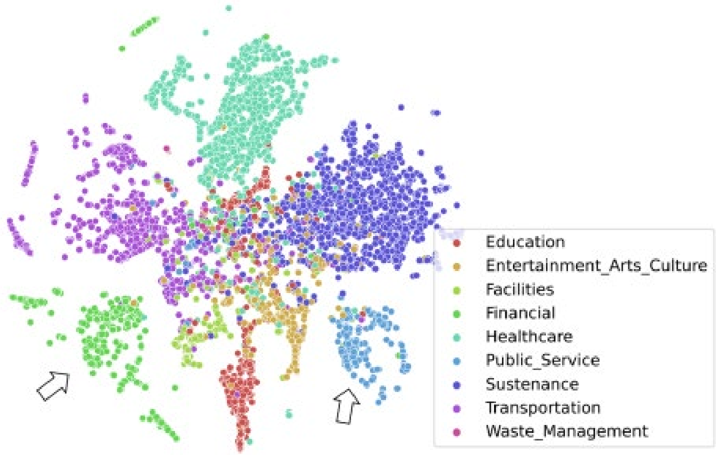
  - Masked Entity Prediction (MEP)
    - Predict the full entity name given spatial coordinates and context.

```
[CLS] University of Minnesota [SEP]
Minneapolis [SEP] ### ### ### [SEP] Bloom
Island Park [SEP] Bell Museum [SEP]
```

  - Pretraining Data
    - OpenStreetMap(OSM), randomly select entities as pivots and construct pseudo sentences

SpaBERT

- Downstream tasks
  - Geo-entity classification
  - Geo-entity link prediction
- Experiments:
  - Entity Classification



Classes →	Edu.	Ent.	Fac.	Fin.	Hea.	Pub.	Sus.	Tra.	Was.	Micro Avg
BERT$_{Base}$	.674	.634	**.763**	.929	.856	.872	.856	.862	.678	.835
RoBERTa$_{Base}$	.626	.627	.605	.951	**.869**	.818	.838	.850	.475	.820
SpanBERT$_{Base}$	.633	.589	.608	.916	.859	.882	.824	.867	**.735**	.819
LUKE$_{Base}$	.648	.608	.598	.945	.857	.867	.854	.851	.517	.825
SimCSE$_{BERT\text{-}Base}$	.623	.590	.504	.925	.867	.852	.857	.810	.470	.810
SimCSE$_{RoBERTa\text{-}Base}$	.621	.629	.499	.951	.841	.853	.828	.856	.500	.814
SpaBERT$_{Base}$	**.674**	**.653**	.680	**.959**	.865	**.900**	**.883**	**.888**	.703	**.852**
BERT$_{Large}$	.707	.661	.647	.937	.874	.850	.873	.864	.526	.841
RoBERTa$_{Large}$	.657	.626	.682	.907	.855	.805	.831	.859	.587	.817
SpanBERT$_{Large}$	.683	.652	.661	.931	.868	.853	.851	.848	.624	.829
LUKE$_{Large}$	.665	.607	.660	.899	.855	.809	.813	.844	.587	.808
SimCSE$_{BERT\text{-}Large}$	.693	.661	**.713**	.940	.880	.871	.864	.867	.564	.844
SimCSE$_{RoBERTa\text{-}Large}$	.683	.630	.648	.916	.865	.802	.807	.848	.587	.811
SpaBERT$_{Large}$	**.731**	**.690**	.710	**.956**	**.901**	**.892**	**.893**	**.903**	**.677**	**.871**

Classes	California	London
Education	6,222	618
Entertainment_Arts_Culture	1,380	601
Facilities	574	179
Financial	2,590	769
Healthcare	3,779	1,779
Public_Service	2,658	393
Sustenance	4,276	1,693
Transportation	4,226	1,618
Waste_Management	167	76
Total	25,872	7,726

SpaBERT

- Experiment
  - Unsupervised Link Prediction
    - A set of entities from Wikidata, and the another larger set from USGS.
    - Do mapping from Wikidata to USGS using cosine similarity.

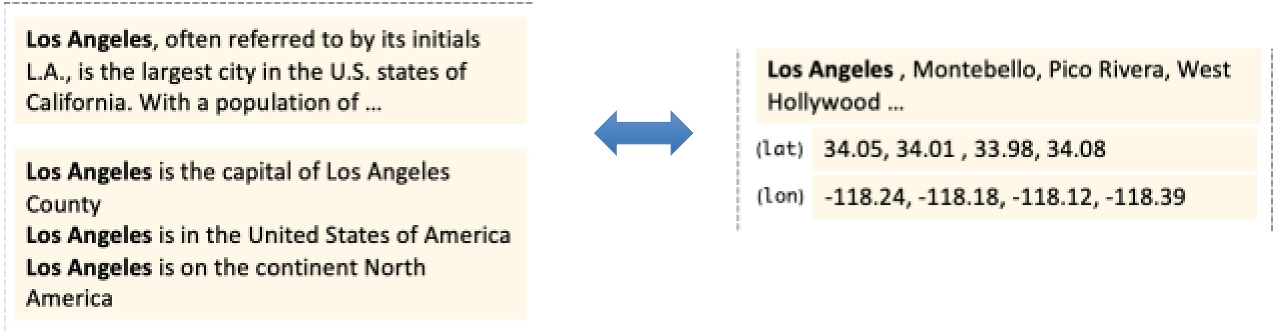
Model	MRR	R@1	R@5	R@10
BERT_{Base}	.400	.289	.559	.635
RoBERTa_{Base}	.326	.232	.446	.540
SpanBERT_{Base}	.164	.138	.201	.213
LUKE_{Base}	.306	.188	.440	.547
$\text{SimCSE}_{BERT\text{-}Base}$	.453	**.371**	.547	.628
$\text{SimCSE}_{RoBERTa\text{-}Base}$	.227	.188	.264	.301
SpaBERT_{Base}	**.515**	.338	**.744**	**.850**
BERT_{Large}	.337	.245	.459	.509
RoBERTa_{Large}	.379	.220	.603	.704
SpanBERT_{Large}	.229	.176	.308	.339
LUKE_{Large}	.402	.232	.635	.767
$\text{SimCSE}_{BERT\text{-}Large}$	.475	**.402**	.559	.616
$\text{SimCSE}_{RoBERTa\text{-}Large}$	.214	.176	.239	.283
SpaBERT_{Large}	**.537**	.383	**.744**	**.864**

Table 3: Geo-entity linking result. Bold and underlined numbers are the highest scores in each column and the highest scores among the baselines, respectively.

$$MRR = \frac{1}{|Q|}\sum_{i=1}^{|Q|}\frac{1}{rank_i}$$

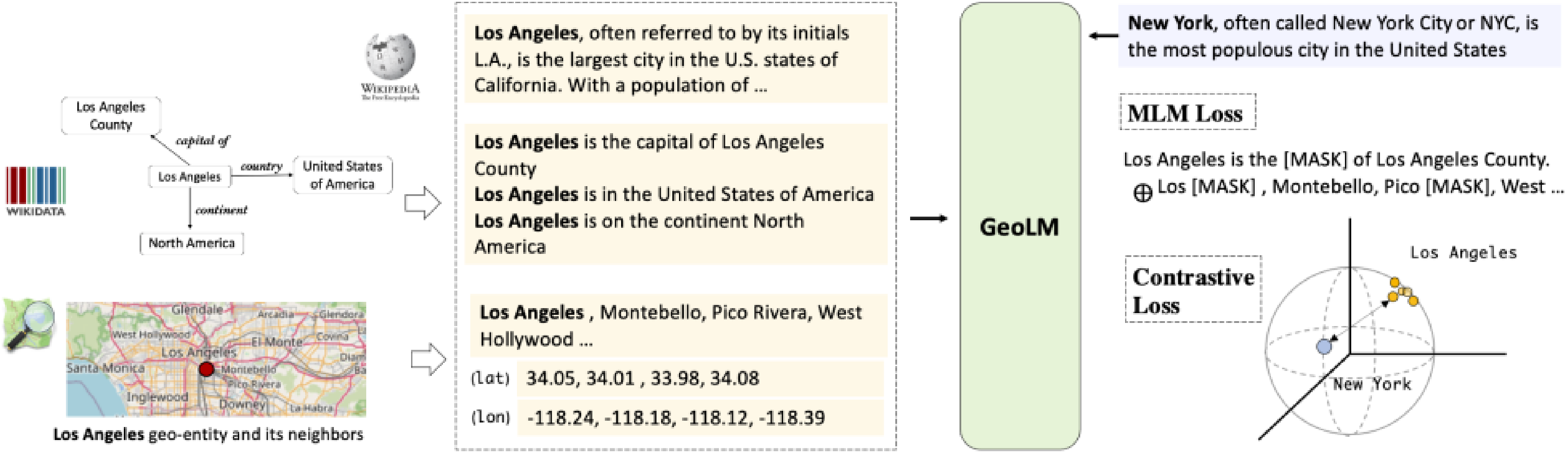
GeoLM

- Main Idea
  - It is unclear if LLM can be strengthened by aligning the pseudo sentences with linguistic descriptions.



GeoLM

- How?
  - Construct geo-corpus from Wikidata.
  - Construct pseudo sentences from OSM following SpaBERT
  - Train LLM using MLM loss
  - Contrastive learning



GeoLM

- How?
  - Tokenize natural language and pseudo sentences in a single framework.

NL Input

Tokens	[CLS]	Los	Angeles	is	the	commercial	,	financial	and	cultural	[SEP]
Position ID	0	1	2	3	4	5	6	7	8	9	10
Segment ID	0	0	0	0	0	0	0	0	0	0	0
X-Coord	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP
Y-Coord	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP	DSEP

$\oplus$ *concat*

Geospatial Input

Tokens	Los	Angeles	[SEP]	Glen	##dale	[SEP]	Pasadena	[SEP]	Al	##ham
Position ID	0	1	2	3	4	5	6	7	8	9
Segment ID	1	1	1	1	1	1	1	1	1	1
X-Coord	34.05	34.05	DSEP	34.17	34.17	DSEP	34.16	DSEP	34.08	34.08
Y-Coord	-118.24	-118.24	DSEP	-118.25	-118.25	DSEP	-118.13	DSEP	-118.13	-118.13

GeoLM

- Pretraining corpus
  - Geographical: OpenStreetMap(OSM)
  - Natural language: Wikidata

- Pretraining tasks
  - **Contrastive learning**

$$\mathcal{L}_i^{contrast} = -\log \frac{e^{\mathrm{sim}(\mathbf{h}_i^{nl}, \mathbf{h}_i^{geo})/\tau}}{\sum_{j=1}^{2N} \mathbb{1}_{[j \neq i]} e^{\mathrm{sim}(\mathbf{h}_i^{nl}, \mathbf{h}_j^{geo})/\tau}},$$

  - Masked Language Modeling(MLM)

GeoLM

- Experiments
  - Entity-name recognition
    - Predict B(begin of entity), I(Inside entity), 0(non-entity) for each token
  - Entity linking
    - Identify the inputs of the same entity from different sources
  - Geo-entity classification

GeoLM

- Entity name recognition

GeoWebNews	Token(B-topo)			Token (I-topo)			micro-	Entity		
	Prec	**Recall**	**F1**	**Prec**	**Recall**	**F1**	**F1**	**Prec**	**Recall**	**F1**
BERT	<u>90.00</u>	89.28	<u>89.64</u>	78.55	79.44	78.99	84.46	<u>77.03</u>	83.42	<u>80.10</u>
SimCSE-BERT	83.86	**90.26**	86.95	74.61	82.07	78.16	82.67	72.76	<u>83.68</u>	77.84
SpanBERT	85.98	88.37	87.16	**86.13**	**89.19**	**87.63**	**87.38**	75.32	81.16	78.13
SapBERT	83.12	88.32	85.64	76.26	81.11	78.61	82.22	72.48	80.16	76.12
GEOLM	**91.15**	<u>90.43</u>	**90.79**	<u>79.16</u>	<u>84.27</u>	<u>81.63</u>	<u>86.33</u>	**82.18**	**85.67**	**83.89**

Table 1: Toponym recognition results on GeoWebNews dataset. **Bolded** and <u>underlined</u> numbers are for best and second best scores respectively.

- Entity Linking

LGL	R@1	R@5	R@10	P@D$_{161}$
BERT	34.6	**67.5**	**78.1**	41.2
RoBERTa	24.2	48.7	60.6	27.9
SpanBERT	25.2	48.8	61.0	28.8
SapBERT	30.8	58.8	72.2	35.1
GeoLM	**38.2**	65.3	72.6	**44.1**

WikToR	P@D$_{20}$	P@D$_{50}$	P@D$_{100}$	P@D$_{161}$
BERT	16.1	16.3	16.9	17.6
RoBERTa	11.7	11.9	12.4	13.0
SpanBERT	5.5	5.7	5.9	6.3
SapBERT	25.9	26.3	27.0	28.3
GeoLM	**32.5**	**33.4**	**34.3**	**35.8**

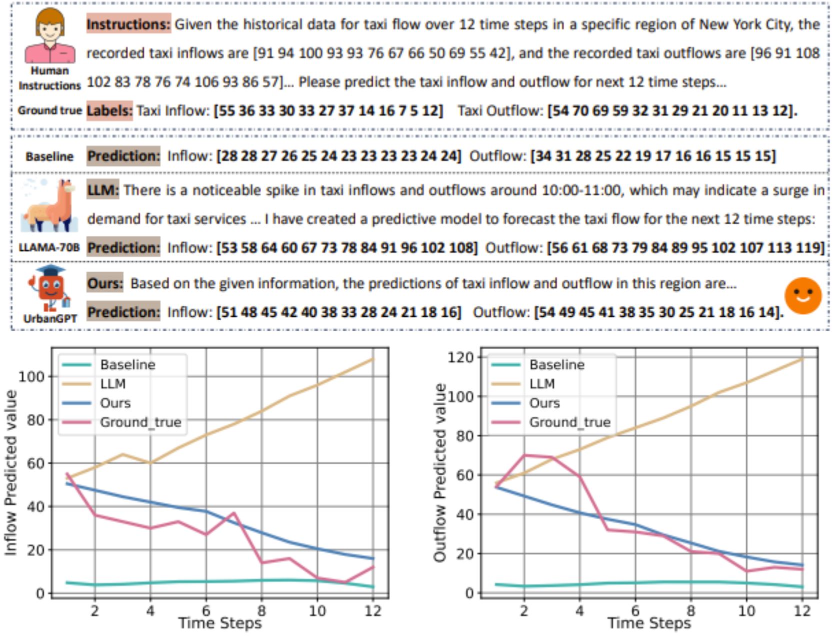
GeoLM

- Entity classification

Classes →	Edu.	Ent.	Fac.	Fin.	Hea.	Pub.	Sus.	Tra.	Was.	Micro F1
BERT	67.4	63.4	**76.3**	92.9	85.6	87.2	85.6	86.2	67.8	83.5
SpanBERT	63.3	58.9	60.8	91.6	85.9	88.2	82.4	86.7	**73.5**	81.9
SimCSE-BERT	62.3	59.0	50.4	92.5	86.7	85.2	85.7	81.0	47.0	81.0
LUKE	64.8	60.8	59.8	94.5	85.7	86.7	85.4	85.1	51.7	82.5
SpaBERT	67.4	65.3	68.0	95.9	86.5	**90.0**	88.3	88.8	70.3	85.2
GeoLM	**72.5**	**70.9**	73.0	**97.8**	**91.5**	83.6	**90.5**	**90.8**	62.2	**87.8**

UrbanGPT

- Main Idea
  - The previous research focus only on spatial-level.
  - Directly applying LLM on sptio-temporal data = inferior zero-shot performance
  - It is necessary to take temporal dependencies into finetuning.



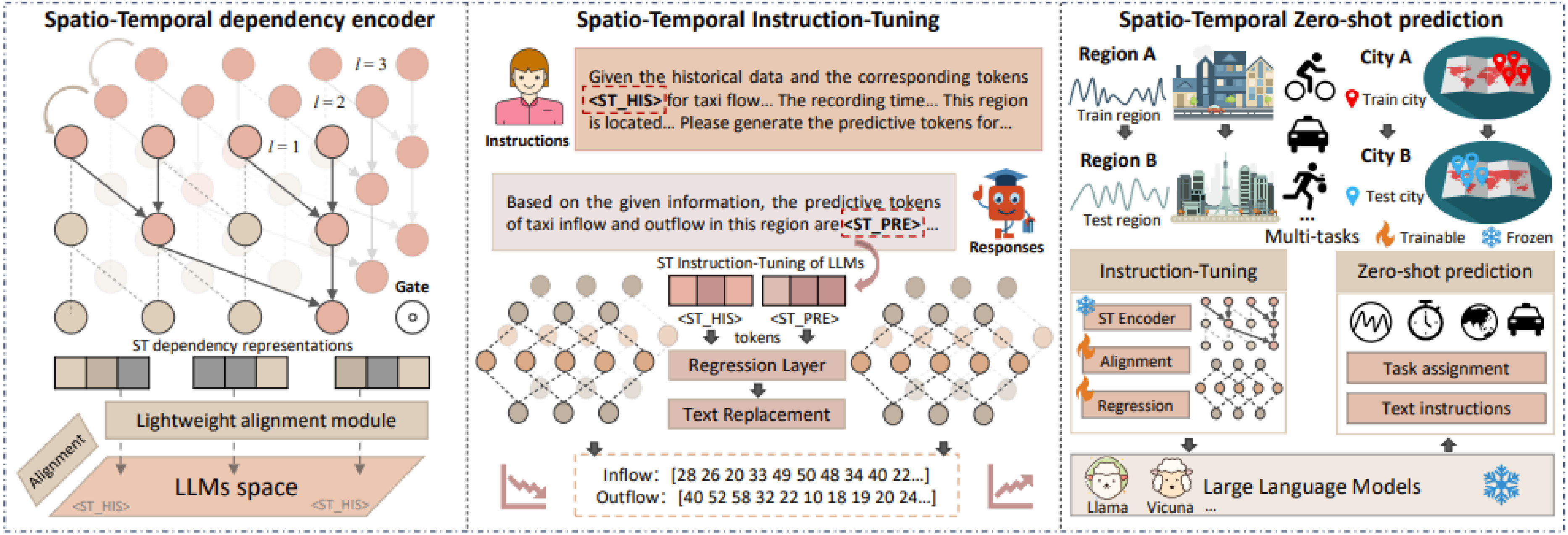
UrbanGPT

- Problem setting
  - Spatio-temporal data: $X \in R^{A \times T \times F}$ (area, time, feature)
  - Spatio-temporal forecast:

$$\mathbf{X}_{t_{K+1}:t_{K+P}} = f(\mathbf{X}_{t_{K-H+1}:t_K}) \quad (1)$$

UrbanGPT

- Overview



UrbanGPT

- Spatio-Temporal Dependency Encoder

$$\Psi_r^{(l)} = (\bar{\mathbf{W}}_k^{(l)} * \mathbf{E}_r^{(l)} + \bar{\mathbf{b}}_k^{(l)}) \cdot \delta(\bar{\mathbf{W}}_g^{(l)} * \mathbf{E}_r^{(l)} + \bar{\mathbf{b}}_g^{(l)}) + \mathbf{E}_r'^{(l)} \quad (3)$$

$$\mathbf{S}_r^{(l)} = (\mathbf{W}_s^{(l)} * \Psi_r^{(l)} + \mathbf{b}_s^{(l)}) + \mathbf{S}_r^{(l-1)} \quad (4)$$

Residual connection

- Model optimization

$$\mathcal{L}_c = -\frac{1}{N} \sum_{i=1}^{N} [\delta(y_i) \cdot \log(\hat{y}_i) + (1 - \delta(y_i)) \cdot \log(1 - \hat{y}_i)]$$

$$\mathcal{L}_r = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|; \quad \mathcal{L} = \mathcal{L}_{LLMs} + \mathcal{L}_r + \mathcal{L}_c \quad (6)$$

L_c for classification tasks

UrbanGPT

- Experiments
    - Zero-shot

Model	Dataset	NYC-taxi				NYC-bike				NYC-crime			
	Type	Inflow		Outflow		Inflow		Outflow		Burglary		Robbery	
	Metrics	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	Macro-F1	Recall	Macro-F1	Recall
AGCRN		10.86	26.51	13.15	36.45	3.41	7.98	3.42	8.08	0.48	0.00	0.49	0.01
ASTGCN		9.75	24.12	12.42	33.28	5.58	11.58	5.78	12.29	0.49	0.01	0.55	0.09
GWN		10.73	26.50	9.67	26.74	3.32	8.17	3.07	7.52	0.48	0.00	0.52	0.04
MTGNN		10.16	25.84	12.59	35.38	3.18	7.62	3.20	7.65	0.64	0.27	0.65	0.30
STWA		11.28	28.97	13.54	38.61	4.59	10.94	4.35	10.67	0.48	0.00	0.51	0.03
STSGCN		18.97	41.38	20.07	45.79	6.85	14.98	6.54	14.77	0.48	0.00	0.48	0.00
STGCN		12.54	30.80	14.32	39.58	4.11	9.21	4.45	9.62	0.48	0.00	0.64	0.30
TGCN		10.04	25.10	10.98	30.03	2.88	6.55	2.91	6.42	0.56	0.10	0.58	0.13
DMVSTNET		11.00	28.29	10.59	29.20	3.80	9.87	3.65	9.21	0.48	0.01	0.59	0.15
ST-LSTM		16.97	34.43	18.93	44.10	7.78	15.41	6.92	17.12	0.48	0.00	0.49	0.03
GPT4TS		9.72	24.51	10.85	31.00	3.16	7.45	3.23	7.53	0.48	0.00	0.49	0.02
UrbanGPT		**6.16**	**16.92**	**6.83**	**21.78**	**2.02**	**5.16**	**2.01**	**5.03**	**0.67**	**0.34**	**0.69**	**0.42**

- Supervised Learning

Table 2: Evaluation of performance in the end-to-end supervised setting on the NYC-taxi and NYC-bike datasets.

Model	NYC-taxi				NYC-bike			
	Inflow		Outflow		Inflow		Outflow	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
AGCRN	2.83	8.35	2.62	9.21	3.30	7.65	3.38	7.73
ASTGCN	5.41	18.04	5.00	19.29	3.87	7.93	3.66	7.69
GWN	3.91	11.93	2.89	10.85	4.30	9.04	3.88	8.29
MTGNN	3.09	10.13	2.61	10.96	3.31	7.47	3.26	7.61
STWA	3.90	12.64	3.15	11.32	4.23	9.07	4.18	9.18
STSGCN	4.57	13.93	4.41	15.87	5.10	12.23	4.72	10.78
STGCN	3.45	9.82	3.17	10.53	3.88	9.23	3.90	9.08
TGCN	3.99	11.47	3.31	11.58	4.12	7.92	4.11	7.84
DMVSTNET	3.83	11.55	2.76	9.88	3.71	7.95	3.69	7.92
ST-LSTM	7.78	15.41	6.92	17.12	5.00	11.52	4.96	11.41
UrbanGPT	**2.50**	**6.78**	**1.71**	**6.68**	**3.11**	**7.10**	**3.01**	**6.94**

Rethinking

- How to appropriately represent data is the key question when applying LLM for specific domain.
- Typically, aligning domain-specific data with natural language description could enhance the model performance.
- If you have a self-designed encoder, finetuning encoder only is all you need.