ST-MoE-BERT: A Spatial-Temporal Mixture-of-Experts Framework for Long-Term Cross-City Mobility Prediction

Haoyu He, Haozheng Luo, Qi (Ryan) Wang

PhD Student in Civil and Environmental Engineering

Northeastern University.

he.haoyu1@northeastern.edu

Outline

Motivation &
Problem Definition

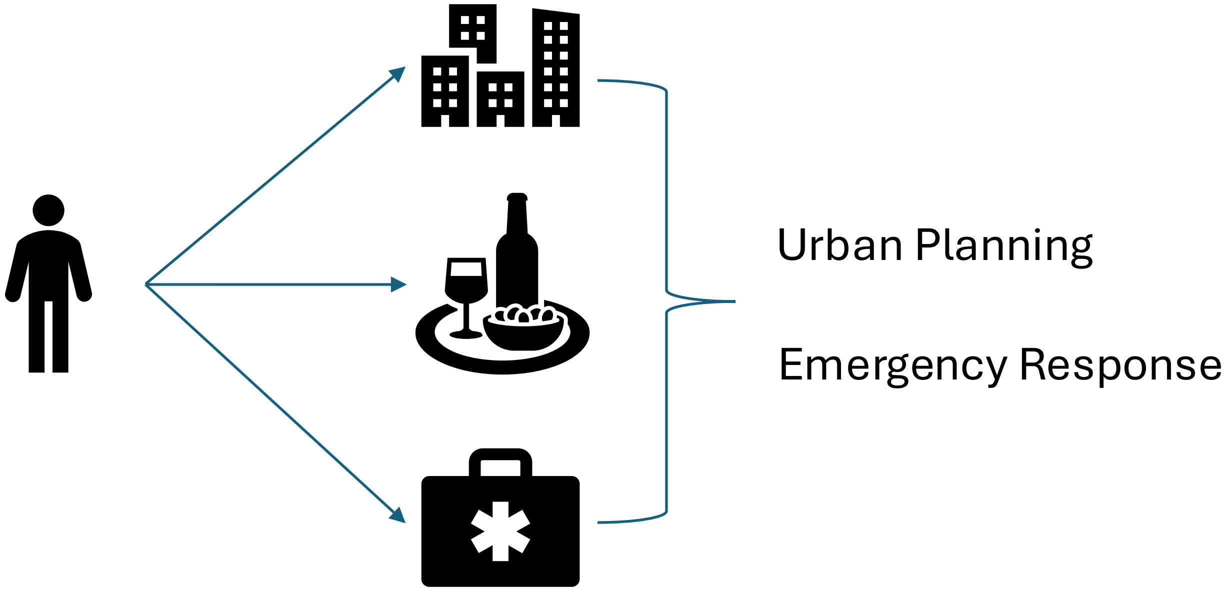
Our Approach

Experiments

1. Motivation & Problem Definition

Human Mobility Prediction Challenge (HuMob Challenge) 2024

Photo by FLY:D on Unsplash



1. Motivation & Problem Definition

Challenges in Human Mobility Prediction

- Data quality: Sparse and unevenly distributed spatially and temporarily
- Complexity of human mobility patterns
- Transfer model between different cities

1. Motivation & Problem Definition

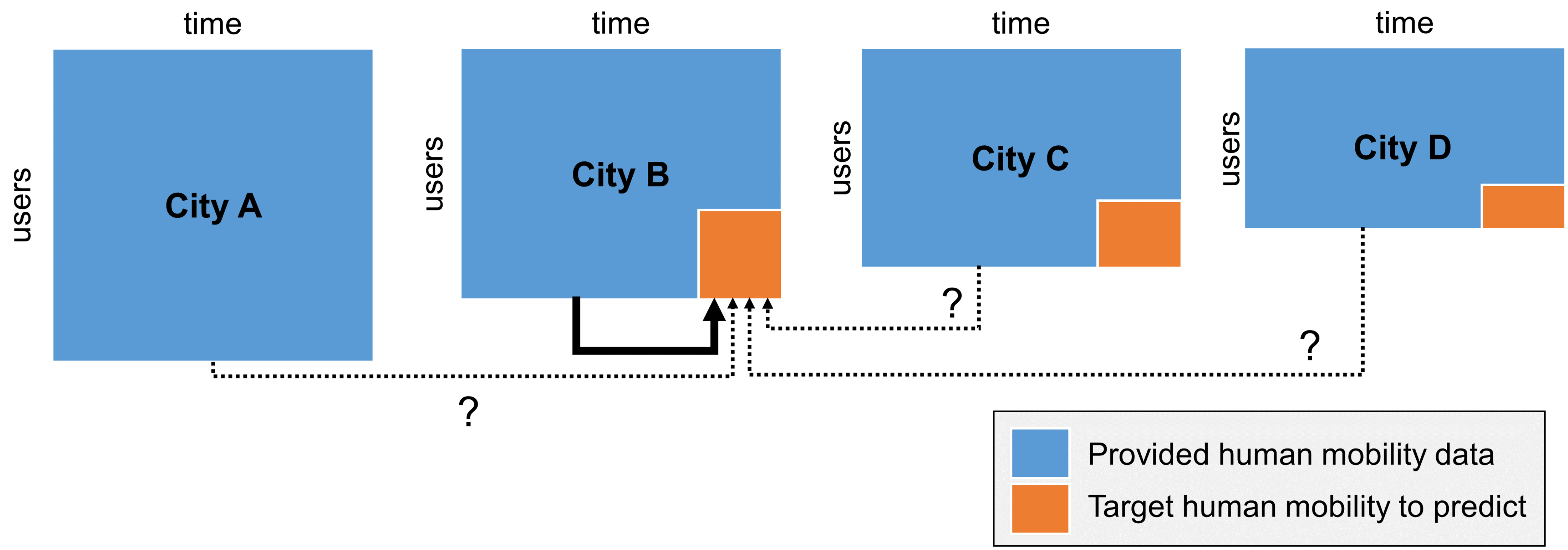
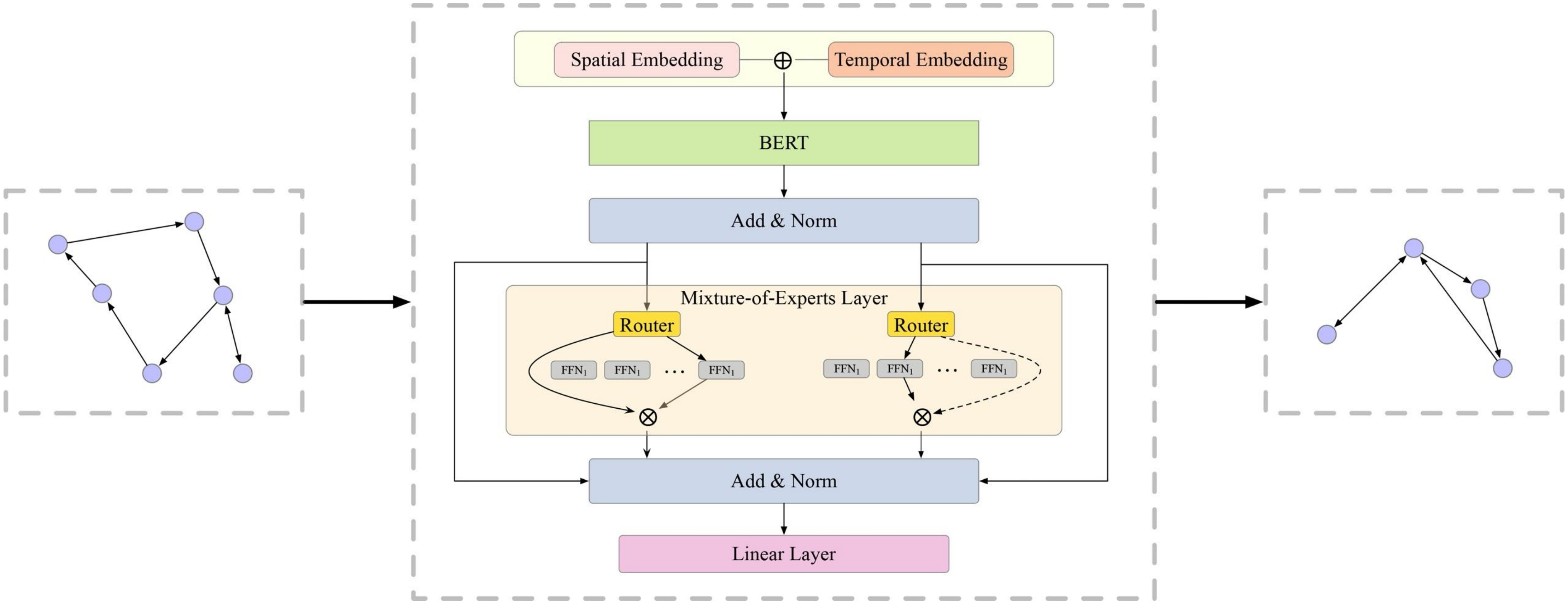


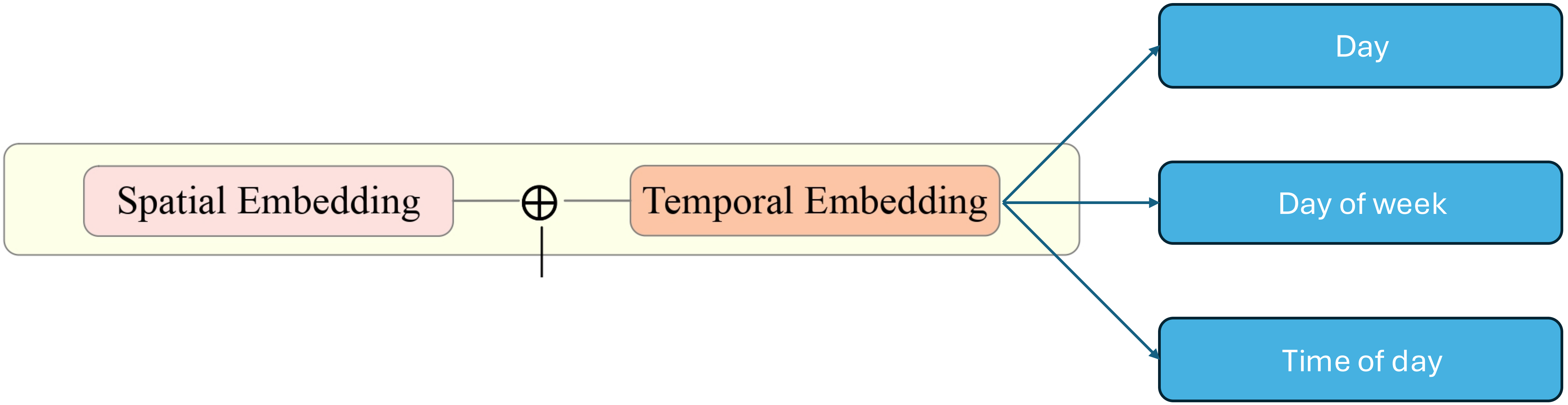
Figure from https://wp.nyu.edu/humobchallenge2024

2. Our Approach



2. Our Approach

Embeddings



2. Our Approach

BERT

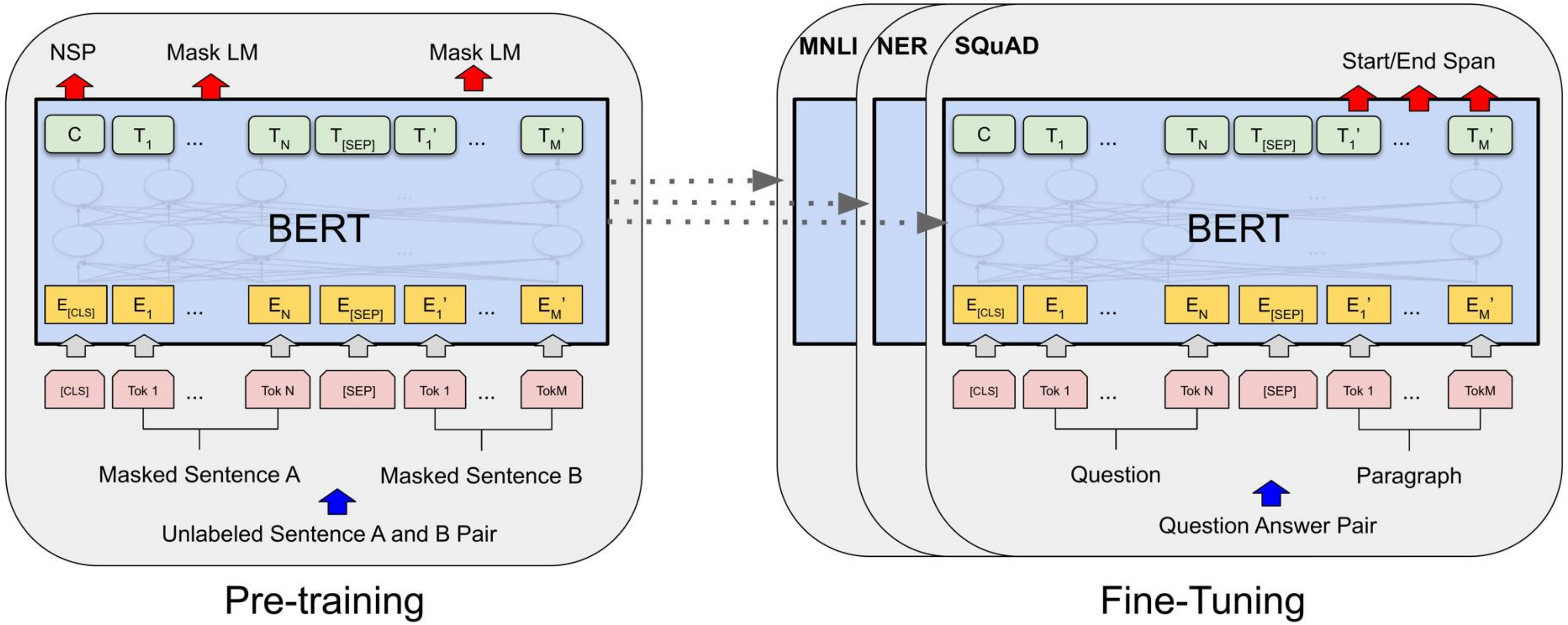
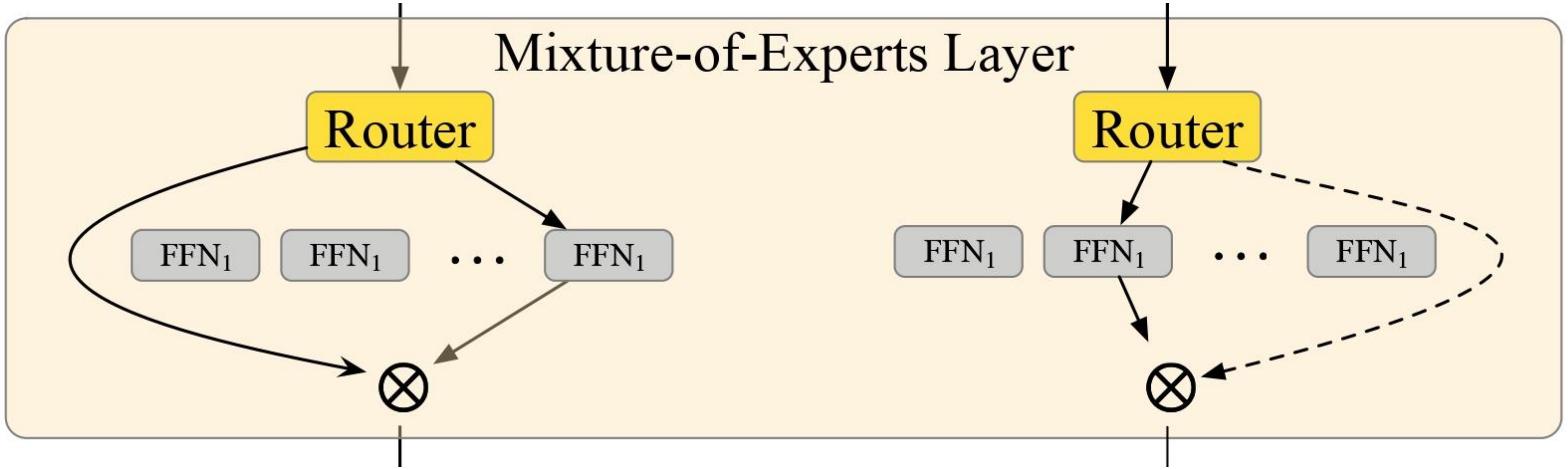


Figure from [1]

[1] Devlin, Jacob. "Bert: Pre-training of deep bidirectional transformers for language understanding." *arXiv preprint arXiv:1810.04805* (2018).

2. Our Approach

Mixture-of-Experts



FFN: Feed-forward network

2. Our Approach

Transfer Learning

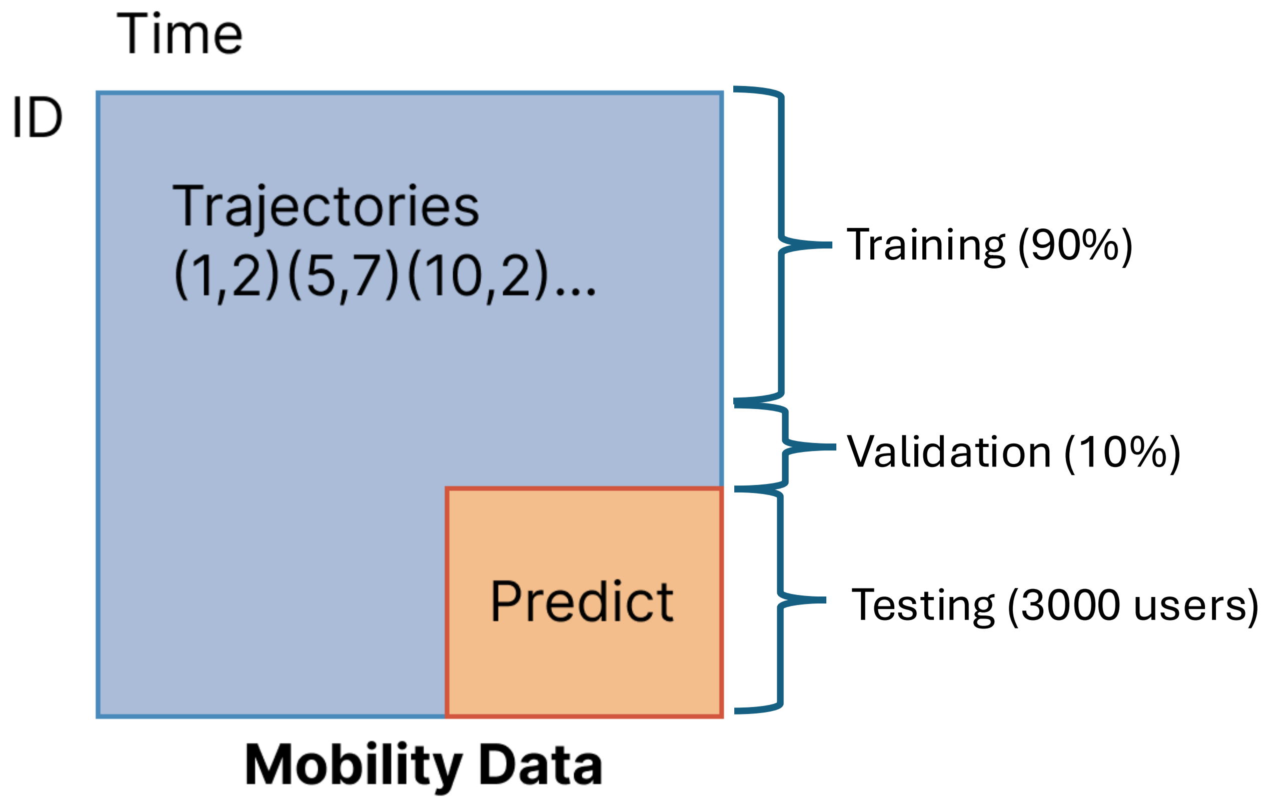
Adapt learned representations from a large-scale mobility dataset of one city to predict mobility patterns in different cities.

Differential Learning Rates:

- **Spatial embeddings:** Higher learning rate (10x the base rate) to quickly adapt to unique spatial characteristics of a new city.
- **General Parameters:** Lower learning rate to retain broad knowledge from the initial training phase.

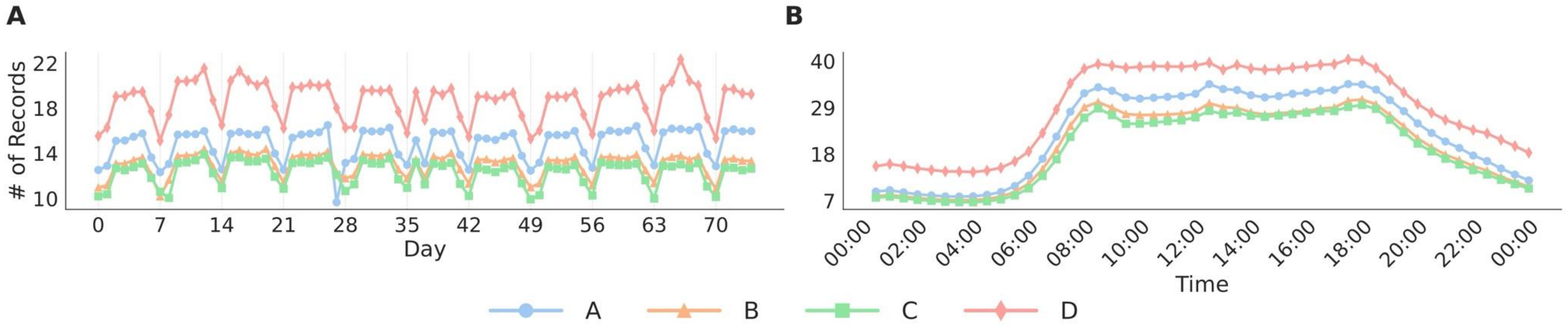
3. Experiments

Data from [1]

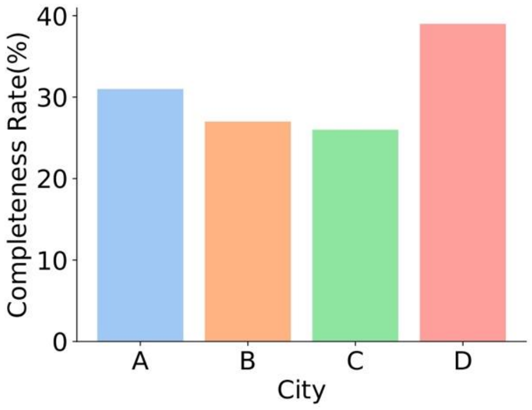


[1] Yabe, Takahiro, et al. "YJMob100K: City-scale and longitudinal dataset of anonymized human mobility trajectories." *Scientific Data* 11.1 (2024): 397.

3. Experiments



Average number of user records per **A** day, **B** timewindow



City	**A**	**B**	**C**	**D**
Number of Users	100,000	25,000	20,000	6,000

3. Experiments

Accuracy

- the percentage of correct location predictions

GEO-BLEU

- evaluates the similarity of geospatial sequences on place n-gram accuracy

Dynamic Time Wrapping (DTW)

- Minimize the cumulative distance over all points, adjusted for temporal shift

3. Experiments

Results

Historical frequency (HF): predicts future locations using historical visit patterns based on time and weekday

Naïve BERT

Method	A			B			C			D		
	GEO-BLEU ↑	DTW ↓	Acc. ↑	GEO-BLEU ↑	DTW ↓	Acc. ↑	GEO-BLEU ↑	DTW ↓	Acc. ↑	GEO-BLEU ↑	DTW ↓	Acc. ↑
HF	0.266	80.3	20.4%	0.265	56.4	21.0%	0.251	42.4	20.8%	0.295	80.0	21.0%
BERT	0.256	35.7	23.8%	0.284	**20.6**	27.0%	0.253	65.6	18.2%	0.253	65.6	18.2%
ST-MoE-BERT	**0.286**	**30.2**	**27.9%**	**0.297**	29.3	**28.7%**	**0.297**	**19.7**	**28.9%**	**0.300**	**48.1**	**26.5%**

= best result

3. Experiments

Ablation study

Impact of Transfer Learning on Prediction Performance

Method	A			B			C			D		
	GEO-BLEU ↑	DTW ↓	Acc. ↑	GEO-BLEU ↑	DTW ↓	Acc. ↑	GEO-BLEU ↑	DTW ↓	Acc. ↑	GEO-BLEU ↑	DTW ↓	Acc. ↑
ST-MoE-BERT w/o PT	**0.286**	**30.2**	**27.9%**	0.286	28.2	27.5%	0.294	20.7	27.9%	0.250	67.6	21.4%
ST-MoE-BERT	-	-	-	**0.297**	29.3	**28.7%**	**0.297**	**19.7**	**28.9%**	**0.300**	**48.1**	**26.5%**

3. Experiments

Hyperparameters

Pretrained Model	
Learning Rate	0.0003
Weight Decay	0.001
Number of Hidden Layers	12
Hidden Size	768
Number of Attention Heads	16
Number of Experts	8
Dropout	0.1
Day Embedding Size	64
Time Embedding Size	64
Day of Week Embedding Size	64
Weekday Embedding Size	32
Location Embedding Size	256
Fine-Tuned Model	
Learning Rate	0.00005
Location Embedding Learning Rate	0.0005

Our Contribution

- We introduce ST-MoE-BERT, a transformer-based method that combines BERT with an MoE layer to predict long-term cross-city mobility
- Transfer learning strategy that employs different learning rates, enhancing prediction accuracy in different cities
- We demonstrate that ST-MoE-BERT outperforms the baseline methods with an average improvement of 8.29%

Thank you!

Haoyu He
he.haoyu1@northeastern.edu
https://he-h.github.io/

Northeastern University