POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

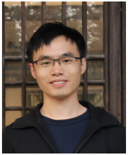
IVADO

Spatiotemporal Traffic Data Imputation and Forecasting with Tensor Learning

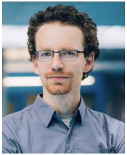
Ph.D. Research Project

Xinyu Chen

May 24, 2022



Ph.D. candidate
Xinyu Chen
Polytechnique Montréal



Supervisor
Prof. Nicolas Saunier
Polytechnique Montréal

Co-supervisor
Prof. Lijun Sun
McGill University

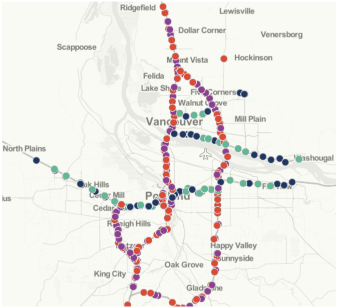
Outline

- **Motivation**
  - Multivariate traffic time series
  - Multidimensional traffic time series
  - Multiple data behaviors
- **Literature Review**
  - Spatiotemporal traffic data imputation
  - Spatiotemporal traffic forecasting
  - Low-rank tensor learning
- **Objective A**
  - Spatiotemporal traffic data imputation
- **Objective B**
  - High-dimensional traffic forecasting
  - Multidimensional traffic forecasting
- **Objective C**
  - Multivariate traffic forecasting on sparse data
  - Multidimensional traffic forecasting on sparse data
- **Conclusion**

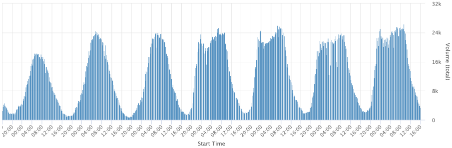
Multivariate Traffic Time Series

Many spatiotemporal traffic time series data are in the form of **matrix**.

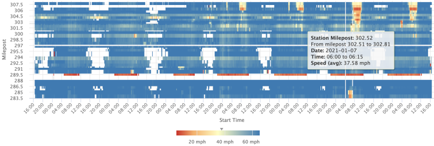
- **Example**: Portland highway traffic data[1].



Detector locations



Traffic volume



Traffic speed

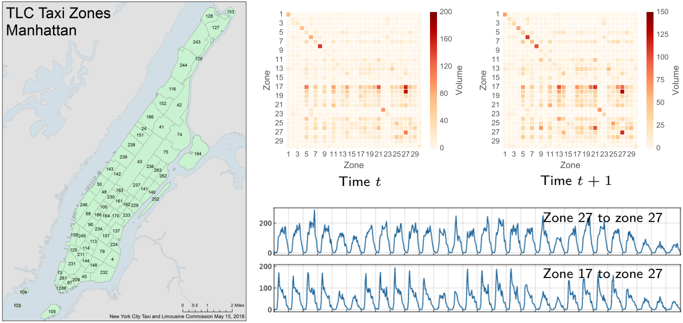
- $\boldsymbol{X} \in \mathbb{R}^{N \times T}$ with N spatial locations $\times$ T time steps

[1] https://portal.its.pdx.edu/home

Multidimensional Traffic Time Series

Many spatiotemporal traffic time series data are in the form of **tensor**.

- **Example:** NYC (hourly) taxi flow data[2].



- $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M \times N \times T}$ with M zones $\times$ N zones $\times$ T time steps

[2] https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page

Multiple Data Behaviors

Spatiotemporal traffic data are time series, but they involve multiple data behaviors.

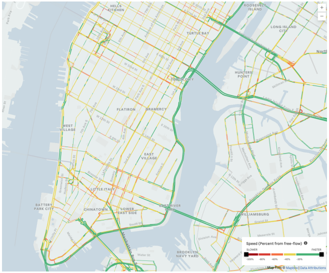
- Incompleteness & sparsity
- High-dimensionality
- Multidimensionality
- Noises & outliers
- Time-varying behavior
- Nonstationarity
- ......

In addition, spatiotemporal correlations are also very important.

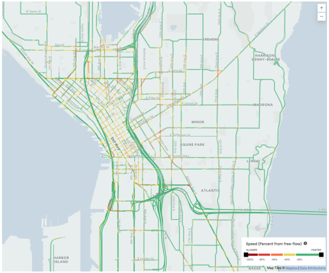
Multiple Data Behaviors

Sparsity & high-dimensionality

- Uber (hourly) movement speed data[3]



NYC movement



Seattle movement

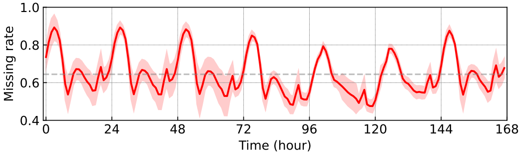
- The average speed on a given road segment for each hour of each day.
- Hourly speeds are computed when road segments have 5+ unique trips.
- **Issue**: insufficient sampling of ridesharing vehicles on the road network.

[3] https://movement.uber.com/

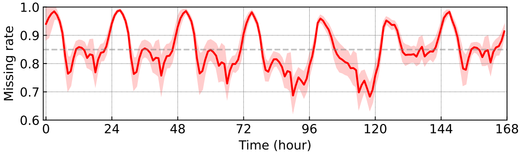
Multiple Data Behaviors

Sparsity & high-dimensionality

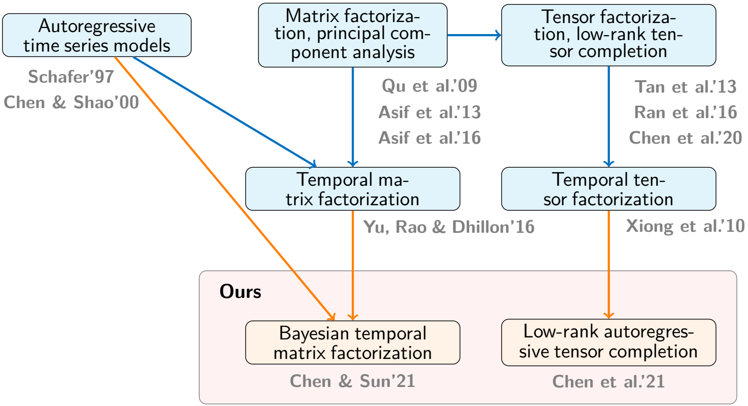
- **NYC** movement speed data (2019)
  - **98,210** road segments & 8,760 time steps (hours)
  - Overall missing rate: **64.43%**



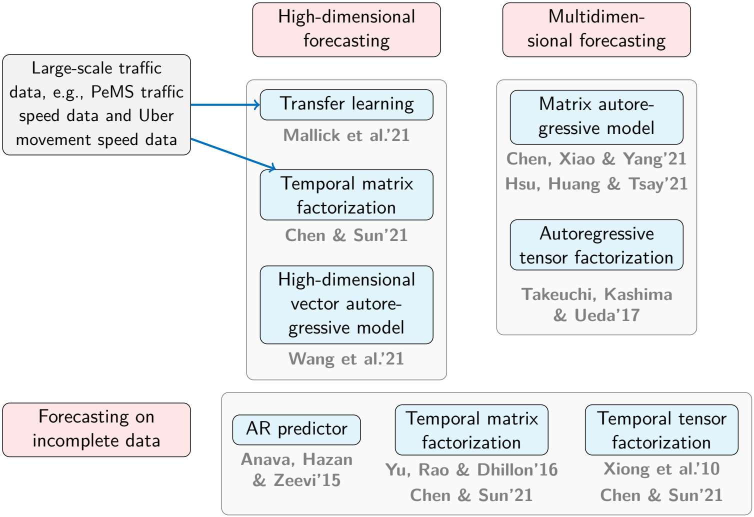
- **Seattle** movement speed data (2019)
  - **63,490** road segments & 8,760 time steps (hours)
  - Overall missing rate: **84.95%**



Spatiotemporal Traffic Data Imputation



Spatiotemporal Traffic Forecasting



Low-Rank Tensor Learning

1927 — Higher-Order SVD

Frank Lauren Hitchcock

1960s — Tucker Decomposition

Ledyard R. Tucker

1970 — CP Decomposition

J. Douglas Carroll
Jih-Jie Chang
Richard A. Harshman

2009 — Tensor Decompositions and Applications

Tamara G. Kolda

2011 — Tensor-Train Decomposition

Ivan Oseledets

Low-Rank Tensor Learning

- Low-rank matrix/tensor completion

Candès & Recht'09: Convex nuclear norm minimization for matrix completion.

$$\min_{\boldsymbol{X}} \|\boldsymbol{X}\|_* \quad \text{s.t. } \mathcal{P}_\Omega(\boldsymbol{X}) = \mathcal{P}_\Omega(\boldsymbol{Y})$$

Cai, Candès & Shen'10: Singular value thresholding algorithm.

$$\begin{cases} \boldsymbol{X}^\ell = \mathcal{D}_\tau\left(\boldsymbol{Z}^{\ell-1}\right) \\ \boldsymbol{Z}^\ell = \boldsymbol{Z}^{\ell-1} + \delta_\ell \mathcal{P}_\Omega\left(\boldsymbol{Y} - \boldsymbol{X}^\ell\right) \end{cases}$$

Zhang et al.'12: Nonconvex truncated nuclear norm minimization.

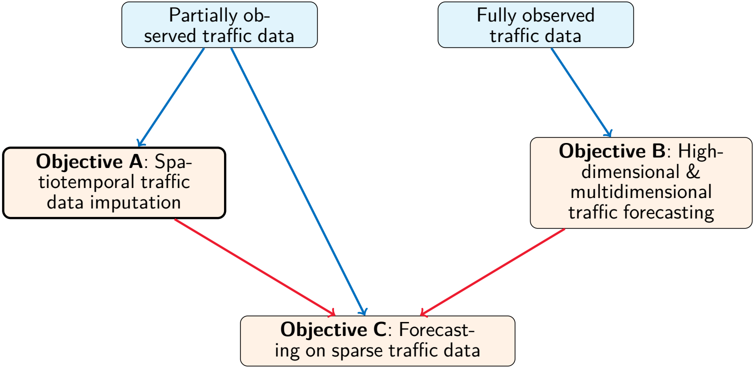
Liu et al.'13: Convex nuclear norm minimization for tensor completion.

$$\min_{\boldsymbol{\mathcal{X}}} \|\boldsymbol{\mathcal{X}}\|_* \quad \text{s.t. } \mathcal{P}_\Omega(\boldsymbol{\mathcal{X}}) = \mathcal{P}_\Omega(\boldsymbol{\mathcal{Y}})$$

Lu, Peng & Wei'19: Tensor nuclear norm induced by linear transform.

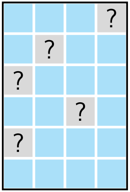
A Whole Picture of Objectives

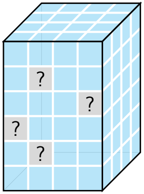
We are working on **spatiotemporal traffic data modeling**.



Spatiotemporal Traffic Data Imputation

- **Objective A:** Given a multivariate time series data like $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$ or a multidimensional time series data like $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{M \times N \times T}$, impute the missing values of the data.





[Q]

- How to reconstruct missing values from observed data?
- How to make use of spatiotemporal correlations?
- How to make use of traffic time series dynamics?

Spatiotemporal Traffic Data Imputation

Low-rank matrix completion (Candès & Recht'09)

For any partially observed data matrix $\boldsymbol{Y} \in \mathbb{R}^{N \times T}$ with observed index set Ω, then low-rank matrix completion takes the form of

$$\begin{aligned} \min_{\boldsymbol{X}} \; & \|\boldsymbol{X}\|_* \\ \text{s.t. } & \mathcal{P}_{\Omega}(\boldsymbol{X}) = \mathcal{P}_{\Omega}(\boldsymbol{Y}). \end{aligned} \tag{1}$$

Low-rank tensor completion (Liu et al.'13)

For any partially observed data matrix $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{M \times N \times T}$ with observed index set Ω, then low-rank matrix completion takes the form of

$$\begin{aligned} \min_{\boldsymbol{\mathcal{X}}} \; & \|\boldsymbol{\mathcal{X}}\|_* \\ \text{s.t. } & \mathcal{P}_{\Omega}(\boldsymbol{\mathcal{X}}) = \mathcal{P}_{\Omega}(\boldsymbol{\mathcal{Y}}). \end{aligned} \tag{2}$$

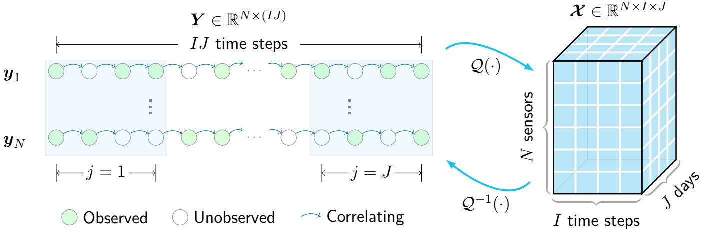
- **Limitation:** Only cover the global consistency.
- **Comment:** For modeling spatiotemporal traffic data, local consistency (e.g., temporal correlations) is also important.

Spatiotemporal Traffic Data Imputation

Low-rank autoregressive tensor completion (LATC)

$$\min_{\boldsymbol{\mathcal{X}}} \ \|\boldsymbol{\mathcal{X}}\|_* + \frac{\lambda}{2}\sum_{n=1}^{N}\sum_{t=d+1}^{T}\Big(z_{n,t} - \sum_{k=1}^{d} a_{n,k} z_{n,t-k}\Big)^2$$
$$\text{s.t.} \ \begin{cases} \boldsymbol{\mathcal{X}} = \mathcal{Q}(\boldsymbol{Z}), \\ \mathcal{P}_{\Omega}(\boldsymbol{Z}) = \mathcal{P}_{\Omega}(\boldsymbol{Y}). \end{cases} \quad (3)$$

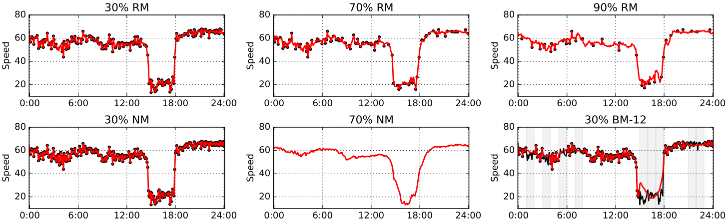
- **Advantage:** Global consistency + local consistency.



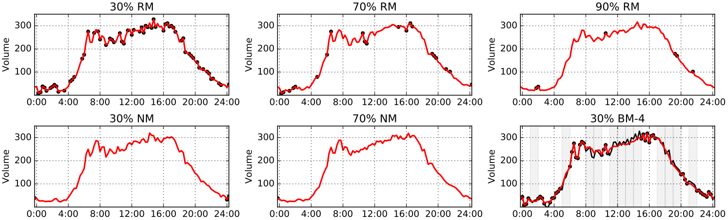
Spatiotemporal Traffic Data Imputation

LATC imputation

- Seattle freeway traffic speed data

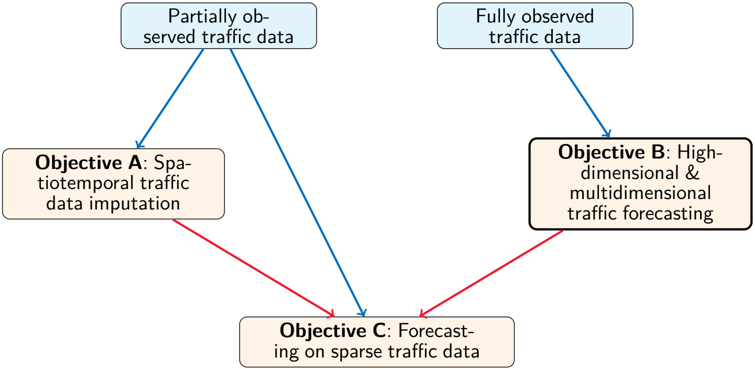


- Portland highway traffic volume data



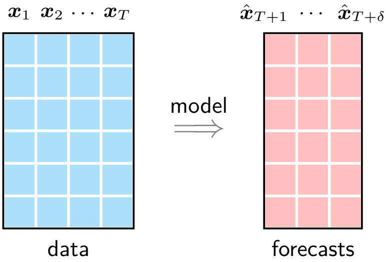
A Whole Picture of Objectives

We are working on **spatiotemporal traffic data modeling**.



High-Dimensional Traffic Forecasting

- **Objective B-1:** Given a multivariate traffic time series $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T \in \mathbb{R}^N$ with $N \gg T$ ("tall-skinny"), forecast data points $\boldsymbol{x}_{T+\delta}, \delta \in \mathbb{N}^+$.

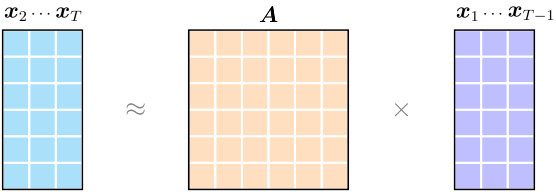


- **Solution:** For time series $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T \in \mathbb{R}^N$, the dth-order vector autoregressive (VAR(d)) model: $\boldsymbol{x}_t = \sum_{k=1}^{d} \boldsymbol{A}_k \boldsymbol{x}_{t-k} + \boldsymbol{\epsilon}_t$.
- **Advantage:** Co-evolution patterns
- **Limitation:** Over-parameterization

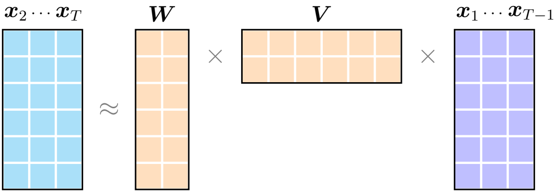
High-Dimensional Traffic Forecasting

VAR(1) model

- Over-parameterization in the case of $N \gg T$.



- Reduced-rank autoregression: $\boldsymbol{A} = \boldsymbol{W}\boldsymbol{V}$ with $\boldsymbol{W} \in \mathbb{R}^{N \times R}, \boldsymbol{V} \in \mathbb{R}^{R \times N}$.



High-Dimensional Traffic Forecasting

VAR(d) model

- Recall that $\boldsymbol{x}_t = \sum_{k=1}^{d} \boldsymbol{A}_k \boldsymbol{x}_{t-k} + \boldsymbol{\epsilon}_t$.
- Coefficients $\boldsymbol{A}_k \in \mathbb{R}^{N \times N}, k = 1, \ldots, d$ are tensor, e.g., $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{N \times N \times d}$.

VAR(d) with Tucker decomposition (Wang et al.'21)

For VAR(d) on the multivariate time series $\boldsymbol{x}_t \in \mathbb{R}^N, t = 1, \ldots, T$, the reduced-rank VAR via Tucker decomposition is given by

$$\min_{\boldsymbol{\mathcal{G}}, \boldsymbol{W}, \boldsymbol{U}, \boldsymbol{H}} \frac{1}{2} \sum_{t=d+1}^{T} \|\boldsymbol{x}_t - (\boldsymbol{\mathcal{G}} \times_1 \boldsymbol{W} \times_2 \boldsymbol{U} \times_3 \boldsymbol{H})_{(1)} \boldsymbol{z}_t\|_2^2 \tag{4}$$

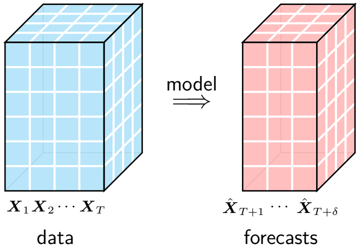
where $\boldsymbol{z}_t = (\boldsymbol{x}_{t-1}^\top, \cdots, \boldsymbol{x}_{t-d}^\top)^\top \in \mathbb{R}^{dN}$. The multilinear rank is (R_1, R_2, R_3). $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{R_1 \times R_2 \times R_3}$ is the core tensor, while $\boldsymbol{W} \in \mathbb{R}^{N \times R_1}$, $\boldsymbol{U} \in \mathbb{R}^{N \times R_2}$, and $\boldsymbol{H} \in \mathbb{R}^{d \times R_3}$ are the component matrices.

Advantage: High compression rate.

Limitations: Nonconvex optimization; the model is failing in nonstationary time series.

Multidimensional Traffic Forecasting

- **Objective B-2:** Given a multidimensional traffic time series $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_T \in \mathbb{R}^{M \times N}$, forecast data points $\boldsymbol{X}_{T+\delta}, \delta \in \mathbb{N}^+$.



[Q]

- How to perform forecasting on this kind of data?
- How to preserve the intrinsic tensor representation of data?

Multidimensional Traffic Forecasting

Matrix autoregressive model (Chen, Xiao & Yang'21)

Given matrix-variate time series $\boldsymbol{X}_t \in \mathbb{R}^{M \times N}, t = 1, \ldots, T$, then the dth-order matrix autoregressive (MAR(d)) model takes the form of

$$\boldsymbol{X}_t = \sum_{k=1}^{d} \boldsymbol{A}_k \boldsymbol{X}_{t-k} \boldsymbol{B}_k^\top + \boldsymbol{E}_t \tag{5}$$

where $\boldsymbol{A}_k \in \mathbb{R}^{M \times M}, \boldsymbol{B}_k \in \mathbb{R}^{N \times N}, k = 1, \ldots, d$ are the coefficient matrices.

Advantages:

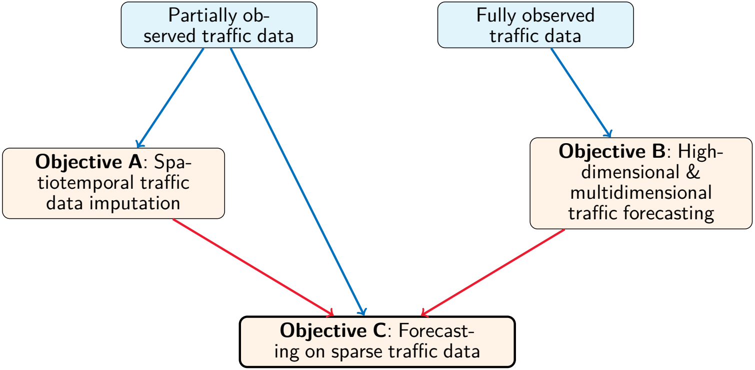
- Preserve the intrinsic tensor representation.
- Reduce parameters in autoregressive models (if $n = \max\{M, N\}$), e.g.,

$$\mathcal{O}(n^4) \text{ in VAR}(1) \quad \text{vs.} \quad \mathcal{O}(n^2) \text{ in MAR}(1)$$

Limitation: Failing in nonstationary time series.

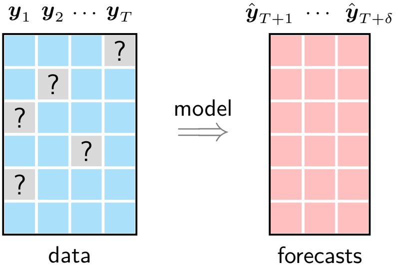
A Whole Picture of Objectives

We are working on **spatiotemporal traffic data modeling**.



Multivariate Traffic Forecasting on Sparse Data

- **Objective C-1:** Given a partially observed data $\boldsymbol{Y} \in \mathbb{R}^{N\times T}$ consisting of time series $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T \in \mathbb{R}^N$, forecast data points $\boldsymbol{y}_{T+\delta}, \delta \in \mathbb{N}^+$.



[Q]

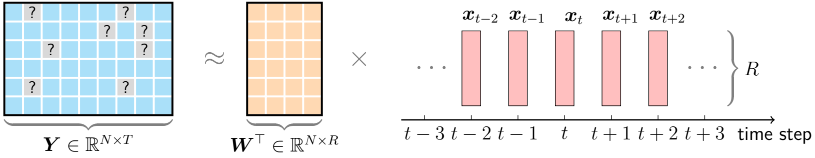
- How to learn from *high-dimensional* and *sparse* data?
- How to model *nonstationarity* in time series?
- How to perform forecasting on these time series?

Multivariate Traffic Forecasting on Sparse Data

Temporal matrix factorization (Yu et al.'16; Chen & Sun'21)

Given any partially observed time series data $\boldsymbol{Y}\in\mathbb{R}^{N\times T}$ with observed index set Ω, then temporal matrix factorization assumes a dth-order vector autoregressive (VAR) process on the temporal factor matrix:

$$\min_{\boldsymbol{W},\boldsymbol{X},\{\boldsymbol{A}_k\}_{k=1}^{d}} \frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^\top\boldsymbol{X})\|_F^2+\frac{\rho}{2}(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2) + \frac{\lambda}{2}\sum_{t=d+1}^{T}\|\boldsymbol{x}_t-\sum_{k=1}^{d}\boldsymbol{A}_k\boldsymbol{x}_{t-k}\|_2^2 \tag{6}$$



🙁 VAR is usually built on stationary time series (temporal factors).

Multivariate Traffic Forecasting on Sparse Data

Nonstationary temporal matrix factorization (NoTMF)

Given any partially observed time series data $\boldsymbol{Y}\in\mathbb{R}^{N\times T}$ with observed index set Ω, then we assume a season-m differencing on the latent temporal factors:

$$\begin{aligned}\min_{\boldsymbol{W},\boldsymbol{X},\{\boldsymbol{A}_k\}_{k=1}^{d}}\quad&\frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^\top\boldsymbol{X})\|_F^2+\frac{\rho}{2}(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2)\\&+\frac{\lambda}{2}\sum_{t=d+m+1}^{T}\|(\boldsymbol{x}_t-\boldsymbol{x}_{t-m})-\sum_{k=1}^{d}\boldsymbol{A}_k(\boldsymbol{x}_{t-k}-\boldsymbol{x}_{t-m-k})\|_2^2\end{aligned}\tag{7}$$

- First-order differencing $\boldsymbol{x}_t'=\boldsymbol{x}_t-\boldsymbol{x}_{t-1}$.
- Second-order differencing $\boldsymbol{x}_t''=(\boldsymbol{x}_t-\boldsymbol{x}_{t-1})-(\boldsymbol{x}_{t-1}-\boldsymbol{x}_{t-2})$.
- Twice-differenced series $\boldsymbol{x}_t''=(\boldsymbol{x}_t-\boldsymbol{x}_{t-m})-(\boldsymbol{x}_{t-1}-\boldsymbol{x}_{t-m-1})$.

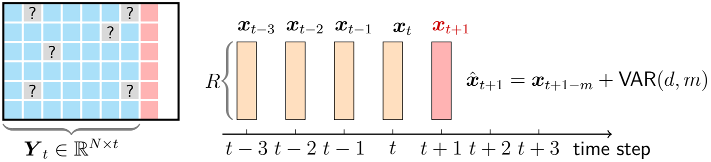
🙂 Stationarizing a time series with differencing can improve the prediction.[4]

[4] Stationarity and differencing: https://otexts.com/fpp2/stationarity.html

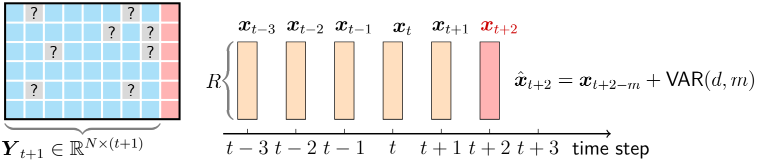
Multivariate Traffic Forecasting on Sparse Data

NoTMF forecasting on streaming data?

- NoTMF: Use $\boldsymbol{Y}_t$ to estimate $\{\boldsymbol{W}, \boldsymbol{X}, \boldsymbol{A}\}$.



- Online forecasting (Gultekin & Paisley'18): Fix $\boldsymbol{W}$ and use $\boldsymbol{Y}_{t+1}$ to update $\{\boldsymbol{X}, \boldsymbol{A}\}$.

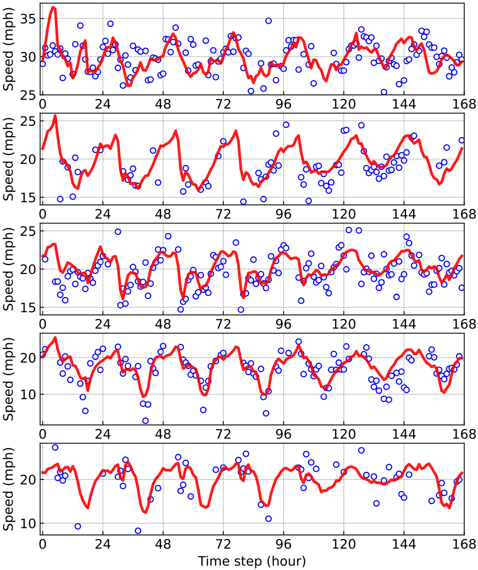


Multivariate Traffic Forecasting on Sparse Data

Forecasting performance on NYC Uber movement speed data (MAPE/RMSE):

- NoTMF outperforms other models.
- Seasonal differencing in NoTMF is important.

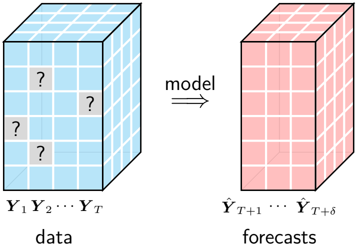
δ	d	NoTMF ($m=24$)	NoTMF ($m=168$)	NoTMF-twice ($m=168$)	TMF	TRMF	BTMF
1	1	13.63/2.88	13.53/2.86	**13.45**/**2.85**	13.74/2.90	14.50/3.12	14.94/3.13
	2	13.47/**2.84**	**13.41**/**2.84**	13.42/**2.84**	13.53/2.85	14.14/3.05	15.70/3.41
	3	13.46/2.84	**13.39**/**2.83**	13.43/2.84	13.47/**2.83**	13.87/2.96	15.80/3.34
	6	13.41/**2.83**	**13.39**/**2.83**	13.41/**2.83**	13.40/**2.83**	14.00/2.98	15.45/3.27
2	1	13.91/2.96	13.76/2.94	**13.70**/**2.92**	14.24/3.00	15.85/3.43	15.33/3.21
	2	13.77/2.92	**13.63**/**2.89**	13.72/2.92	13.87/2.91	15.04/3.31	15.87/3.32
	3	13.72/2.91	**13.61**/**2.89**	13.73/2.92	13.81/**2.89**	15.25/3.36	15.69/3.33
	6	13.59/2.87	**13.57**/2.88	13.68/2.91	13.63/**2.86**	14.92/3.24	15.91/3.39
3	1	14.30/3.05	14.06/3.02	**14.02**/**3.00**	14.81/3.12	17.52/3.83	15.86/3.32
	2	14.01/2.98	**13.84**/**2.94**	13.96/2.98	14.26/2.98	17.32/4.00	16.30/3.40
	3	13.95/2.97	**13.79**/**2.93**	13.98/2.98	14.04/2.94	16.91/3.71	16.56/3.49
	6	13.78/**2.92**	**13.73**/**2.92**	13.91/2.96	13.94/**2.92**	16.72/3.65	15.49/3.27
6	1	**14.61**/**3.11**	14.67/3.20	14.98/3.32	15.41/3.21	21.20/4.70	15.99/3.32
	2	**14.30**/**3.03**	14.33/3.09	14.90/3.28	14.85/3.07	20.87/5.01	16.04/3.33
	3	**14.26**/3.03	14.28/3.09	14.86/3.26	14.57/**3.01**	20.08/4.65	15.67/3.28
	6	**14.06**/**2.97**	14.16/3.06	14.80/3.23	14.47/3.00	20.40/4.35	16.38/3.50



Speed (mph)
Time step (hour)
0
24
48
72
96
120
144
168

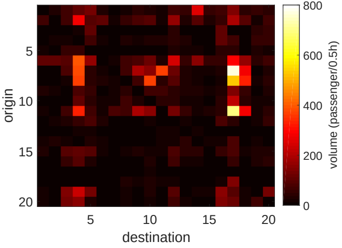
Multidimensional Traffic Forecasting on Sparse Data

- **Objective C-2:** Given a partially observed data $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{M \times N \times T}$ consisting of time series $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_T \in \mathbb{R}^{M \times N}$, forecast data points $\boldsymbol{Y}_{T+\delta}, \delta \in \mathbb{N}^+$.

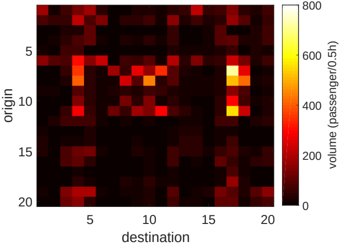


- **Solution:** Temporal tensor factorization, e.g., tensor factorization + VAR (on latent temporal factors).

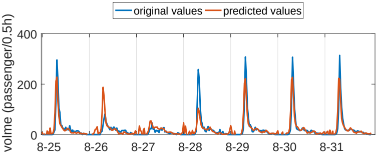
Multidimensional Traffic Forecasting on Sparse Data



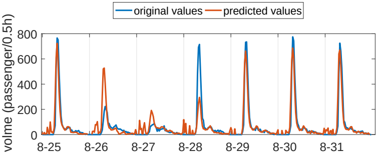
Original OD volume matrix.



Predicted OD volume matrix.



Volume of the #(6,17) OD pair.



Volume of the #(7,17) OD pair.

Conclusion

- **Contributions**
  - *Objective A: Spatiotemporal traffic data imputation.* Develop a low-rank temporal modeling framework and improve the imputation accuracy, efficiency, and scalability.
  - *Objective B: High-dimensional and multidimensional forecasting.* Fast and accurate forecasting approach for high-dimensional and large-scale data; tensor representation based autoregressive model for multidimensional data.
  - *Objective C: Forecasting on sparse data.* Low-rank temporal modeling framework for traffic time series forecasting in the presence of missing values.
- **Significance**
  - Improve traffic data quality.
  - Support data-driven intelligent transportation applications.

Research Work during Ph.D. Research

- **Publications**
  - [**J1**] **X. Chen**, M. Lei, N. Saunier, L. Sun (2021). Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. *IEEE Transactions on Intelligent Transportation Systems*. (Early access)
  - [**J2**] **X. Chen**, Y. Chen, N. Saunier, L. Sun (2021). Scalable low-rank tensor learning for spatiotemporal traffic data imputation. *Transportation Research Part C: Emerging Technologies*, 129: 103226.
  - [**C1**] **X. Chen**, M. Lei, N. Saunier, L. Sun (2021). Low-rank autoregressive tensor completion for spatiotemporal traffic data imputation. *The 7th SIGKDD Workshop on Mining and Learning from Time Series (MiLeTS)* at KDD 2021.

- **Preprint** (under review)
  - [**P1**] **X. Chen**, C. Zhang, X.L. Zhao, N. Saunier, L. Sun (2022). Nonstationary temporal matrix factorization for multivariate time series forecasting. arXiv preprint arXiv:2203.10651.

- **Open-source projects**
  - **transdim**: Machine learning for spatiotemporal traffic data imputation and forecasting. (780 stars & 240 forks on GitHub)
    https://github.com/xinychen/transdim
  - **tracebase**: Multivariate time series forecasting on high-dimensional and sparse Uber movement speed data. (20 stars & 5 forks on GitHub)
    https://github.com/xinychen/tracebase

Thanks for your attention!

Any Questions?

About me:

🏛 Homepage: https://xinychen.github.io

🧑‍💻 GitHub: https://github.com/xinychen (**2.4K+ stars**)

✍ Blog: https://medium.com/@xinyu.chen (**30K+ views**)

💌 How to reach me: chenxy346@gmail.com