POLYTECHNIQUE MONTRÉAL

UNIVERSITÉ D'INGÉNIERIE

IVADO

transdim: Machine Learning for Transportation Data Imputation and Prediction

Reproducible Research Workshop

TRB 103rd Annual Meeting · Washington, D.C., USA

Xinyu Chen

January 11, 2024

Open-source & reproducible research:

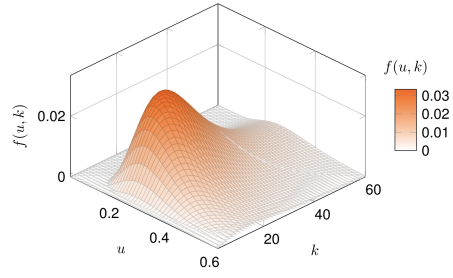
1. GitHub: https://github.com/xinychen
2. Slides: https://xinychen.github.io/slides/transdim.pdf

ML algorithms

transdim

(1.1k stars)

Visualization tools

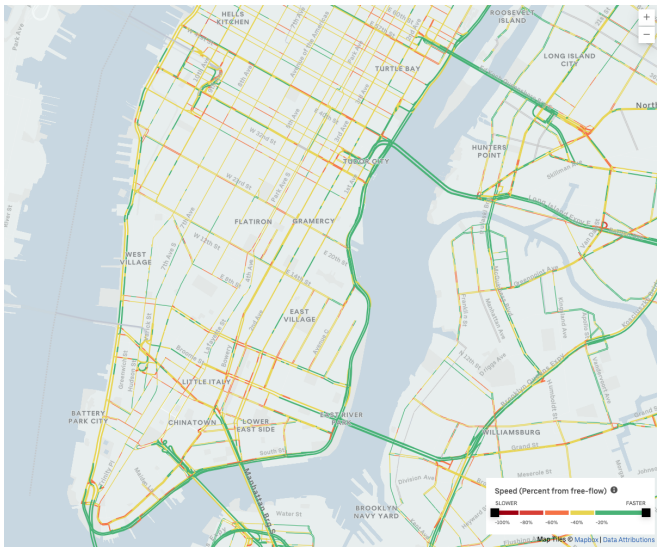


awesome-latex-drawing

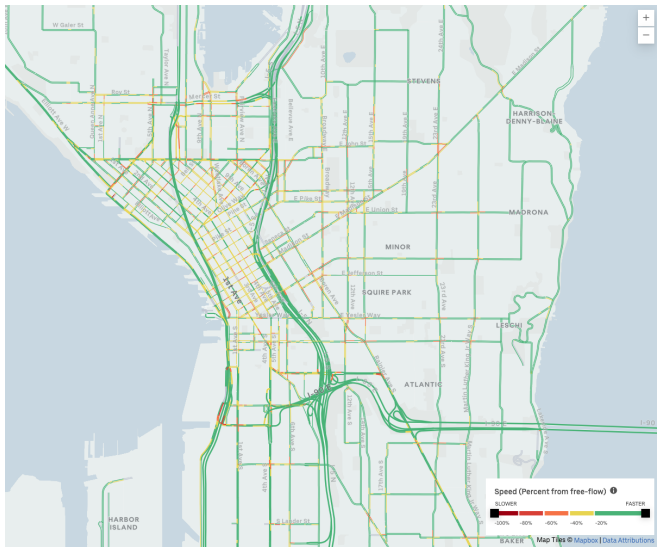
(1.2k stars)

Storytelling with Data

- Uber (hourly) movement speed data



NYC movement



Seattle movement

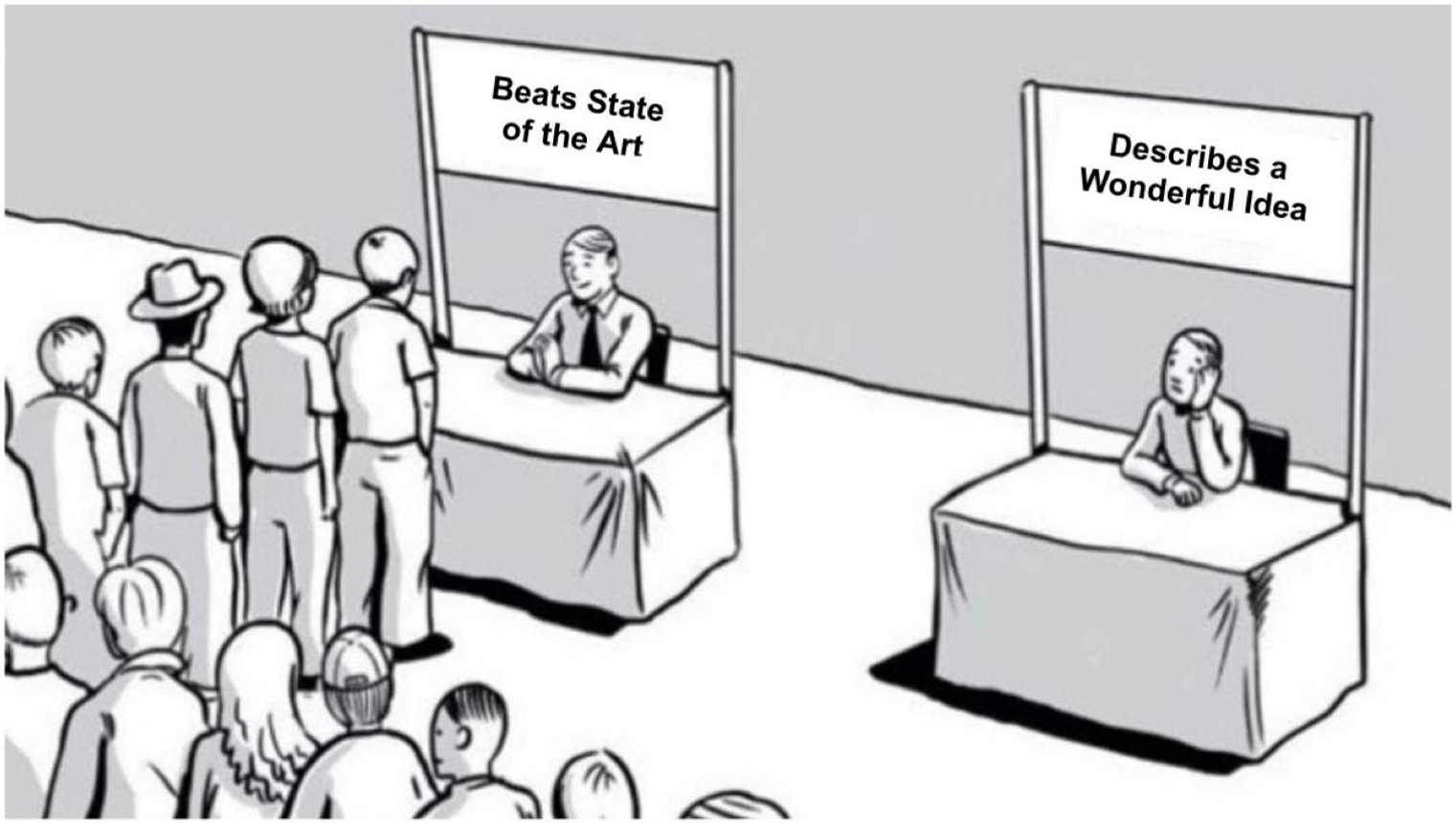
- {road segment, time step (hour), average speed}
- $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$ with N spatial locations $\times$ T time steps
- Computing hourly speed: Road segments have 5+ unique trips.

Issue: Insufficient sampling of ridesharing vehicles on the road network!

Storytelling with Data

- Data
- Quality
- Sparsity
- Estimation
- Imputation
- Interpolation
- Forecasting

Storytelling with Data



Source: Twitter

Traffic Data Processing

Computing with `numpy` (numerical computing in Python)

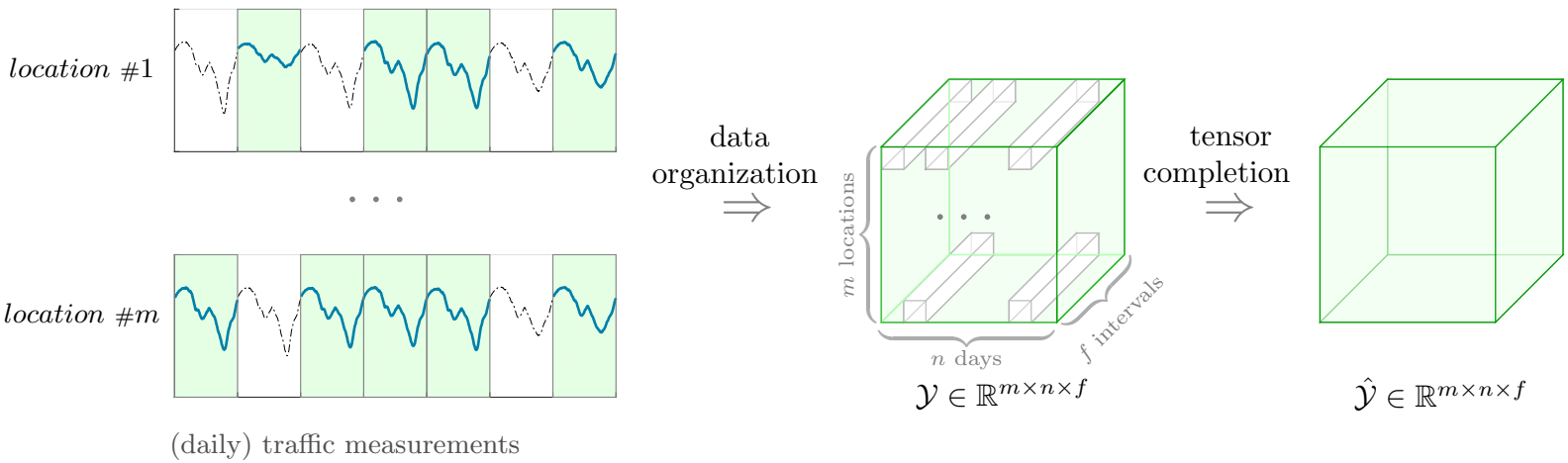
- Data format: `.npz` (compressed format of `.npy`)
- Example
- Easy to connect with `numpy` (in CPU environment) & `cupy` (in GPU environment)

Reformulate Traffic Data Imputation

- Represent traffic data as tensors

$$\text{Tensorization: } \boldsymbol{Y} \in \mathbb{R}^{m \times t} \rightarrow \boldsymbol{\mathcal{Y}} \in \mathbb{R}^{m \times n \times f}$$

w/ m locations, n days, and f time intervals per day.



- Tensor completion (Observed index set Ω)

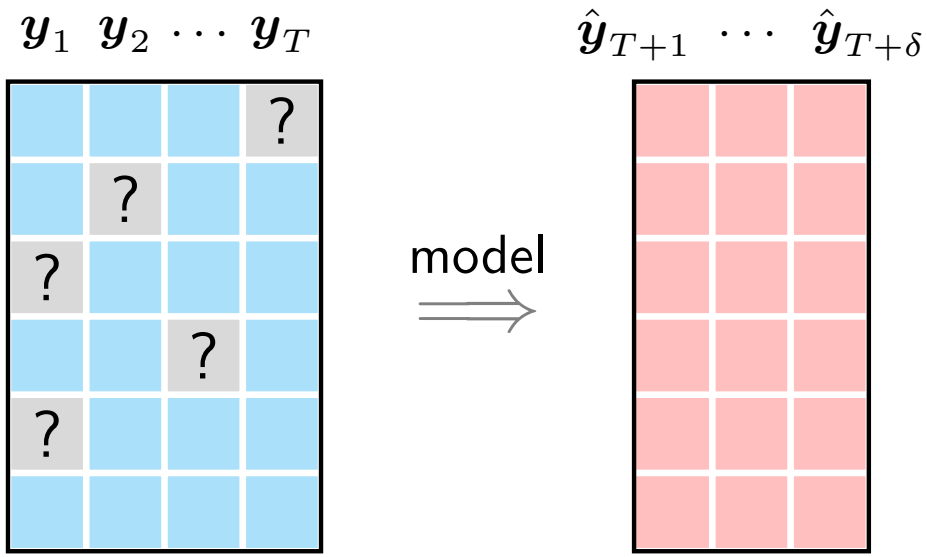
$$\underbrace{\mathcal{P}_\Omega(\boldsymbol{\mathcal{Y}})}_{\text{Partially observed}} \xrightarrow{\text{Estimate}} \underbrace{\mathcal{P}_\Omega^\perp(\boldsymbol{\mathcal{Y}})}_{\text{Unobserved}}$$

Reformulate Traffic Forecasting

Forecasting urban traffic states with sparse data

- Problem definition (δ-step ahead forecasting)

$$\underbrace{\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_T\}}_{\text{Current traffic states}} \xrightarrow{\text{Estimate}} \underbrace{\{\hat{\boldsymbol{y}}_{T+1}, \hat{\boldsymbol{y}}_{T+2}, \ldots, \hat{\boldsymbol{y}}_{T+\delta}\}}_{\text{Future traffic states}}$$



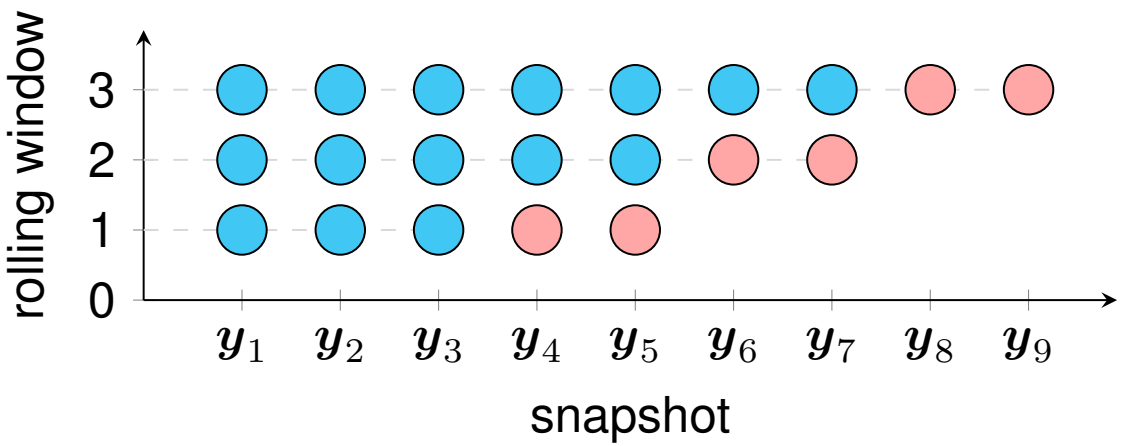
Reformulate Traffic Forecasting

(Rolling) Forecasting urban traffic states with sparse data

1st rolling step: $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3\} \rightarrow \{\boldsymbol{y}_4, \boldsymbol{y}_5\}$

2nd rolling step: $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \boldsymbol{y}_4, \boldsymbol{y}_5\} \rightarrow \{\boldsymbol{y}_6, \boldsymbol{y}_7\}$

3rd rolling step: $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{y}_3, \boldsymbol{y}_4, \boldsymbol{y}_5, \boldsymbol{y}_6, \boldsymbol{y}_7\} \rightarrow \{\boldsymbol{y}_8, \boldsymbol{y}_9\}$



Switch from CPU to GPU

Python implementation of algorithms with the `numpy` package
Easy to convert the codes from CPU to GPU

```
import numpy as np   ⇒   import cupy as np
```

Why?

Academic:

- Sustainable research environment (w.r.t. us & followers)
- Interact with researchers from different fields
- Provide platform and benchmark for comparison
- Stimulate new algorithmic ideas

Industry:

- Solution to ...

Next-step plan:

- Spatiotemporal data modeling:
  https://spatiotemporal-data.github.io

POLYTECHNIQUE MONTRÉAL

UNIVERSITÉ D'INGÉNIERIE

IVADO

Thanks for your attention!

Any Questions?

About me:

- Homepage: https://xinychen.github.io
- GitHub: https://github.com/xinychen
- How to reach me: chenxy346@gmail.com