ICDE2023
Anaheim, CA April 3 – 7, 2023

LHMM: A Learning Enhanced HMM Model for Cellular Trajectory Map Matching

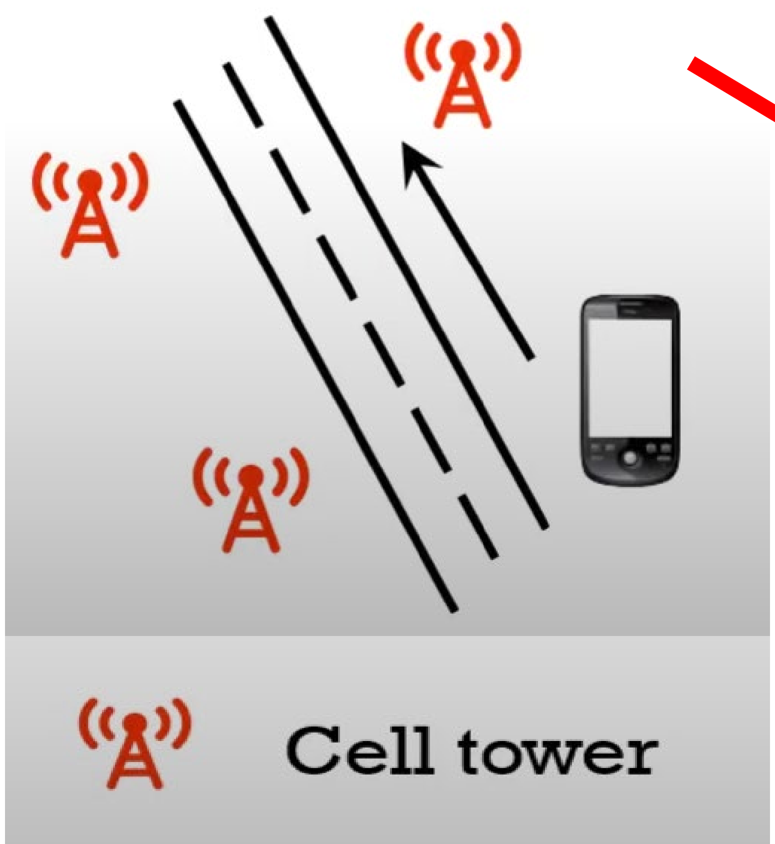
Weijie Shi, Jiajie Xu*, Junhua Fang, Pingfu Chao, An Liu, and Xiaofang Zhou

蘇州大學
SOOCHOW UNIVERSITY

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Two Basic Concepts

Cellular Data



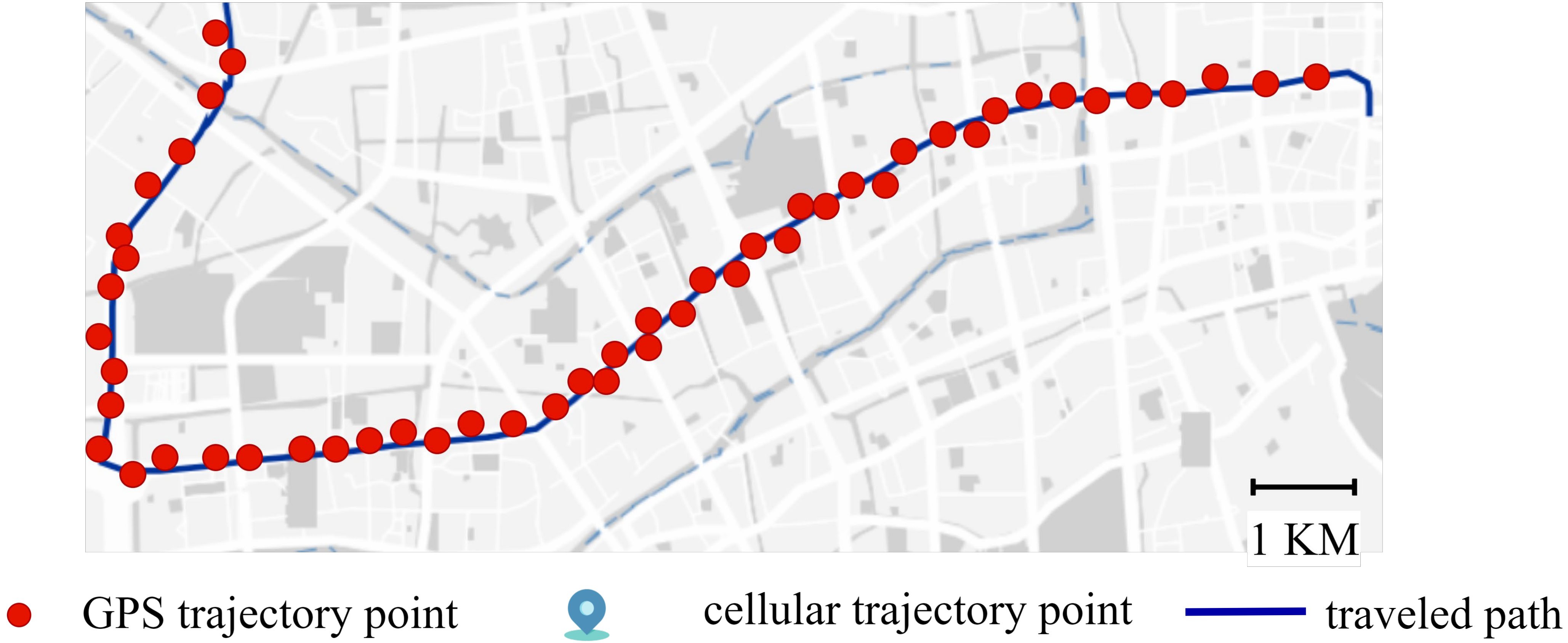
Carrier Server

UserID	Timestamp	Cell towerID
A01	0802081032	37146
A01	0802081046	37149
A01	0802081068	37150
A01	0802081103	37153
A01	0802081148	37121
A01	0802081182	37154

……

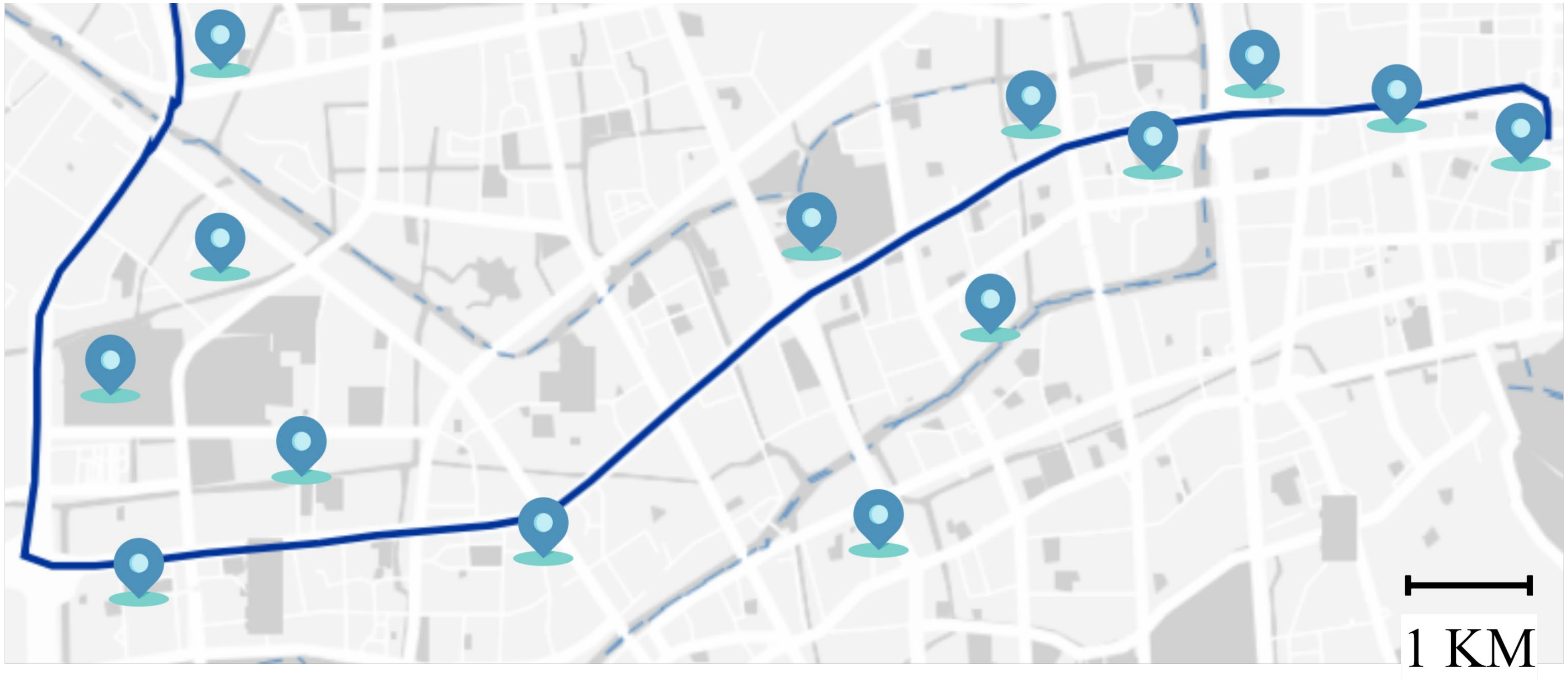
Two Basic Concepts

Cellular Trajectory Map-Matching



Two Basic Concepts

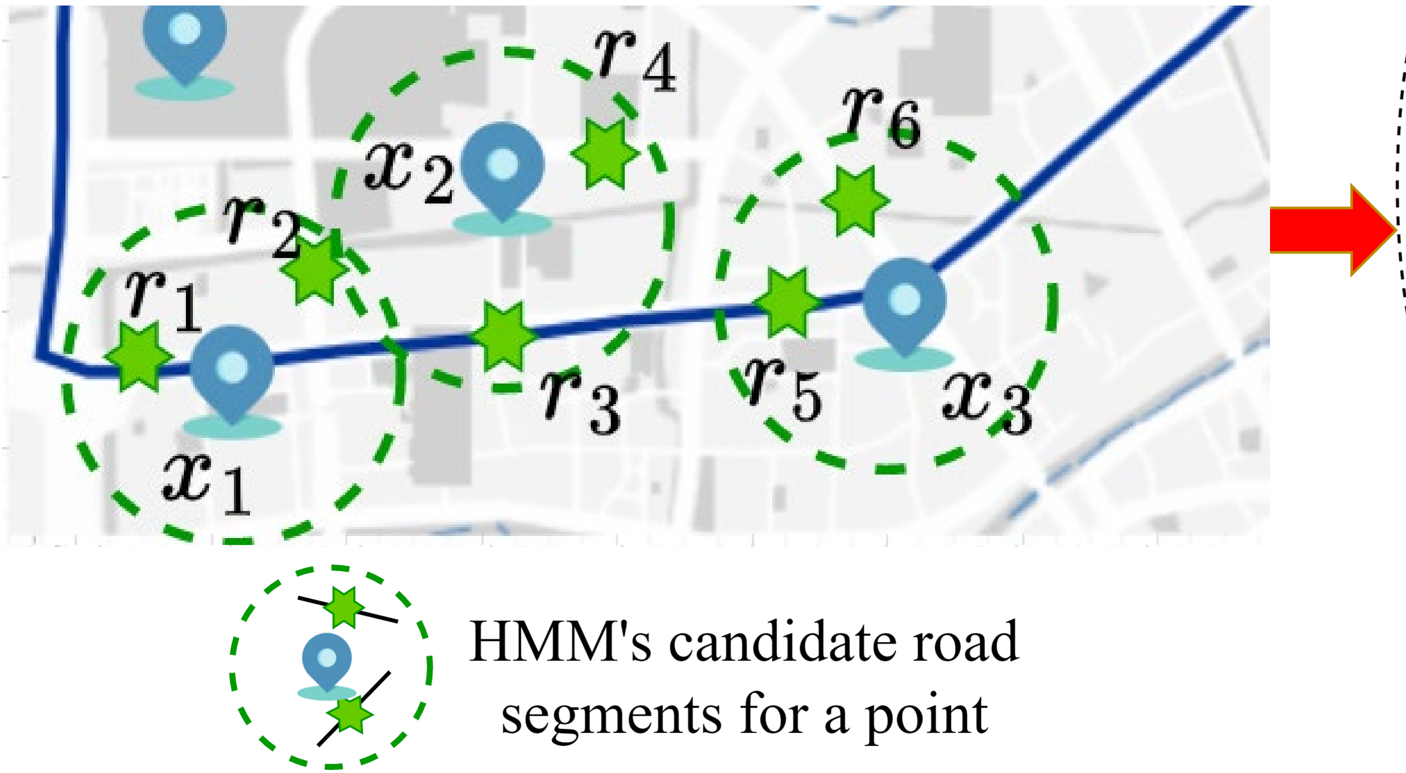
Cellular Trajectory Map-Matching

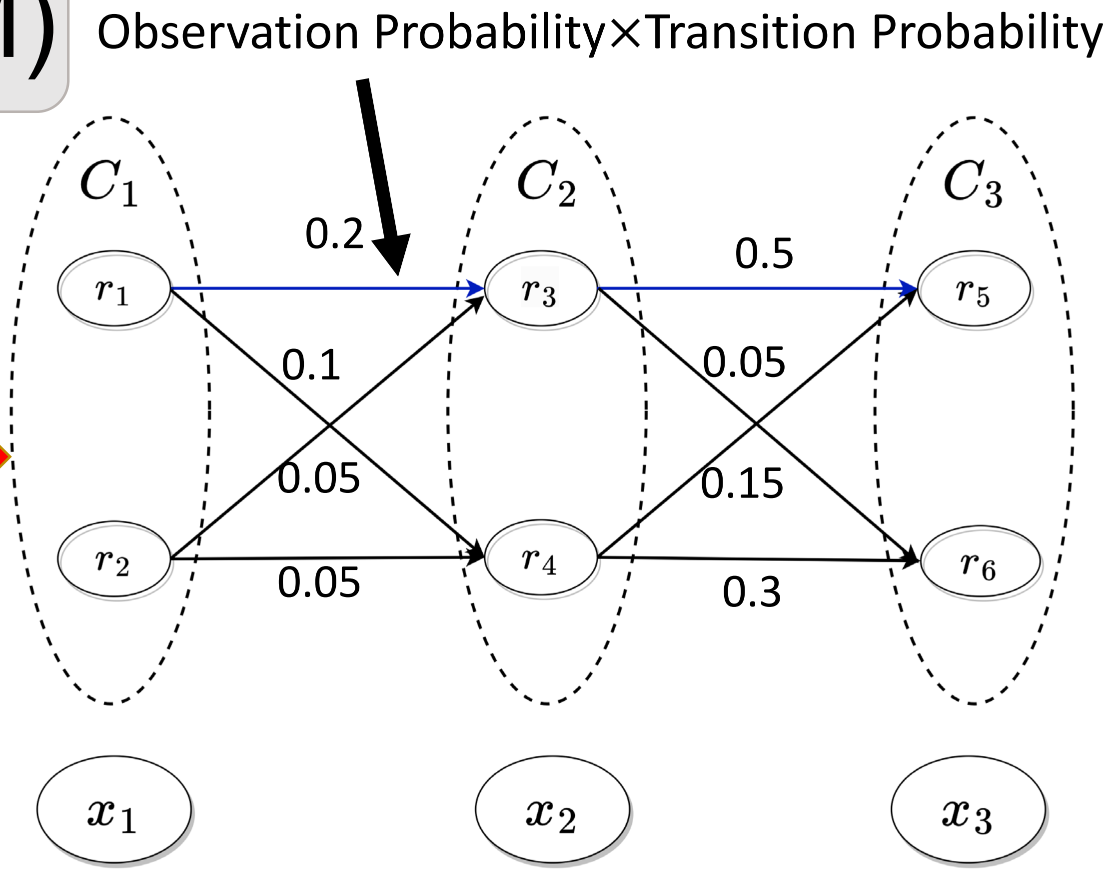


GPS trajectory point   cellular trajectory point   traveled path

Existing Methods

Hidden Markov Model (HMM)



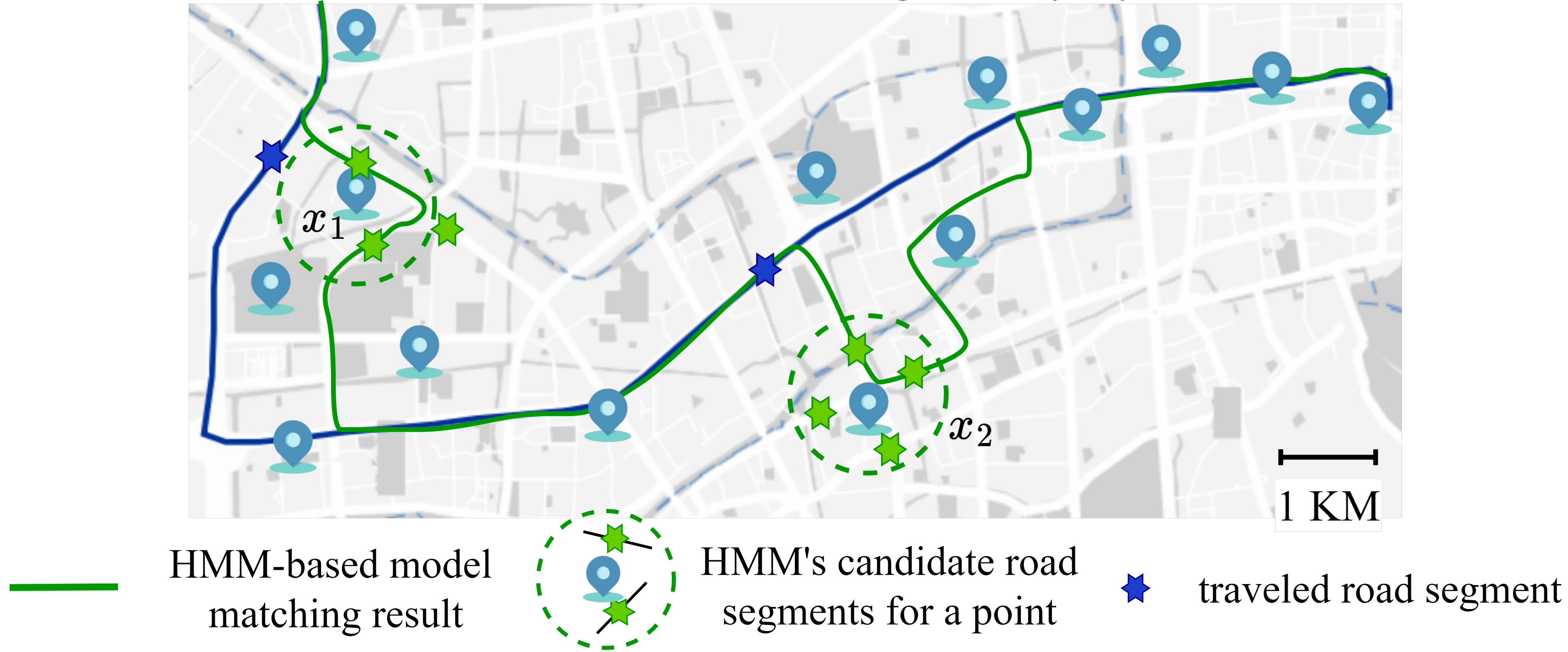


Lattice Graph

Existing Methods

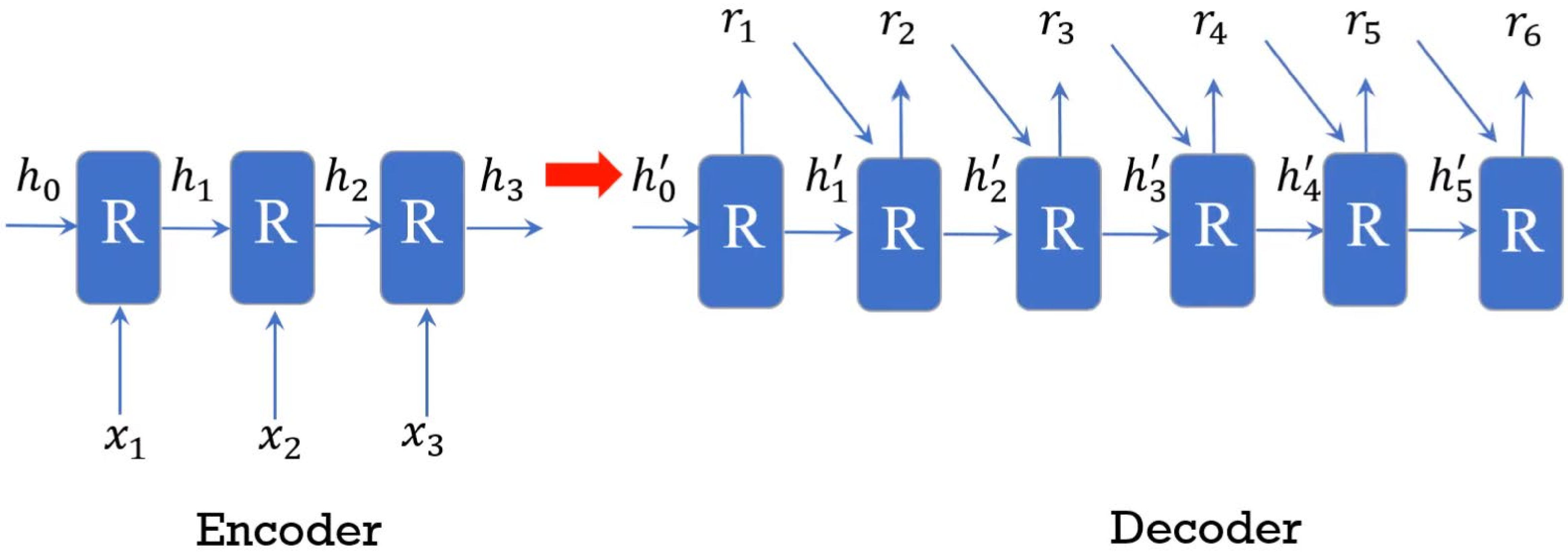
HMM's Drawback

The observation probability P_O
and transition probability P_T
are guided by explicit features (like distance)



Existing Methods

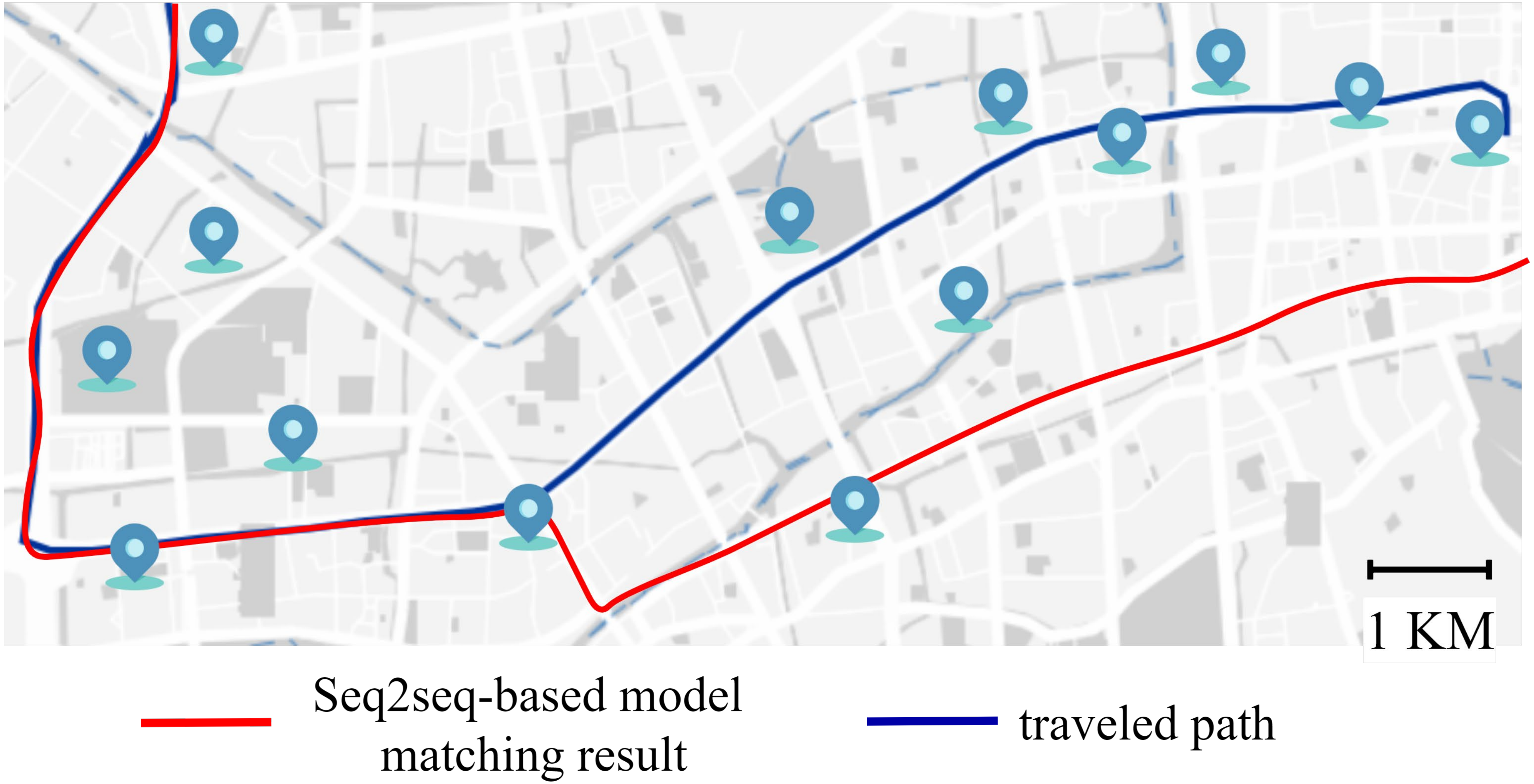
Sequence to Sequence Model (Seq2Seq)



Existing Methods

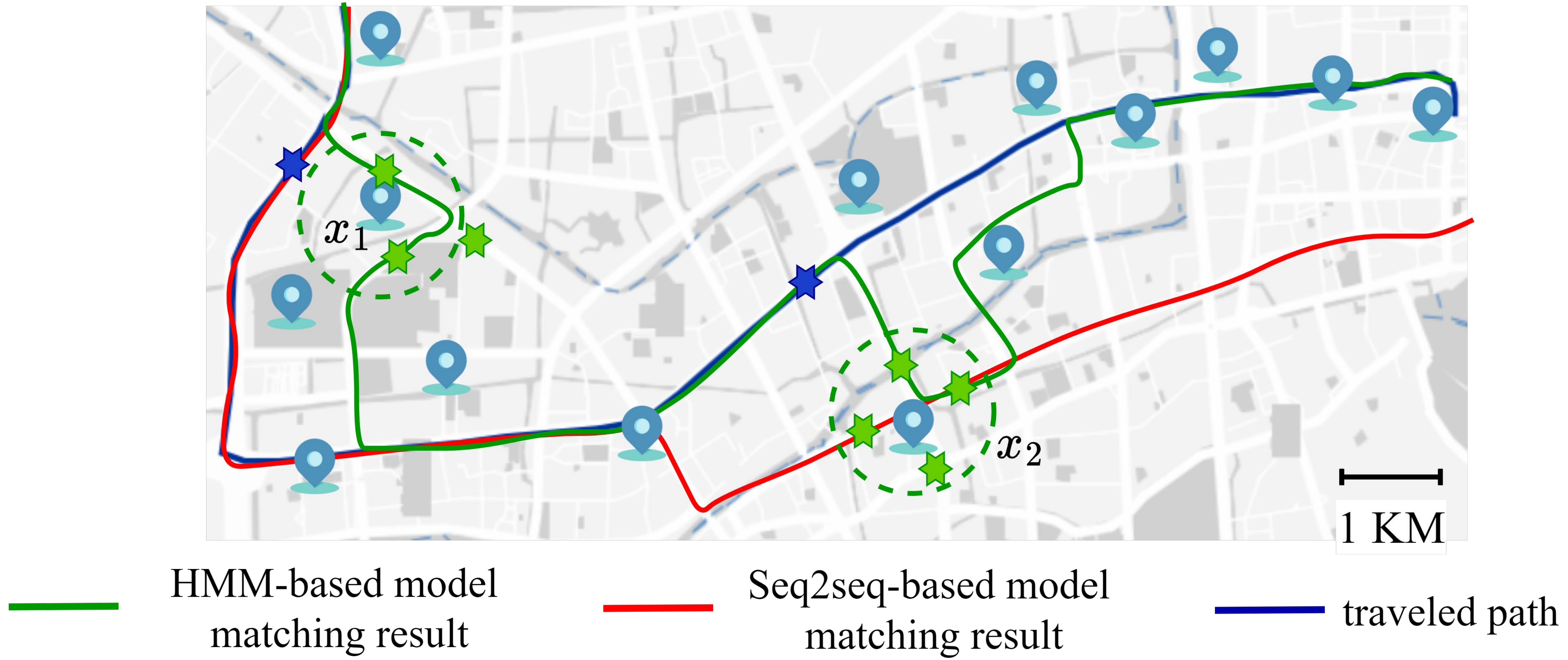
Seq2Seq's Drawback

Facing severe error propagation



Existing Methods

Advantages of HMM and Seq2Seq

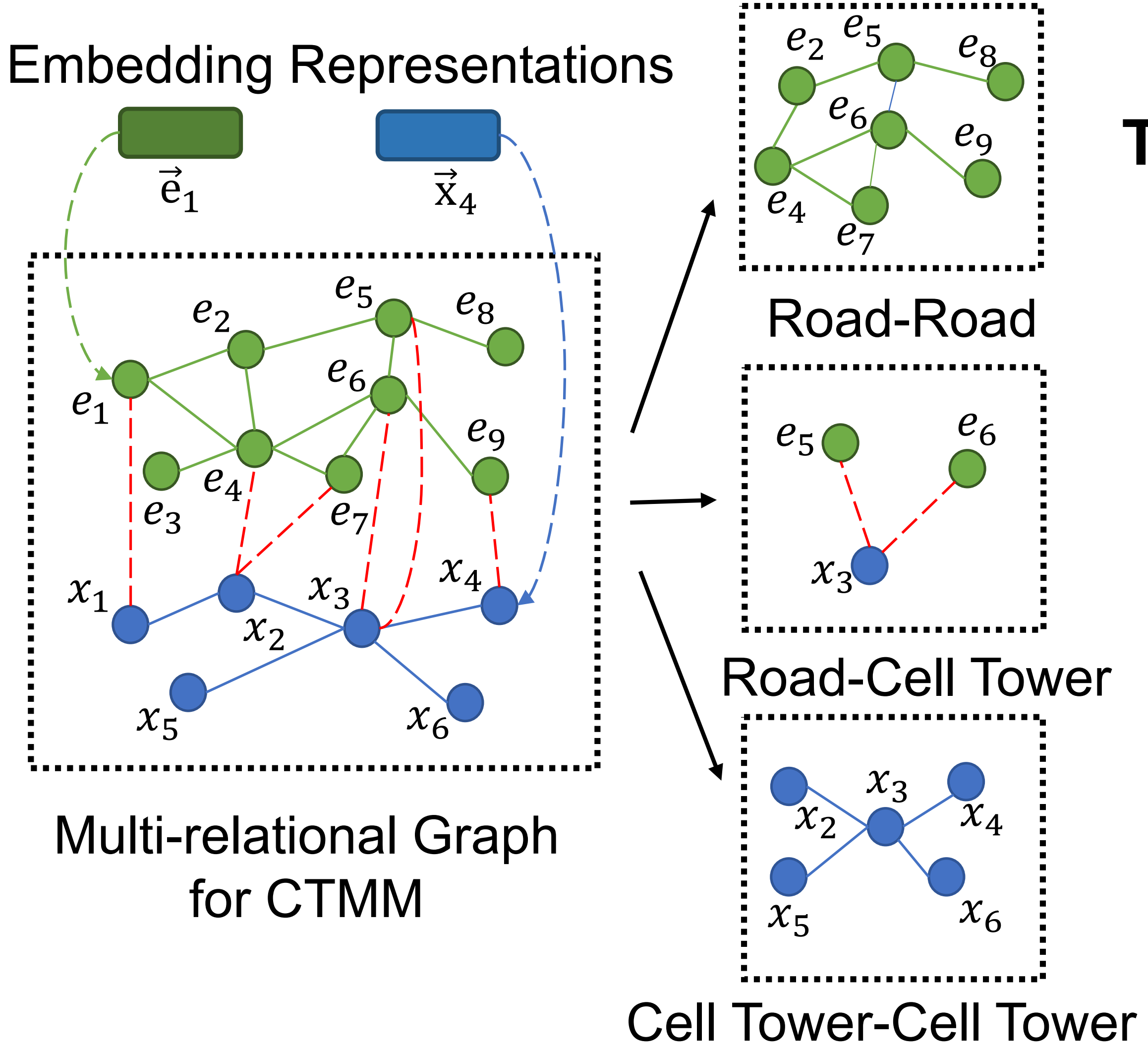


Learning Enhanced Hidden Markov Model

Aim to using **Neural Network** instead of hand-designed heuristic to learn **Observation Probability** P_O **and Transition Probability** P_T

Challenge 1 – Multi-relation Representation Learning

How to effectively capture these relations and embedded them into a shared space?

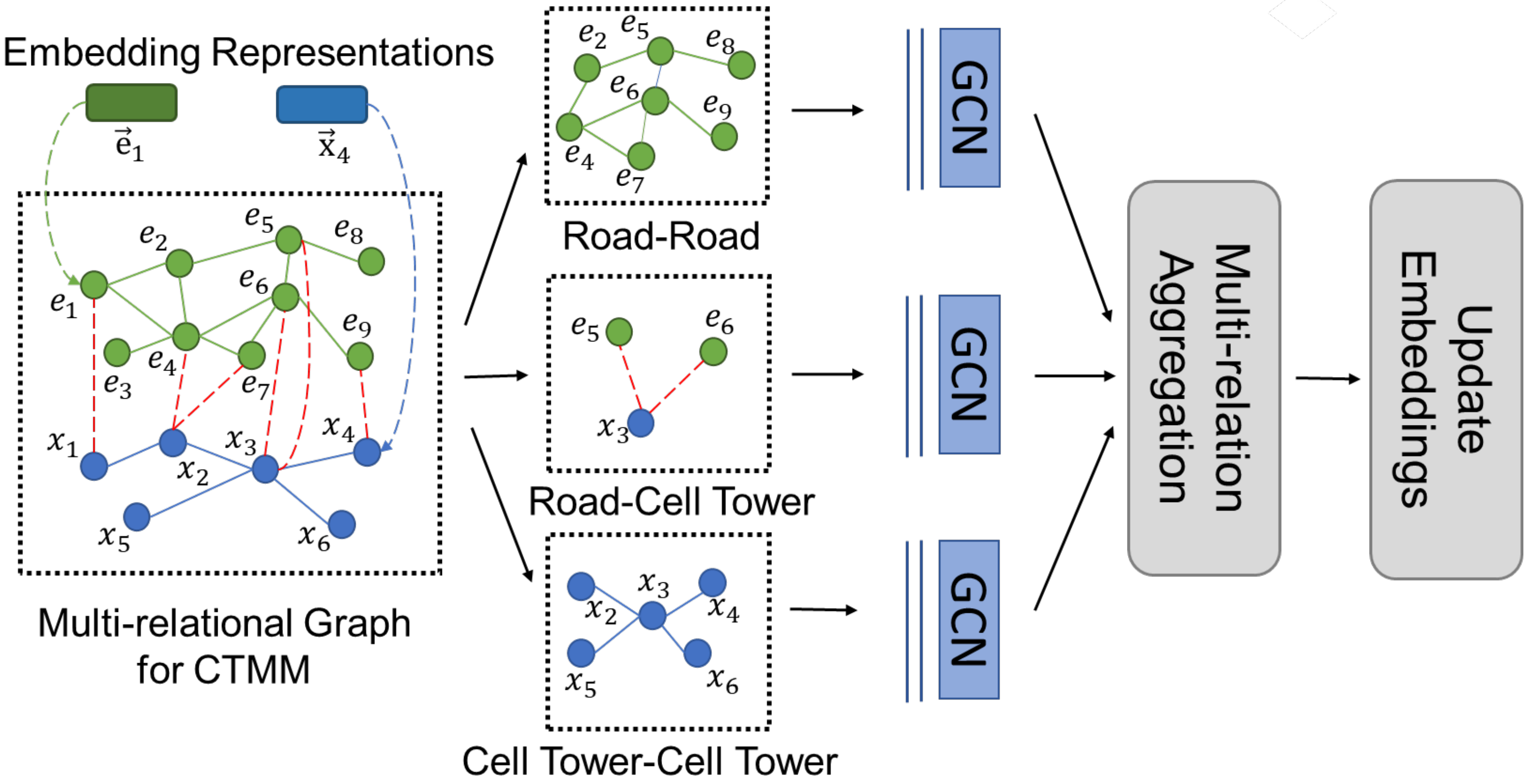


Topological structure

Co-occurrence

Sequentiality

Solution 1 – Multi-relation Representation Learning

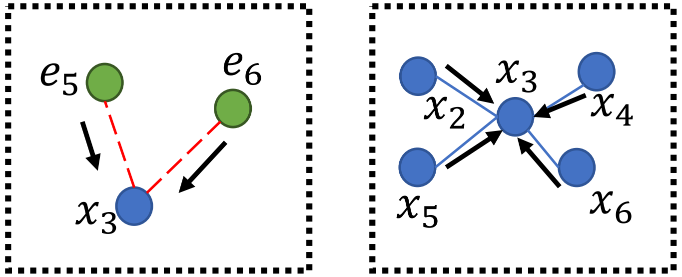


Step1: Initialize

x_3 [0,0,1,0,0,0,…]

[0.052,0.712,0.52,…]

Step2: Message passing



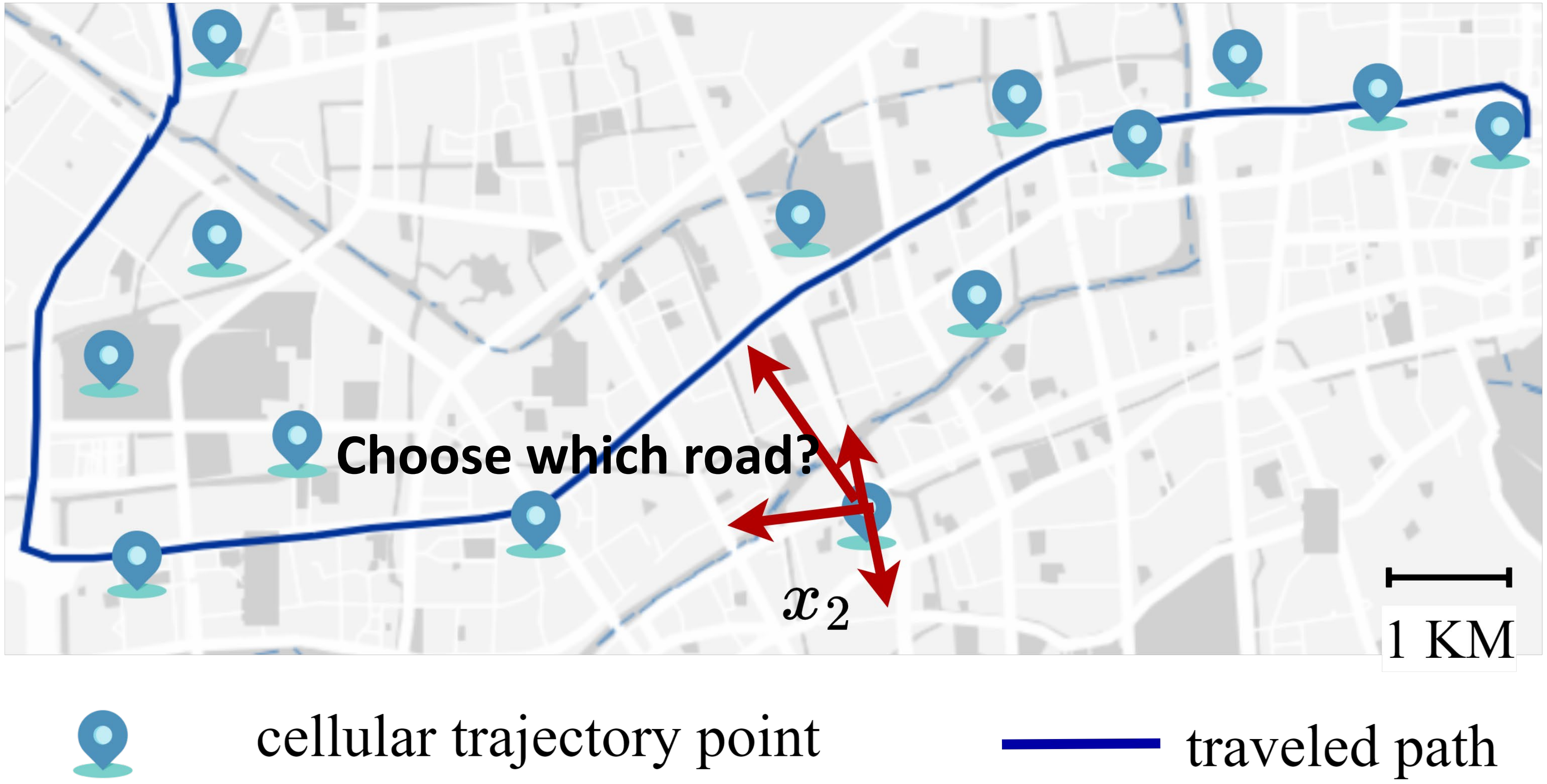
$$z_i^{rel} = \frac{1}{|\mathcal{N}_i^{rel}|} \sum_{v_j \in \mathcal{N}_i^{rel}} W_{rel}^{(l)} h_j^{(l)}$$

Step3: jointly aggregate the message
And update the

$$h_i^{(l+1)} = \sigma\left(\sum_{rel \in \mathcal{R}} W_{agg} z_i^{rel} + W_0^{(l)} h_i^{(l)} \right)$$

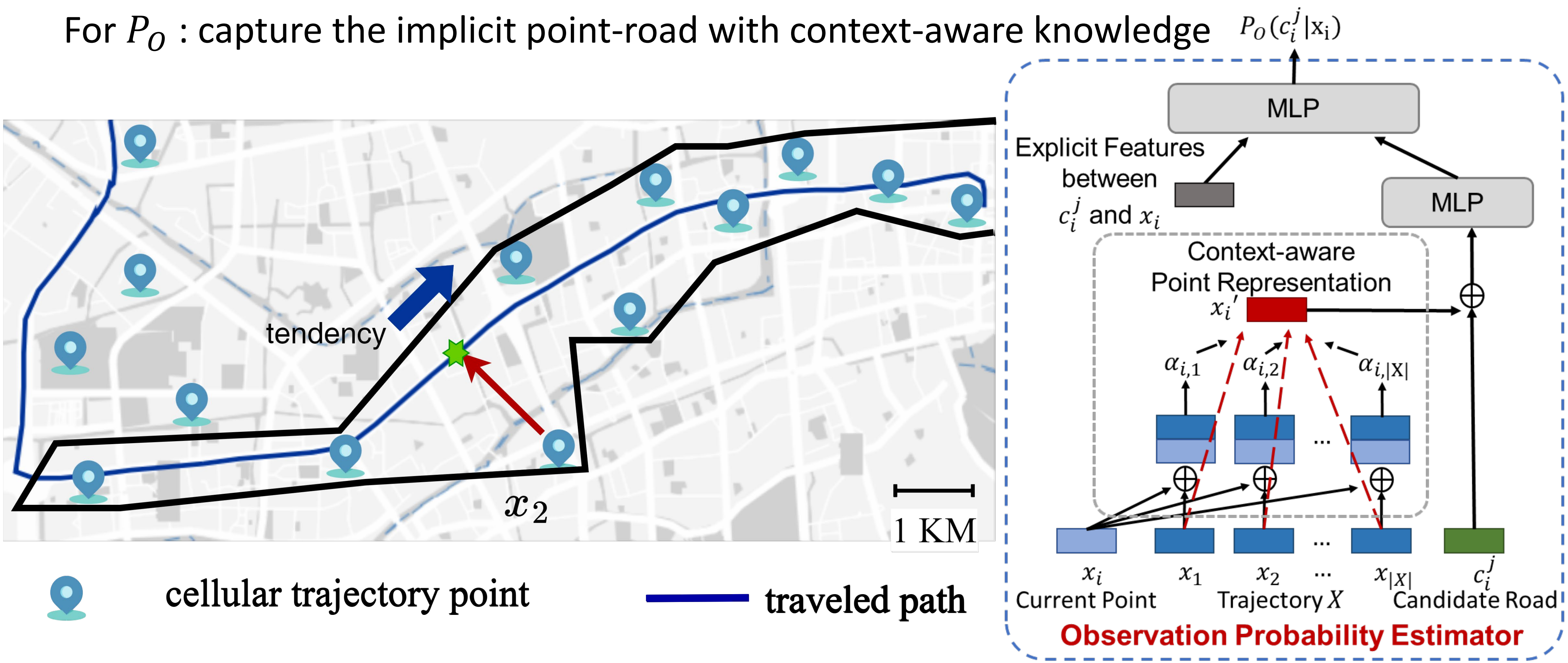
Challenge 2 – Learning for Observation Probability

For P_O : there are too roads with high relevant with the point x_2



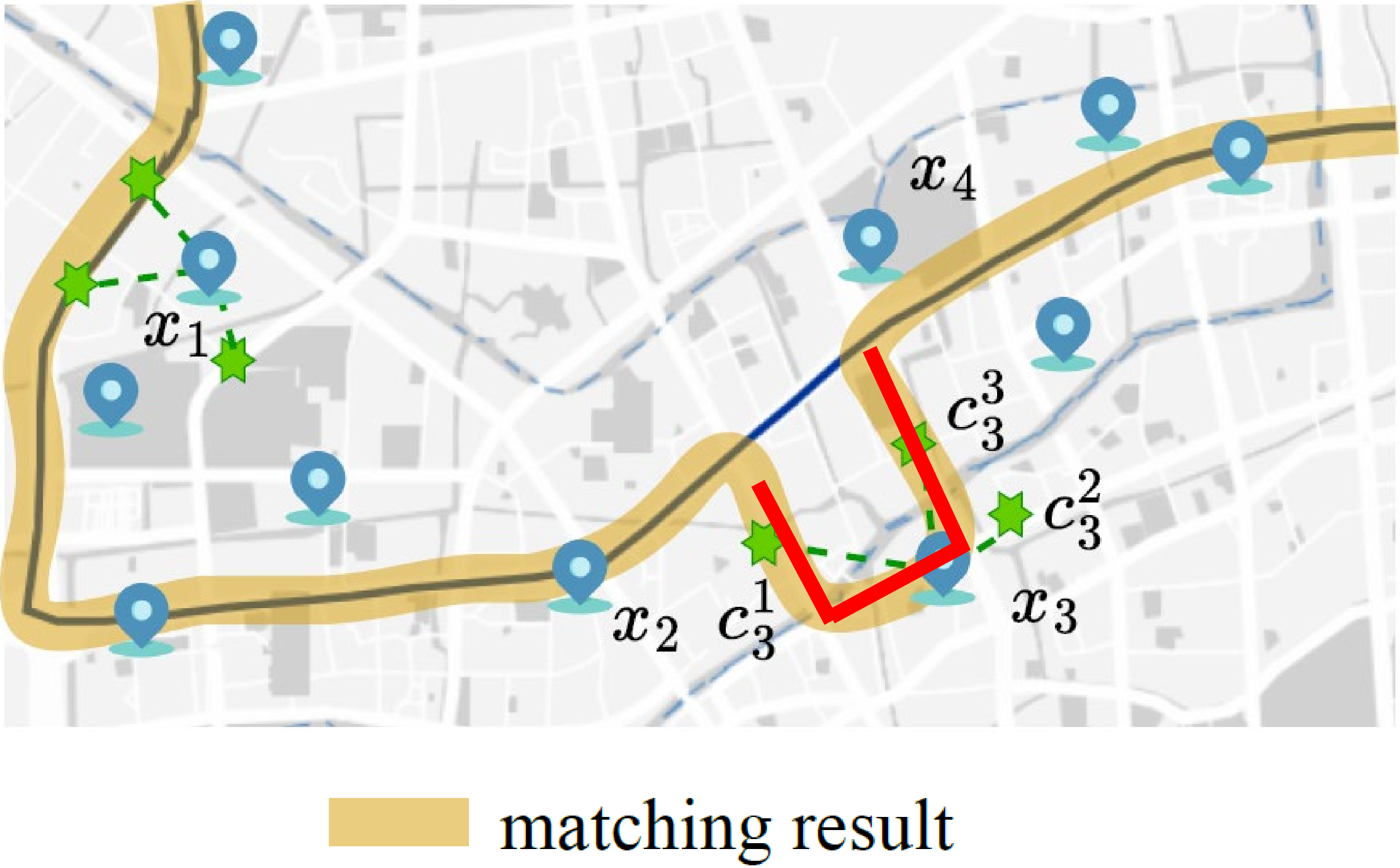
Solution 2 – Learning for Observation Probability

For P_O : capture the implicit point-road with context-aware knowledge



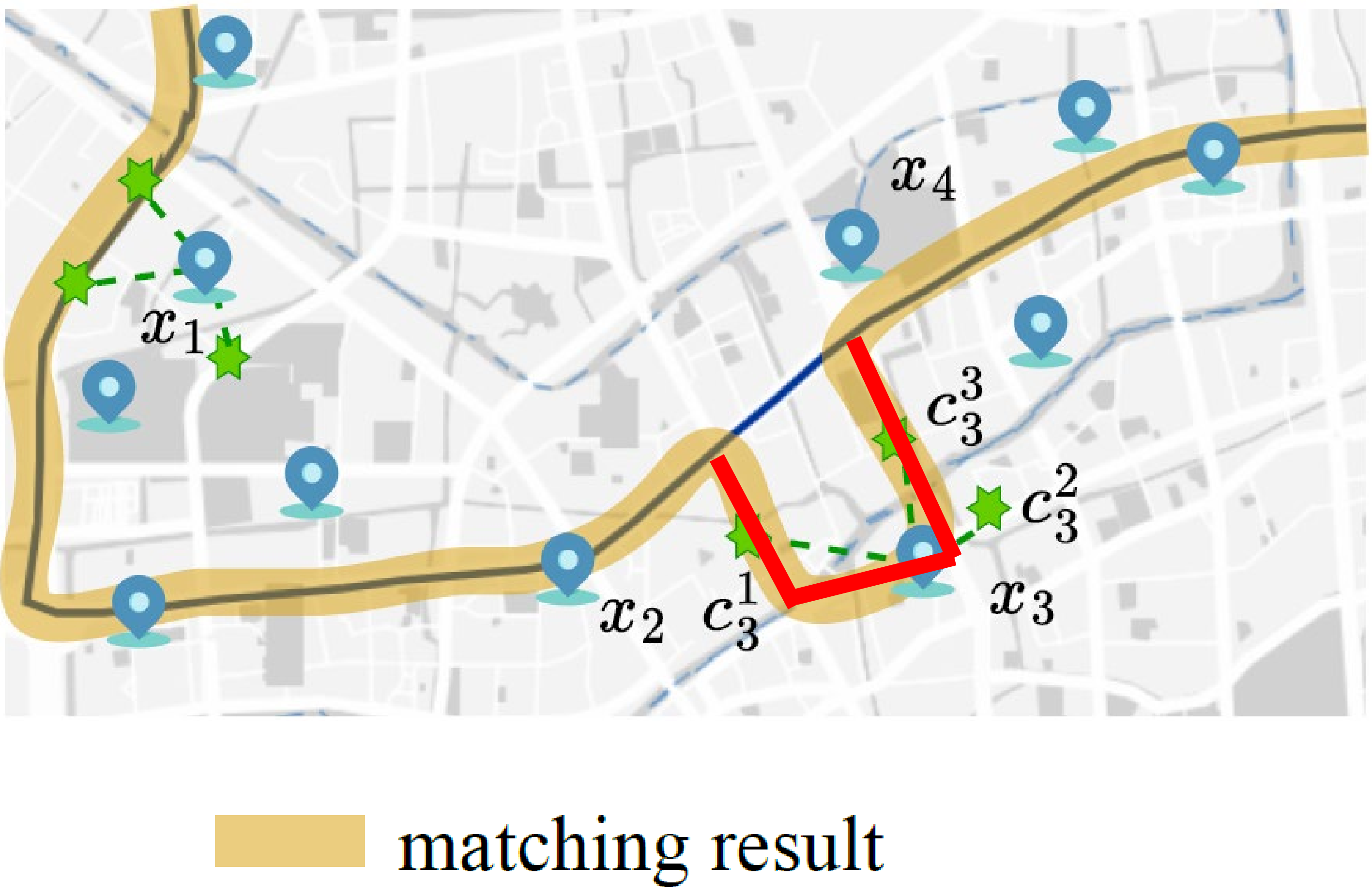
Challenge 3 – Learning for Transition Probability

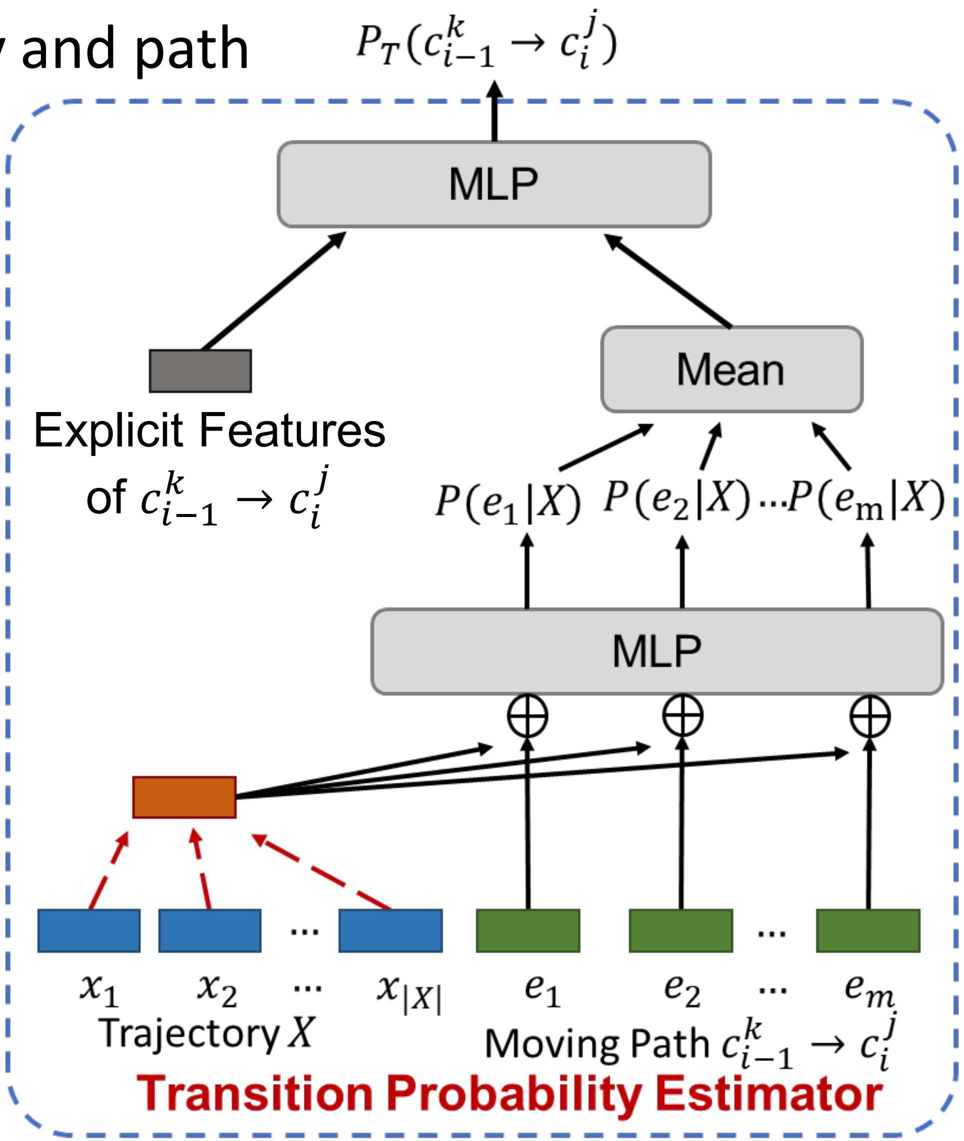
For P_T: how to test the detour of the moving path?



Solution 3 – Learning for Transition Probability

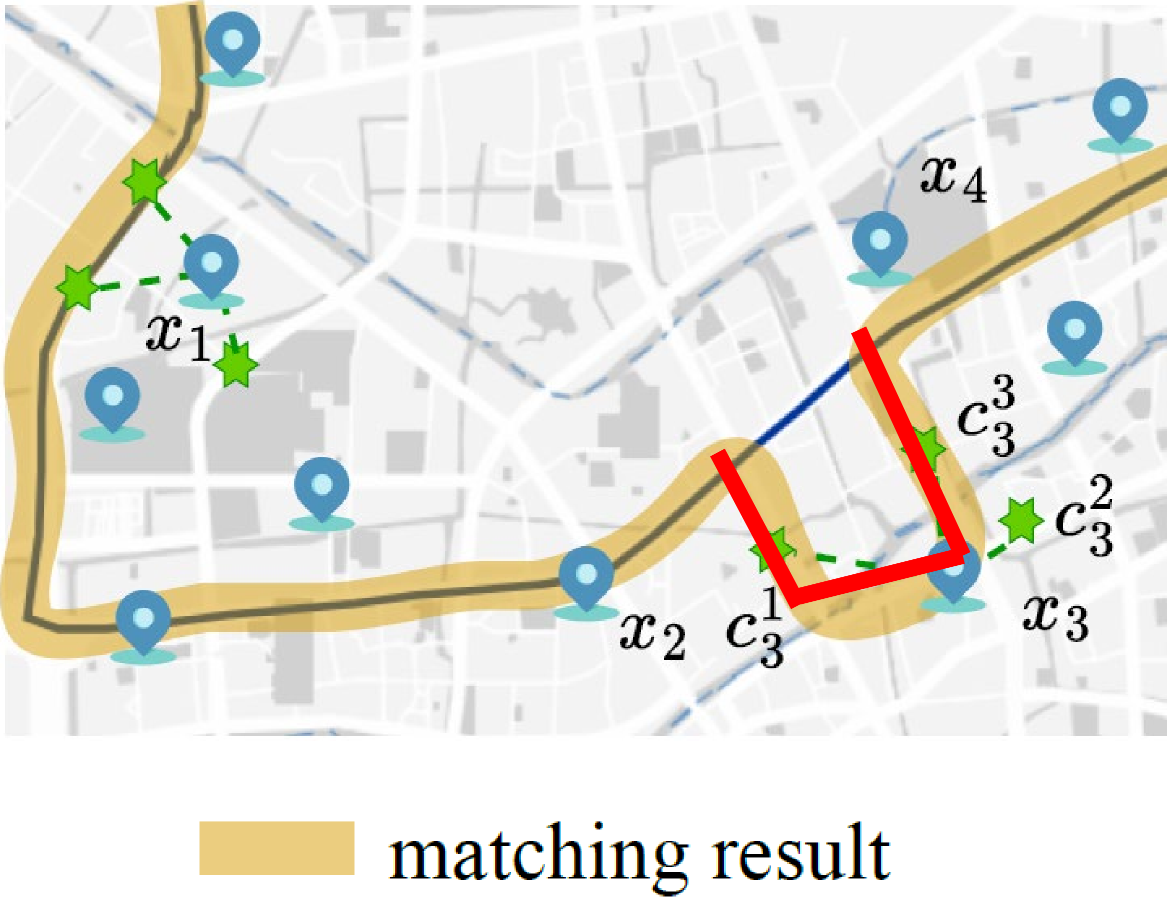
For P_T: capture the implicit features between trajectory and path

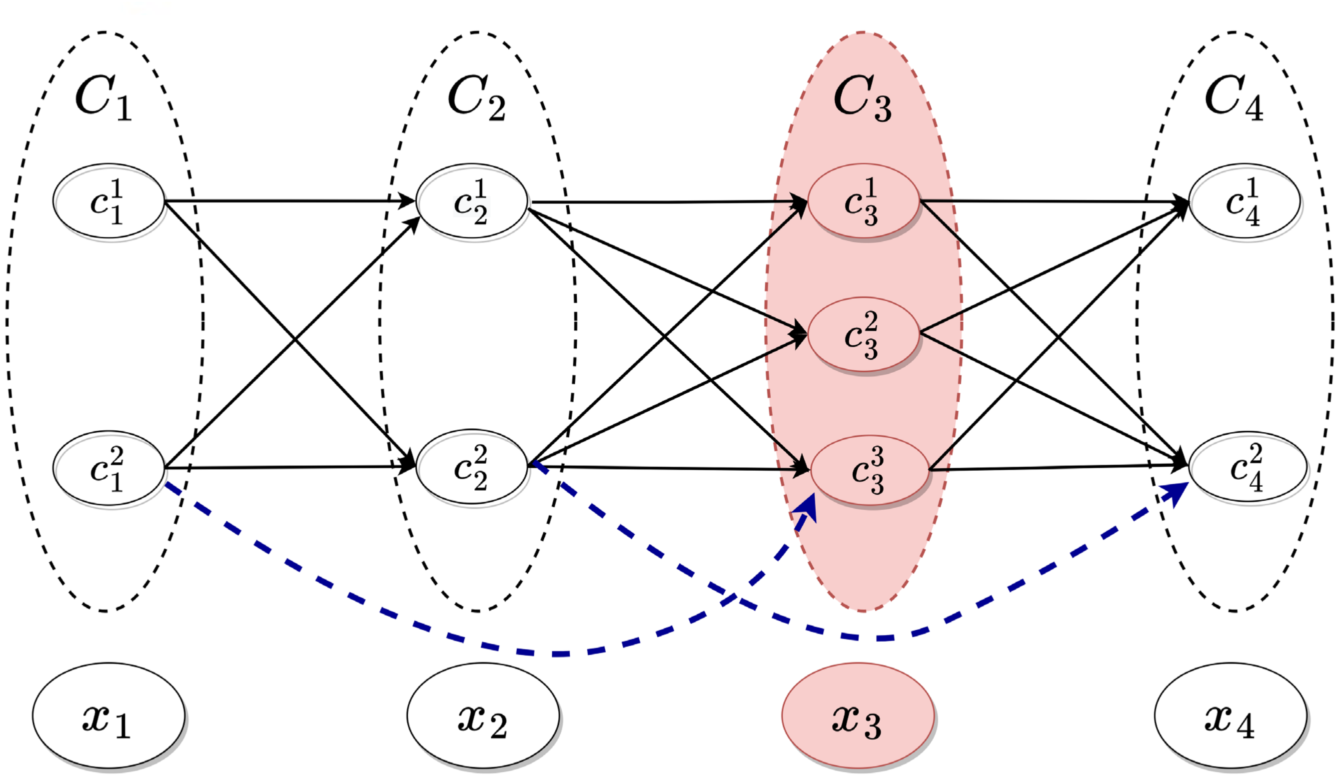




Challenge 4 – Improved Lattice Graph

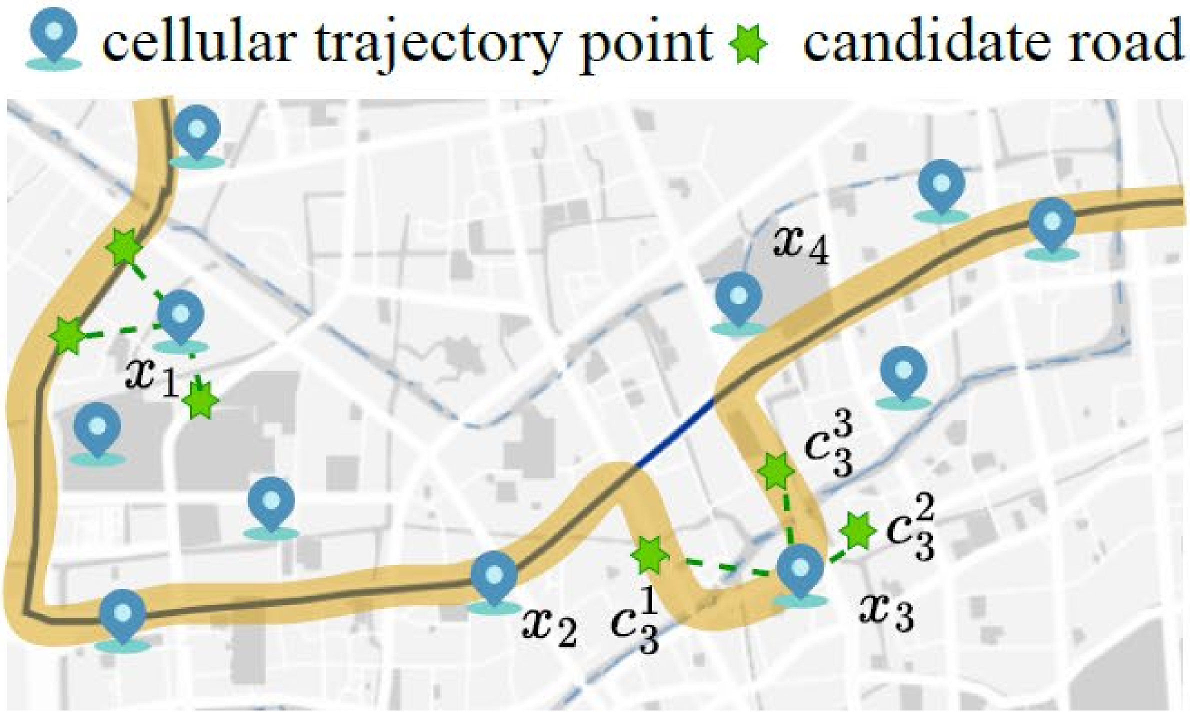
how to remedy the detour of the moving path?

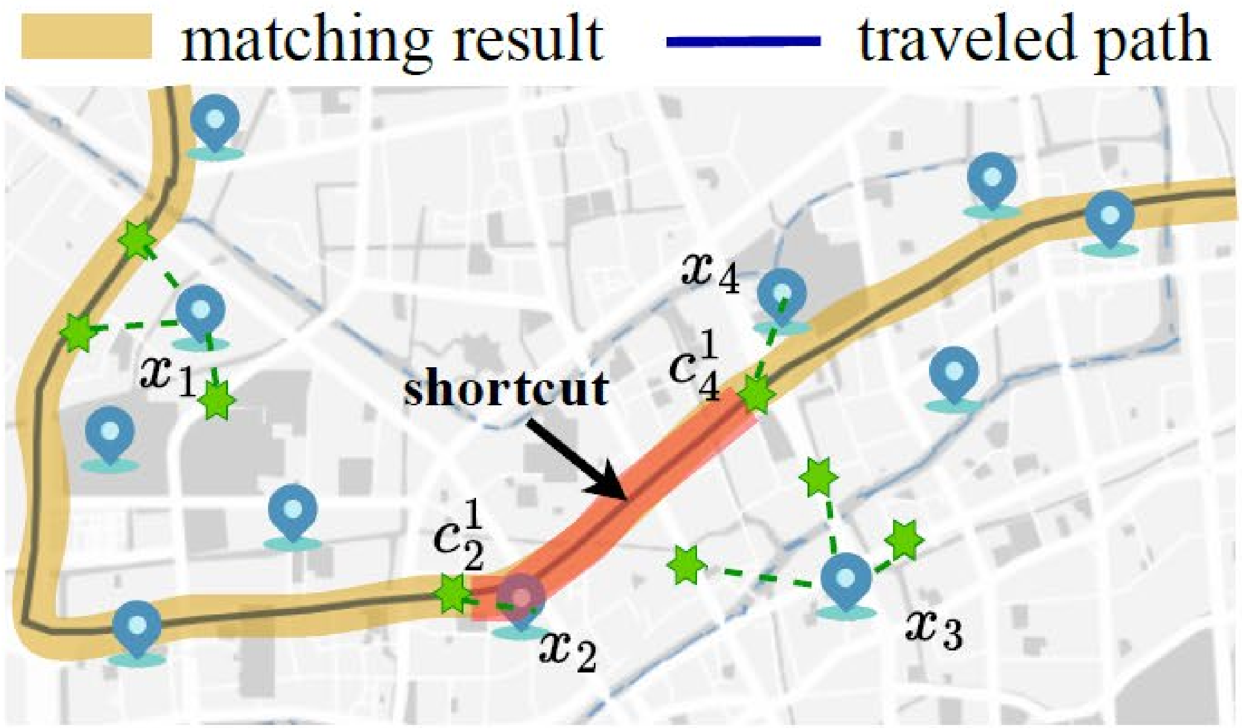




Solution 4 – Improved Lattice Graph

using shortcuts to provide chances to skip the unqualified candidate set





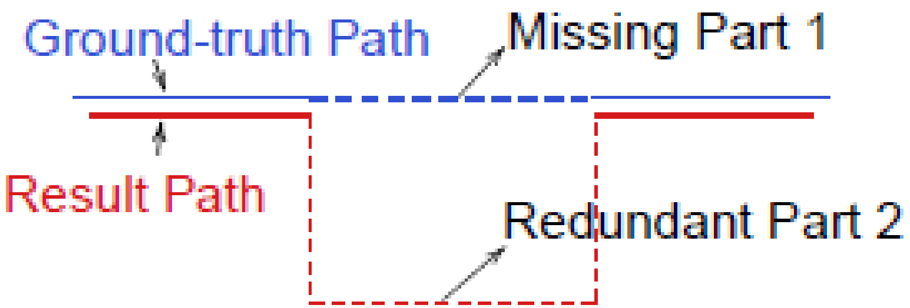
Experiment - Dataset

category	Hangzhou	Xiamen
road segments	92,913	64,828
intersections	67,330	37,591
all cellular trajectory points	3.61 million	1.18 million
all GPS trajectory points	9.73 million	4.98 million
cellular trajectory points per trajectory	34	40
GPS trajectory points per trajectory	81	88
average cellular sampling interval (s)	67	42
maximum cellular sampling interval (s)	247	185
average cellular sampling distance (m)	730	650
median cellular sampling distance (m)	493	455

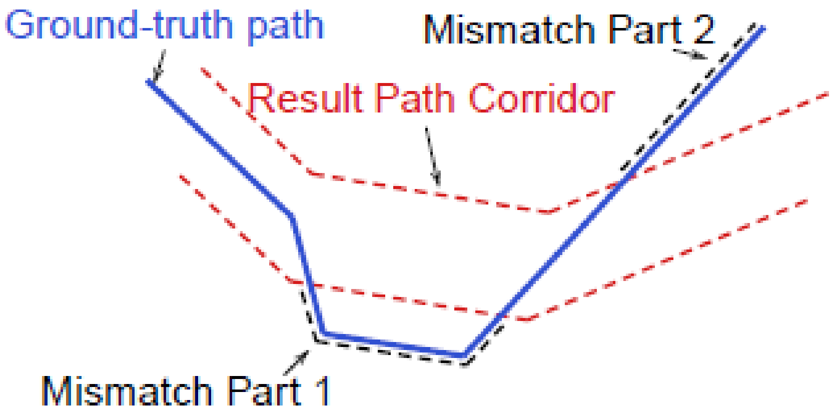
DATASET CHARACTERISTIC

Experiment - Evaluation criteria

- Precision and Recall
- Route Mismatch Fraction (RMF)
- Corridor Mismatch Fraction (CMF)



RMF



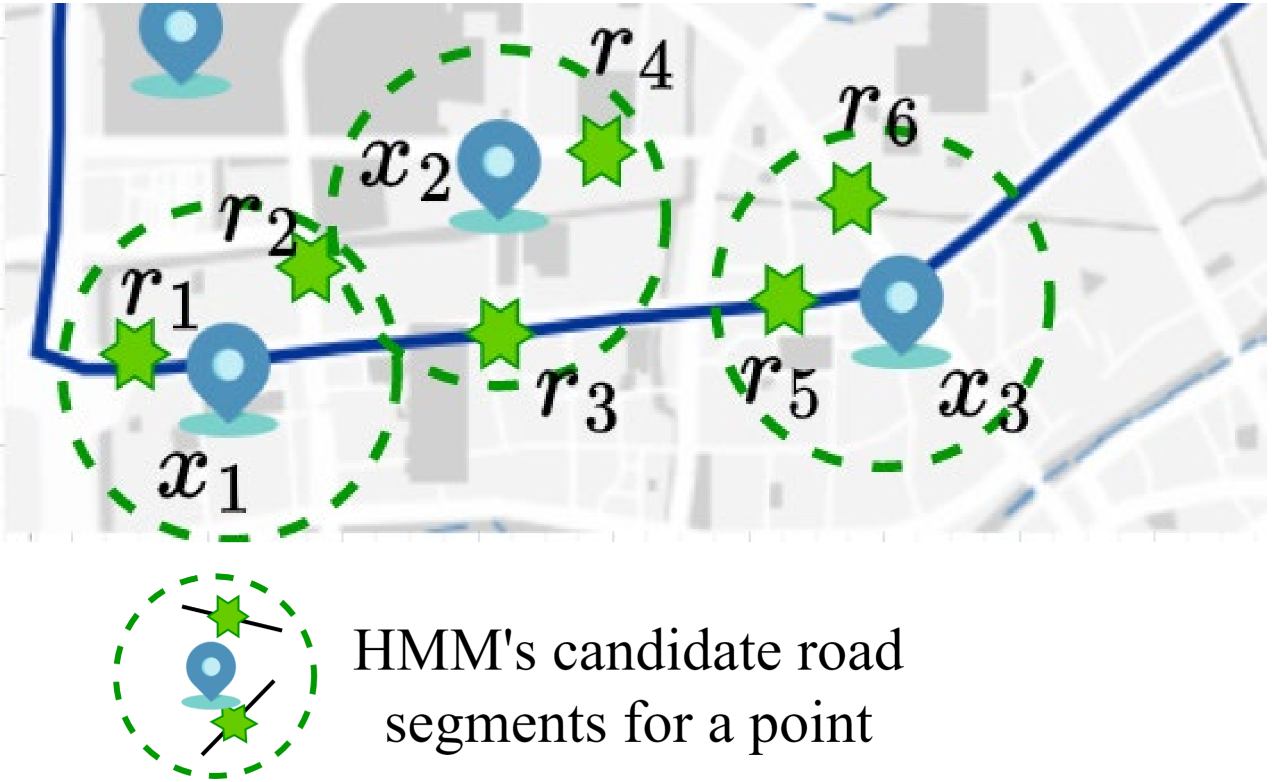
CMF

$$\mathrm{RMF} = \frac{\sum \text{length of mismatched road segments}}{\text{length of the ground-truth path}}$$

$$\mathrm{CMF} = \frac{\sum \text{corridor uncovered length}}{\text{length of the ground-truth path}}$$

Experiment - Evaluation criteria

- Precision and Recall
- Route Mismatch Fraction (RMF)
- Corridor Mismatch Fraction (CMF)
- Hitting Ratio (HR)
- Average inference Time (Avg Time)



Experiment – Overall Performance

HMM based methods

Dataset / Metric	Hangzhou Precision	Hangzhou Recall	Hangzhou RMF	Hangzhou CMF50	Hangzhou Avg Time (s)	Xiamen Precision	Xiamen Recall	Xiamen RMF	Xiamen CMF50	Xiamen Avg Time (s)
Methods designed for GPS trajectory map-matching										
STM [8]	0.388	0.476	1.237	0.225	0.040	0.411	0.498	1.050	0.198	0.044
IVMM [10]	0.409	0.518	1.125	0.188	0.101	0.428	0.529	0.936	0.172	0.136
IFM [32]	0.430	0.522	1.024	0.178	0.045	0.451	0.537	0.889	0.167	0.048
DeepMM [37]	0.446	0.544	0.881	0.172	0.951	0.478	0.568	0.785	0.158	1.284
MCM [34]	0.449	0.552	0.893	0.169	0.033	0.479	0.572	0.780	0.152	0.039
TransformerMM [38]	0.455	0.552	0.838	0.170	1.667	0.483	0.577	0.769	0.153	1.857
Methods designed for CTMM										
CLSTERS [41]	0.443	0.551	0.922	0.173	0.043	0.470	0.563	0.805	0.154	0.048
SNet [12]	0.446	0.555	0.891	0.169	0.034	0.475	0.565	0.792	0.153	0.041
THMM [42]	0.461	0.562	0.815	0.165	0.041	0.486	0.583	0.767	0.148	0.045
DMM [15]	0.467	0.566	0.784	0.163	0.853	0.489	0.594	0.755	0.145	0.916
Our method										
LHMM	**0.516**	**0.613**	**0.670**	**0.126**	**0.032**	**0.547**	**0.667**	**0.641**	**0.124**	**0.037**
Improved	**10.49%**	**8.30%**	**14.54%**	**22.69%**	**3.03%**	**11.86%**	**12.28%**	**8.79%**	**15.09%**	**5.12%**

Seq2Seq based methods

Experiment – Ablation Results

Dataset	Variant	Precision	CMF50	HR
Hangzhou	**LHMM**	**0.516**	**0.126**	**0.953**
	LHMM-E	0.457	0.142	0.931
	LHMM-H	0.489	0.136	0.942
	LHMM-O	0.428	0.178	0.920
	LHMM-T	0.472	0.155	0.926
	LHMM-S	0.484	0.140	0.937
	STM	0.388	0.225	0.874
	STM+S	0.405	0.189	0.911
Xiamen	**LHMM**	**0.545**	**0.125**	**0.965**
	LHMM-E	0.494	0.144	0.938
	LHMM-H	0.517	0.142	0.942
	LHMM-O	0.462	0.158	0.931
	LHMM-T	0.524	0.135	0.952
	LHMM-S	0.516	0.139	0.944
	STM	0.411	0.198	0.882
	STM+S	0.432	0.170	0.915

Experiment – Ablation Results

Dataset	Variant	Precision	CMF50	HR
Hangzhou	**LHMM**	**0.516**	**0.126**	**0.953**
	LHMM-E	0.457	0.142	0.931
	LHMM-H	0.489	0.136	0.942
	LHMM-O	0.428	0.178	0.920
	LHMM-T	0.472	0.155	0.926
	LHMM-S	0.484	0.140	0.937
	STM	0.388	0.225	0.874
	STM+S	0.405	0.189	0.911
Xiamen	**LHMM**	**0.545**	**0.125**	**0.965**
	LHMM-E	0.494	0.144	0.938
	LHMM-H	0.517	0.142	0.942
	LHMM-O	0.462	0.158	0.931
	LHMM-T	0.524	0.135	0.952
	LHMM-S	0.516	0.139	0.944
	STM	0.411	0.198	0.882
	STM+S	0.432	0.170	0.915

Experiment – Ablation Results

Dataset	Variant	Precision	CMF50	HR
Hangzhou	**LHMM**	**0.516**	**0.126**	**0.953**
	LHMM-E	0.457	0.142	0.931
	LHMM-H	0.489	0.136	0.942
	LHMM-O	0.428	0.178	0.920
	LHMM-T	0.472	0.155	0.926
	LHMM-S	0.484	0.140	0.937
	STM	0.388	0.225	0.874
	STM+S	0.405	0.189	0.911
Xiamen	**LHMM**	**0.545**	**0.125**	**0.965**
	LHMM-E	0.494	0.144	0.938
	LHMM-H	0.517	0.142	0.942
	LHMM-O	0.462	0.158	0.931
	LHMM-T	0.524	0.135	0.952
	LHMM-S	0.516	0.139	0.944
	STM	0.411	0.198	0.882
	STM+S	0.432	0.170	0.915

Experiment – Ablation Results

Dataset	Variant	Precision	CMF50	HR
Hangzhou	**LHMM**	**0.516**	**0.126**	**0.953**
	LHMM-E	0.457	0.142	0.931
	LHMM-H	0.489	0.136	0.942
	LHMM-O	0.428	0.178	0.920
	LHMM-T	0.472	0.155	0.926
	LHMM-S	0.484	0.140	0.937
	STM	0.388	0.225	0.874
	STM+S	0.405	0.189	0.911
Xiamen	**LHMM**	**0.545**	**0.125**	**0.965**
	LHMM-E	0.494	0.144	0.938
	LHMM-H	0.517	0.142	0.942
	LHMM-O	0.462	0.158	0.931
	LHMM-T	0.524	0.135	0.952
	LHMM-S	0.516	0.139	0.944
	STM	0.411	0.198	0.882
	STM+S	0.432	0.170	0.915

Summary

- we have developed a learning-enhanced HMM map-matching approach for cellular trajectories.
- A representation learning component is designed to fully capture multi-relational information tailored for the CTMM task.
- A learned observation probability captures the implicit context-aware correlation between roads and points for better positioning denoising, and a learned transition probability models the hidden relevance between moving paths and trajectories.
- These two probabilities then guide the path-finding process on an improved candidate graph.