Explainable Trajectory Representation through Dictionary Learning

Yuanbo Tang, Zhiyuan Peng and Yang Li*
Tsinghua-Berkeley Shenzhen Institute, Tsinghua University

9

Introduction

Background.

- Trajectory representation learning provides great opportunities to understand vehicular traffic patterns.
- It can benefit downstream tasks including trajectory compression, trip time estimation, public transportation route planning.

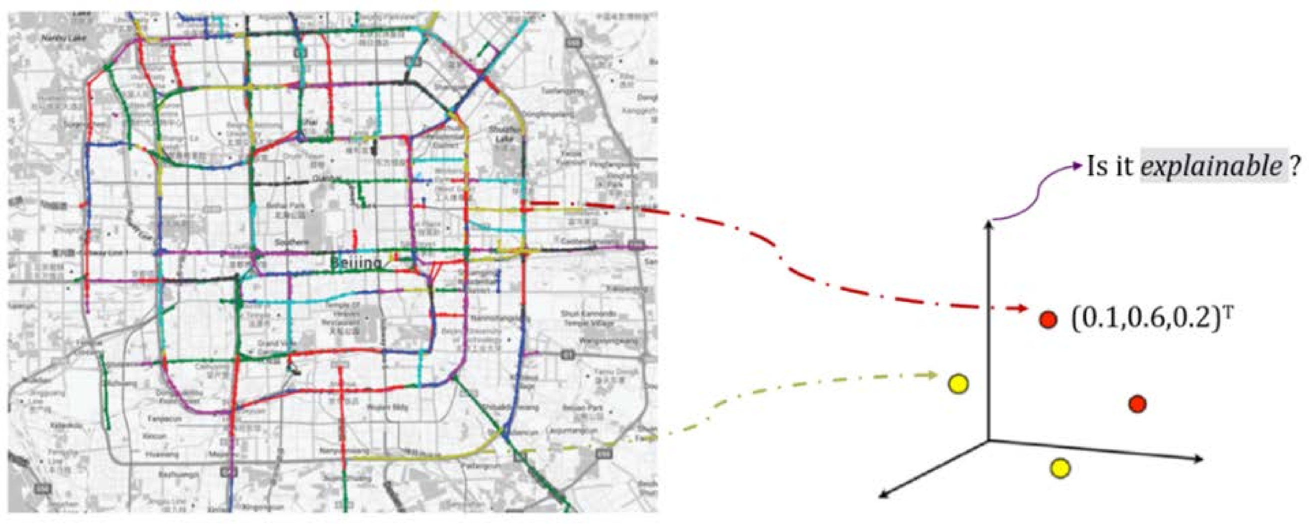


Fig 1. Transforming a trajectory into an embedding vector

Motivation.

- Embeddings generated by deep-learning methods are usually dense vectors whose dimensions lack semantic meaning.
- It is difficult to interpret the learned representation.

Purpose.

- Extract common trajectory segments (called pathlet) as a dictionary.
- Represent trajectory by concatenating pathlets from dictionary.
- Generate semantic trajectory representation vectors, each dimension corresponding to a mobility pattern.

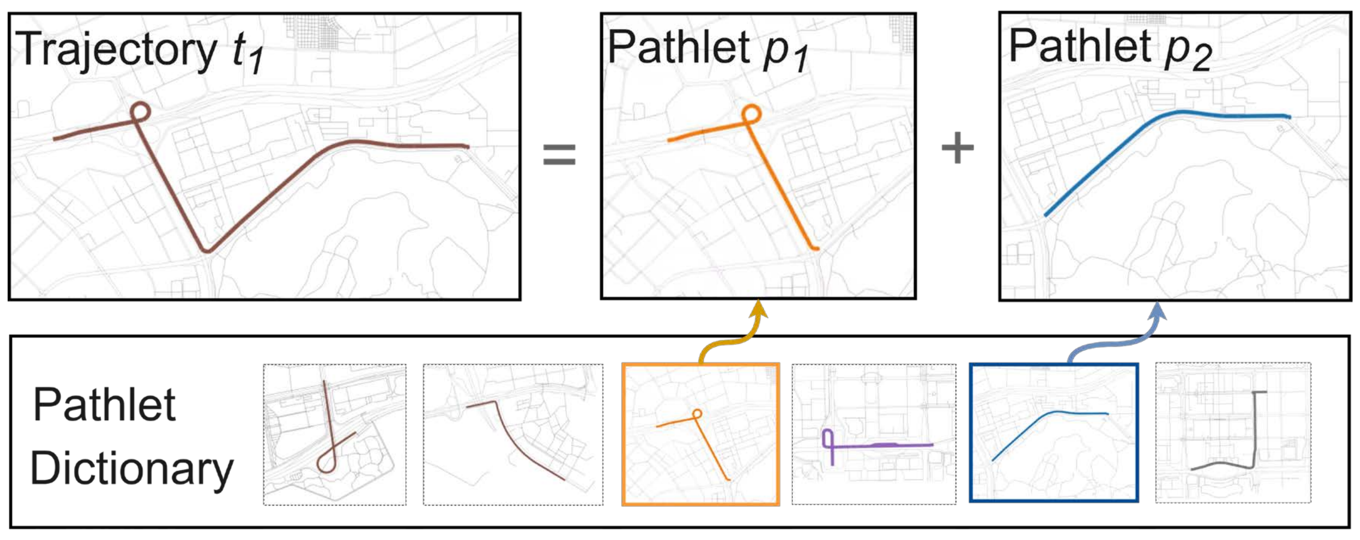
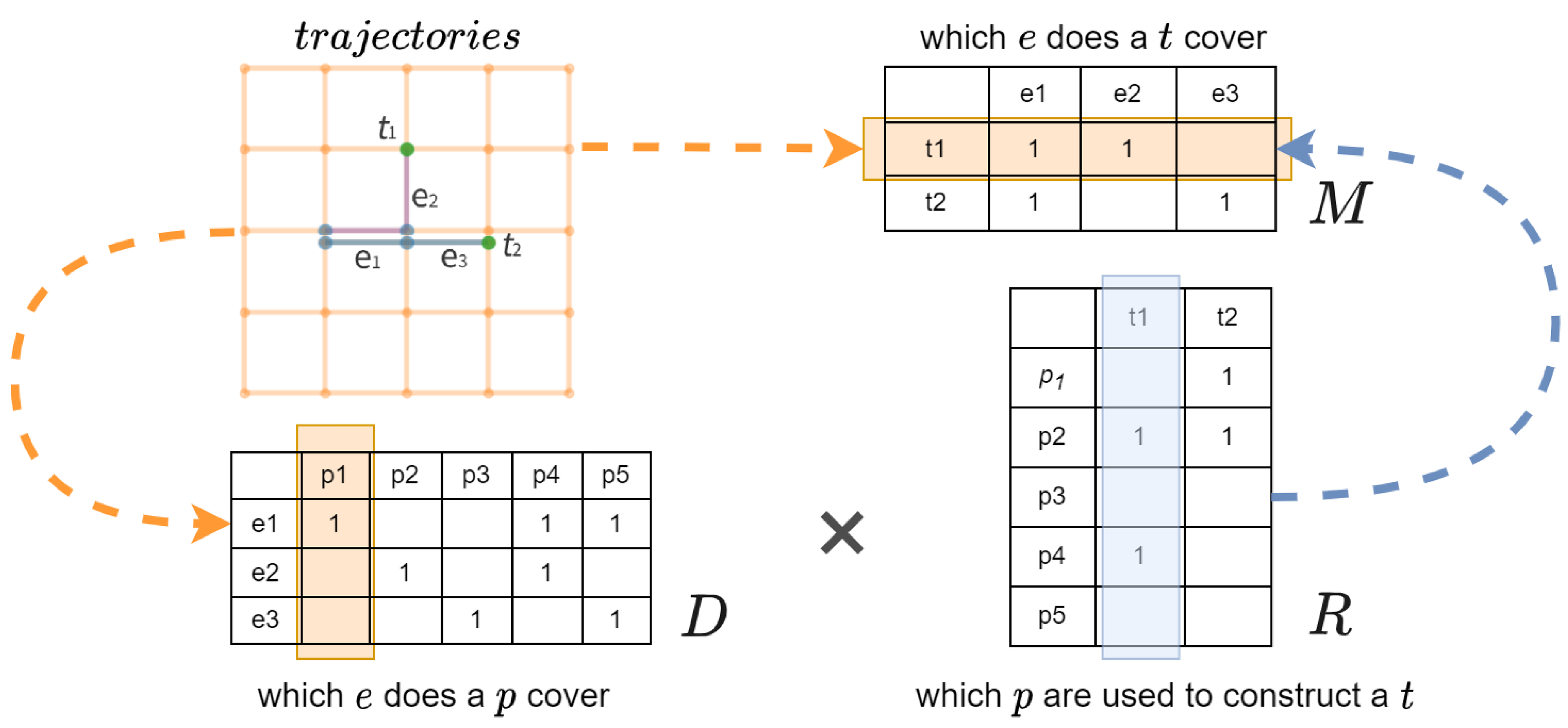


Fig 2. Represent a trajectory by concatenating pathlets from dictionary

Evaluation criteria of a pathlet dictionary.

- This dictionary should be able to reconstruct all trajectories.
- Smaller dictionary is better, which means less redundant information.
- Average number of pathlets used to reconstruct trajectory should be as small as possible.

Problem Formulation



- R is decision matrix, corresponding to dictionary P.
- Matrix D and M record the cover relationship between $T, E, \overline{P}$

Problem formulation.

The size of dictionary

Trade off hyper parameter

$$\min_{R_{i,j}\in\{0,1\}} \sum_i max(R_{i,:}) + \lambda * \sum_i \sum_j |R_{i,j}|$$

$$s.t.\ DR \geq M$$

Constraint: each trajectory should be covered

The number of pathlets used to reconstruct trajectories

Algorithm

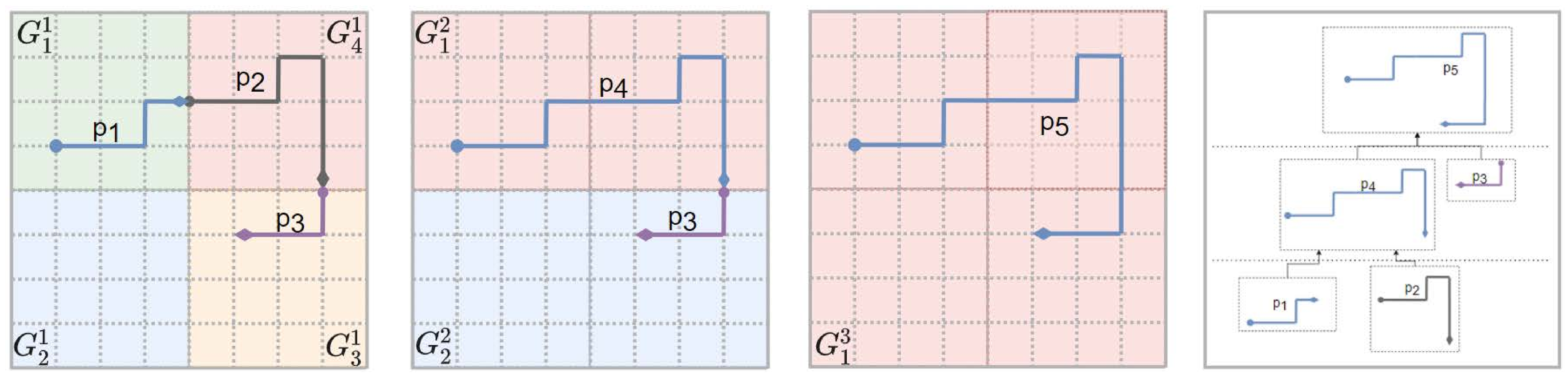
Pathlet sparse dictionary learning with randomized rounding.

1. Relax the binary constraint and find the fractional solution R^*
2. Obtain the rounded solution R^r

Probability bound. Given the size of dataset $|T|$, trajectory matrix M, pathlet matrix D and trade off parameter λ. Then for constant parameter θ, we have the following bound on the cost of R^r:

$$P[C(R^r) \leq 2\theta\frac{\lambda+1}{\lambda}C(R^*)\ and\ DR^* \geq M] \geq \frac{1}{2} - |T|e^{-\theta}$$

Hierarchical Pathlet Learning.



1. Divide the map into small grids and learn local pathlets.
2. Merge small grids and extract high-level pathlets iteratively.

Experiment & Result

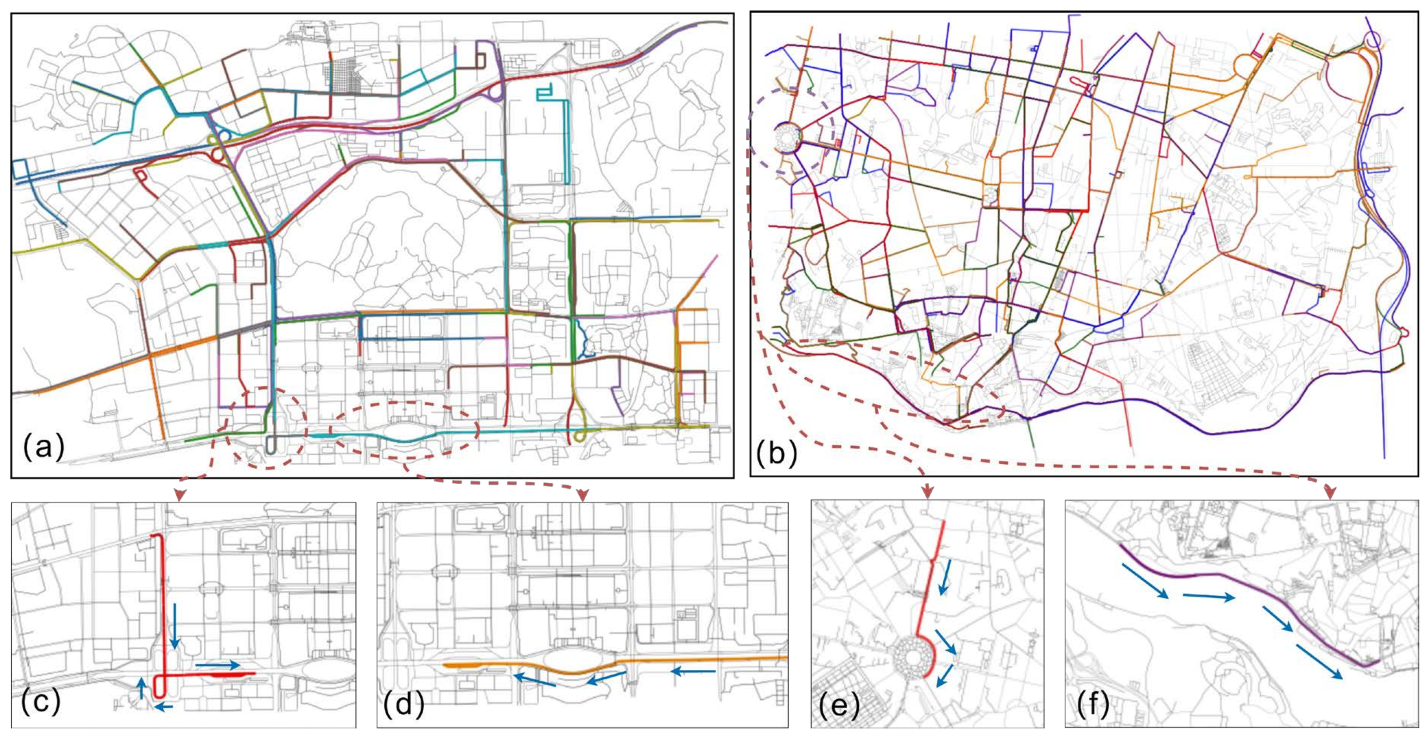


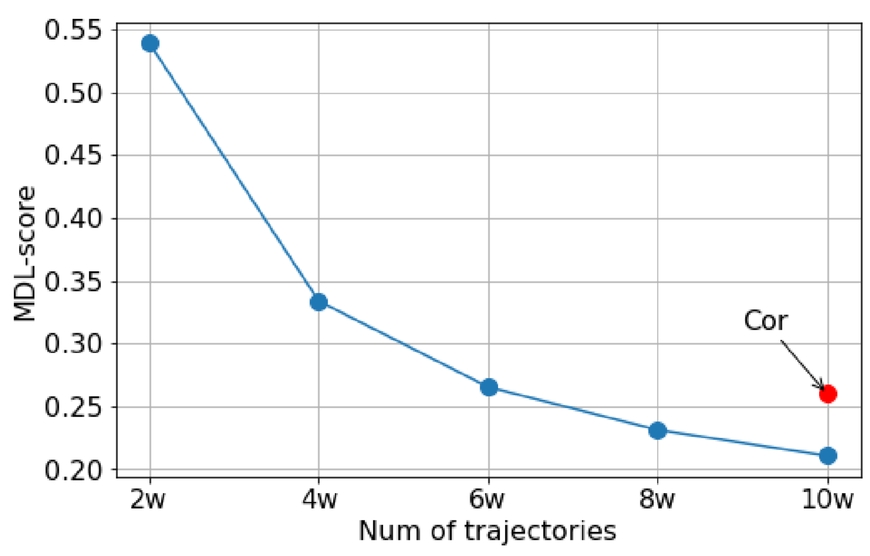
Fig 3. Visualization of Top 300 frequent pathlets.

Tab 1: The performance comparison with previous work.

Dataset	Method	Train Phrase			Test Phrase	
		dictionary size/|T|	rc	cover ratio	rc	cover ratio
Porto	DP	1.79	2.14	100%	4.19	88.03%
	ours	1.02(-43.01%)	2.00(-6.54%)	99.6%	2.75	93.6%(+5.57%)
Shenzhen	DP	1.21	2.88	100%	3.01	93.9%
	ours	0.91(-36.36%)	2.75 (-4.51%)	99.1%	3.02	95.4% (+1.5%)

Tab 2. Application in Trip Time Prediction.

dataset	method	MAE	MAPE	RMSE	RMSLE
Porto	[28]	-	-	-	0.41
	[3]	171.97	-	-	-
	Ours	163.9	26.2	199.74	0.35
Porto (short trips)	[27]	39.25	14.74	52.35	-
	Ours	32.87	9.37	37.59	0.08



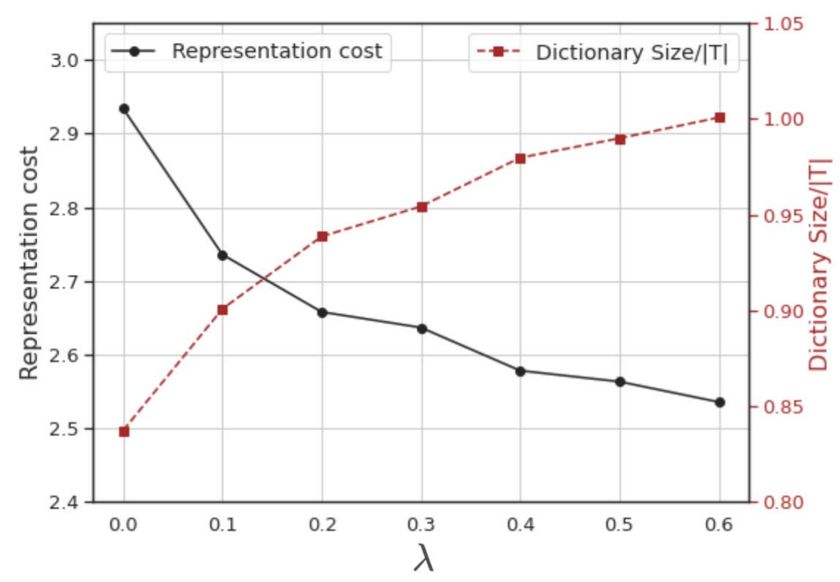


Fig 4. Application in Data Compression. Fig 5. Effect of hyper-parameter.

Conclusion

- A dictionary learning based method with theoretical probability bound analysis is proposed to solve the trajectory representation problem.
- The most frequent pathlets generated are visualized.
- The numerical experiment result demonstrates the effectiveness and broad prospects in downstream tasks.