TBSI 清华-伯克利深圳学院 Tsinghua-Berkeley Shenzhen Institute

A Robust Generative Model for trajectory Modeling and Application

Yuanbo Tang and Yang Li*

Introduction

Background.

- Trajectory generative models learn to synthesize new, realistic paths based on the distribution of existing trajectory data.
- It can benefit downstream tasks including privacy protection, trajectory prediction, anomaly detection, *Conditional Generation*.

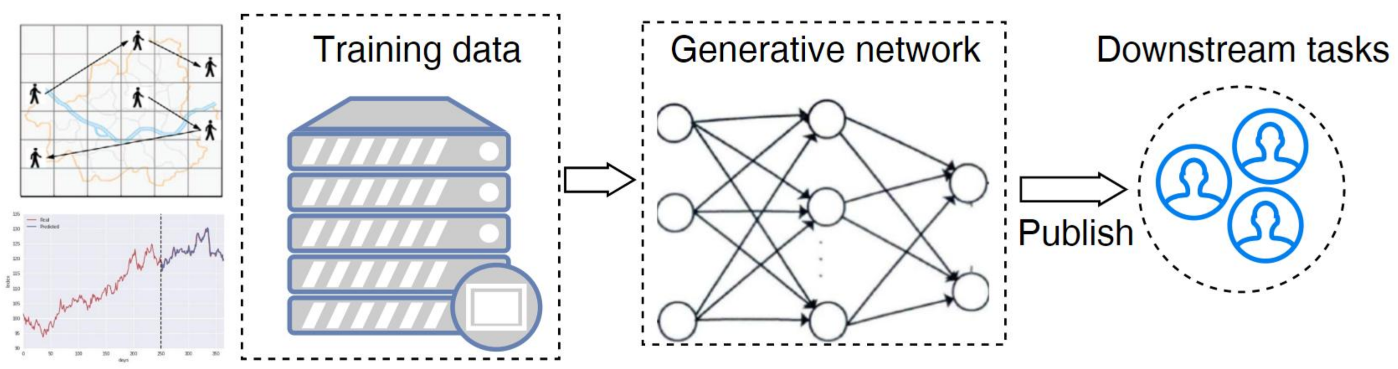


Fig 1. Workflow of a Trajectory Generation Model

Gap & Motivation.

- Few previous studies have tackled the issues of noise and incompleteness in raw trajectory datasets, which are common in real-world scenarios.
- Noise and incompleteness can cause distribution shifts, significantly reducing the performance.

Purpose.

- Develop a generative model framework that can effectively model trajectory distributions while being robust to noise and data incompleteness.
- Use generative model for downstream tasks.

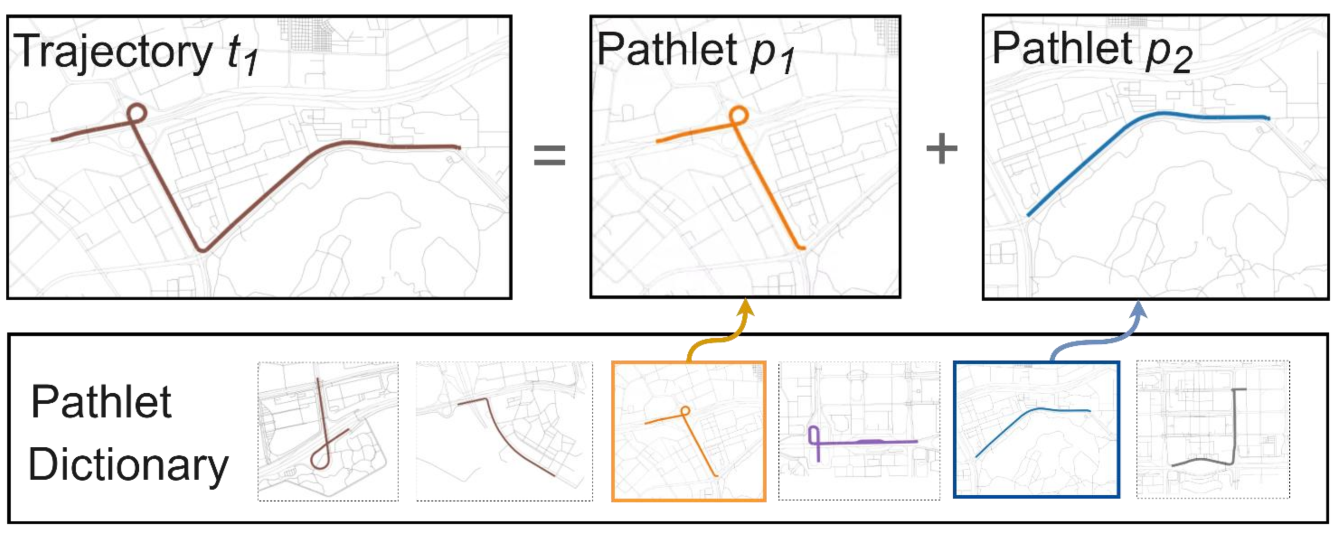


Fig 2. Basic idea of pathlet learning and reconstruction, the inherent sparsity makes algorithm more robust.

Main work

- We propose a probabilistic graphical model that combines CVAE with sparse dictionary learning to model trajectory distribution.
- We conducted experiments on real-world datasets to validate the effectiveness of our approach.

Problem Formulation

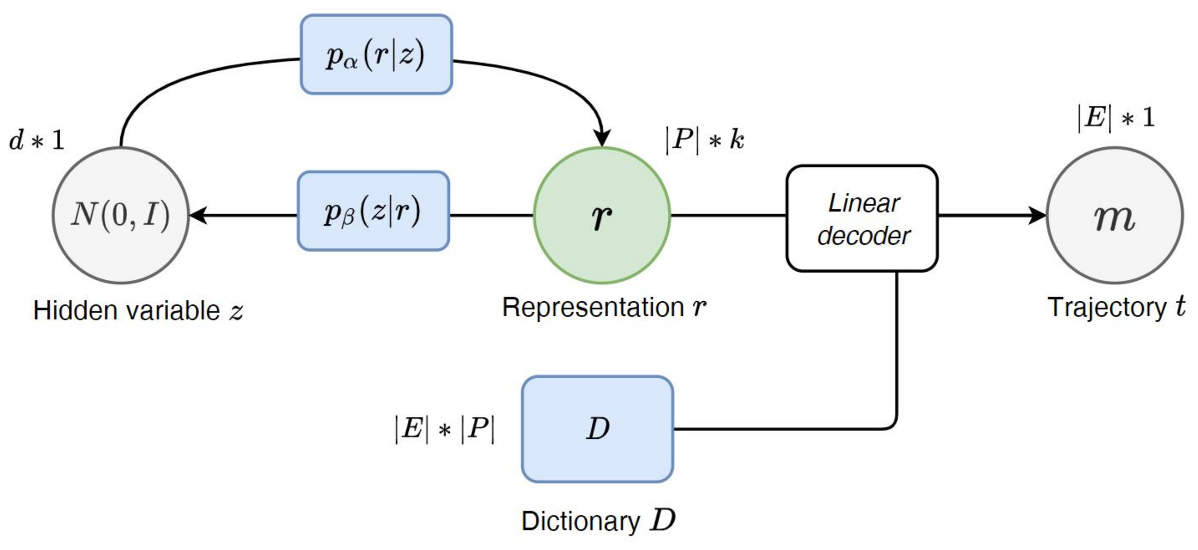


Fig 3. Illustration of the graphical model which describes generative process of trajectory.

Loss function

$$L_{all} = \boxed{L_{VAE}} + \boxed{L_{Dict}}$$

$$\boxed{L_{\mathrm{KL}}} = KL(p_\alpha(z|r)||N(0,I))$$

$$\boxed{L_{\mathrm{recon1}}} = \sum (r-\hat{r})^2$$

- This dictionary should be able to reconstruct all trajectories.
- Smaller dictionary is better.
- Average number of pathlets used to reconstruct trajectory should be as small as possible.

$$\boxed{L_{\mathrm{dict\_size}}} = \sum max(R_{i,:})$$

$$\boxed{L_{\mathrm{repr}}} = \sum ||r||_1$$

$$\boxed{L_{\mathrm{recon2}}} = \sum (m-\hat{m})^2$$

Method

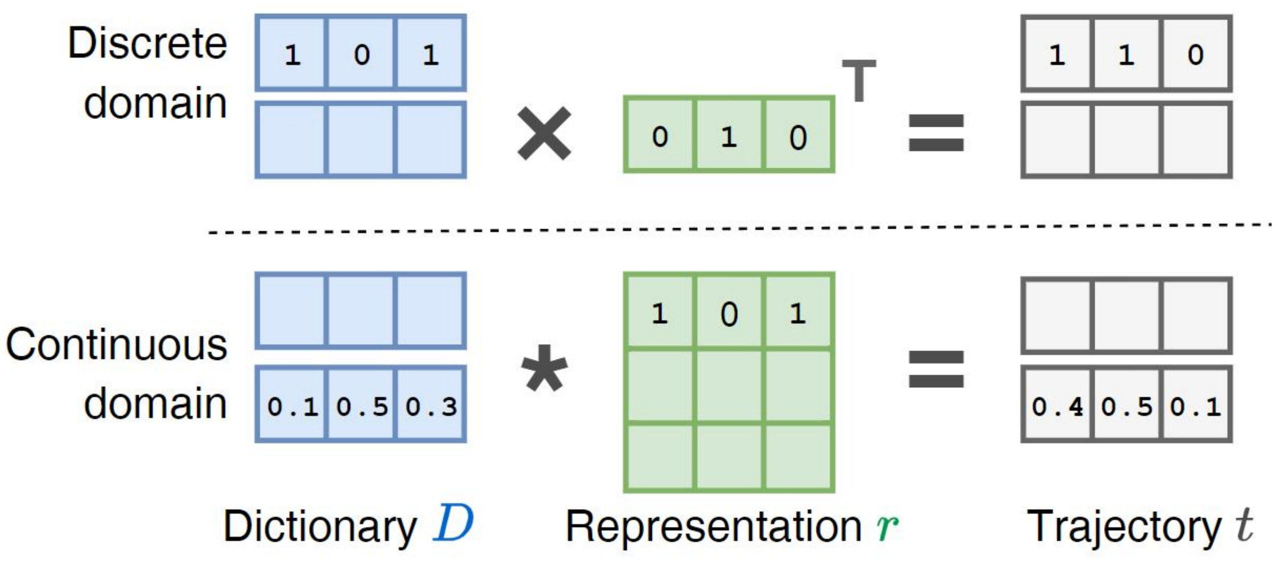
Training strategy

- The entire model is trained end-to-end using the gradient descent method.

Binary VAE

- Representation r are composed of binary elements, binary VAE is used to model its distribution.

Linear decoder



- Discrete domain: each element in r refers to use or not one pathlet; Matrix multiplication operations
- Continuous domain: r records the position to use corresponding pathlet; Convolution operations

Experiment & Result

Table 1: Statistic of trajectory datasets.

Name	#Trajectories	Avg.#Points	Avg.time gap
Porto	1.2M	60.20	15.00Sec.
Shenzhen	510K	43.96	21.66Sec.

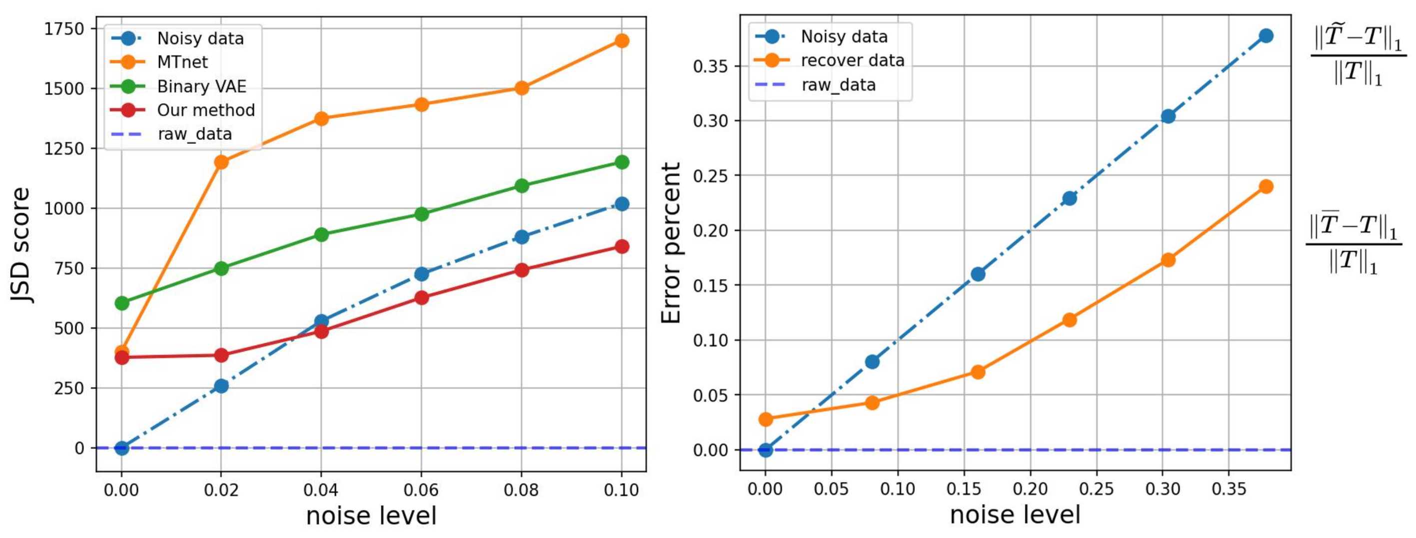


Fig 3. Left: Jsd score between the distribution of generated dataset and the original data. Right: Recovery performance under different noise levels.

Fig 3. Visualization of conditionally generated trajectories.

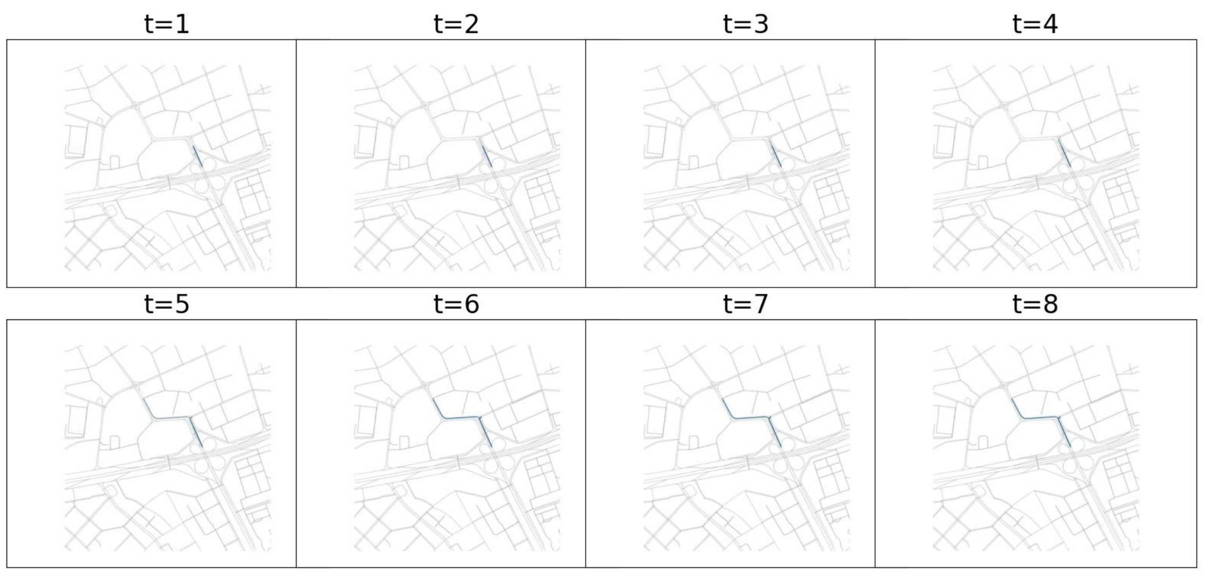


Fig 4. The learning process of pathlets from map-matched trajectories.

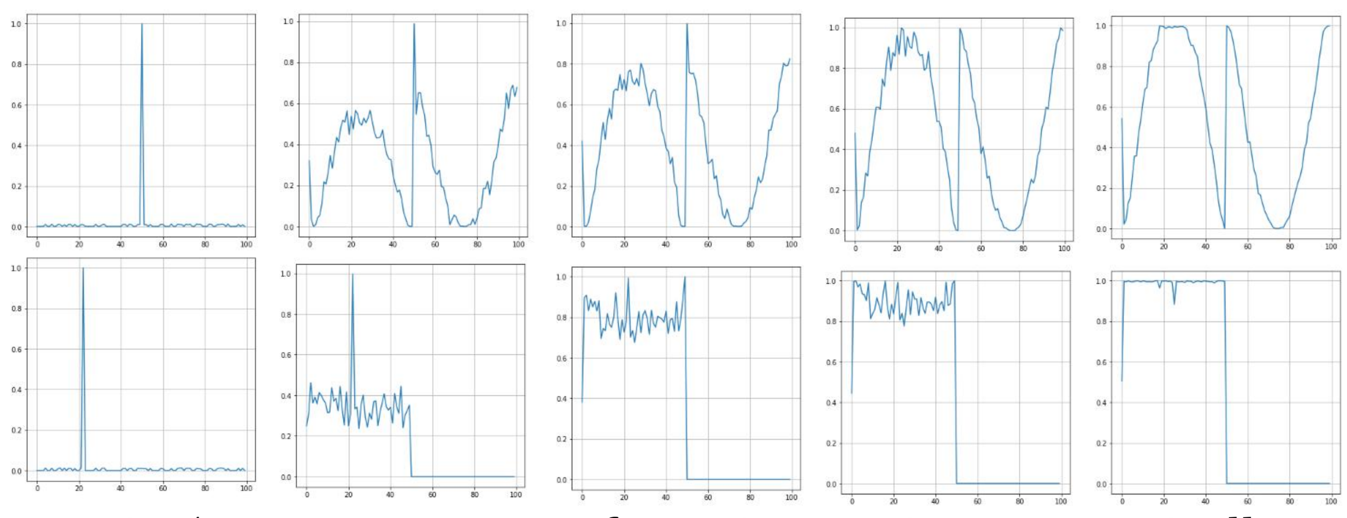


Fig 5. Extract patterns from time series automatically.

Conclusion

- A probabilistic graphical model for trajectory generation is proposed, which integrates pathlet dictionary learning and Variational Autoencoder (VAE).
- The numerical experiment result demonstrates the effectiveness and prospects in downstream tasks.