POLYTECHNIQUE MONTRÉAL
UNIVERSITÉ D'INGÉNIERIE

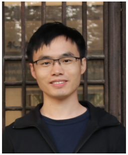
IVADO

Spatiotemporal Traffic Data Imputation and Forecasting with Tensor Learning

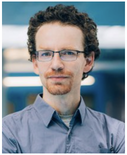
Ph.D. Research Project

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

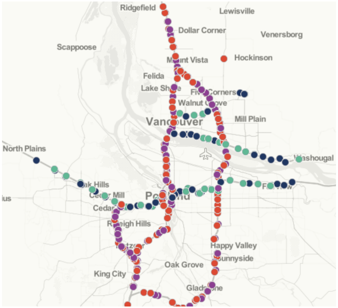
Outline

- **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
- **Objectives**
- **Methodology**
  - Spatiotemporal traffic data imputation
  - High-dimensional traffic forecasting
  - Multidimensional traffic forecasting
  - Traffic forecasting on sparse data
- **Conclusion**

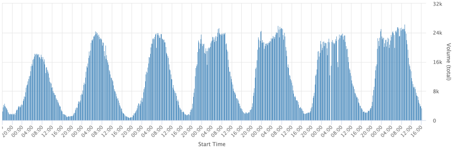
Multivariate Traffic Time Series

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

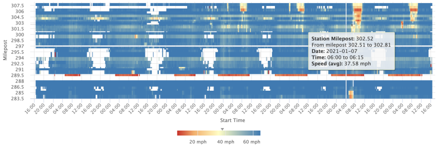
- **Example**: Portland highway traffic data[1].



Detector locations



Traffic volume



Traffic speed

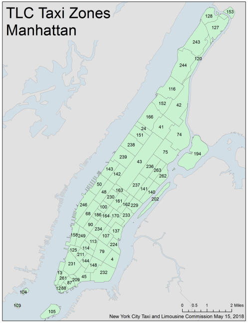
- $\boldsymbol{X} \in \mathbb{R}^{N \times T}$ with N spatial locations $\times$ T time steps

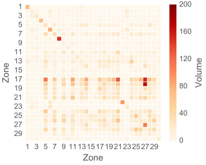
[1] https://portal.its.pdx.edu/home

Multidimensional Traffic Time Series

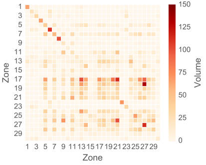
Many spatiotemporal traffic time series data are in the form of **tensor**.

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

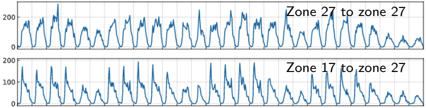




Time t



Time $t = 1$



- $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{M \times N \times T}$ 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.

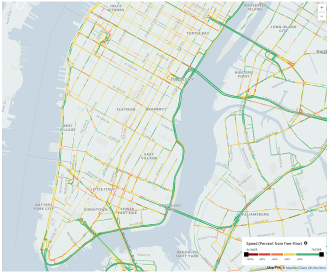
- Incompleteness & sparsity
- High-dimensionality
- Multidimensionality
- Noises & outliers
- Nonstationarity
- ......

In addition, spatiotemporal correlations are also very important.

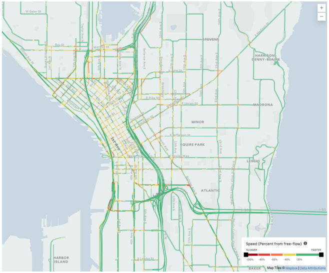
Multiple Data Behaviors

Sparsity & high-dimensionality

- Uber (hourly) movement speed data[3]



NYC movement



Seattle movement

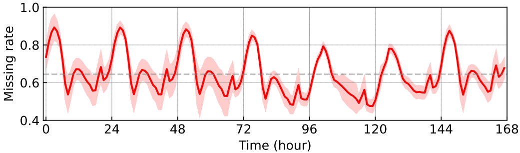
- The average speed on a given road segment for each hour of each day.
- Hourly speed data are computed when road segments have 5+ unique trips.
- **Issue**: insufficient sampling of ridesharing vehicles on the road network.

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

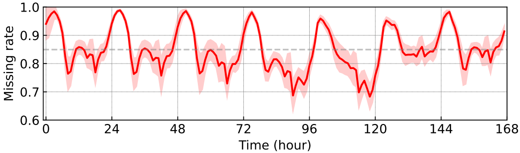
Multiple Data Behaviors

Sparsity & high dimensionality

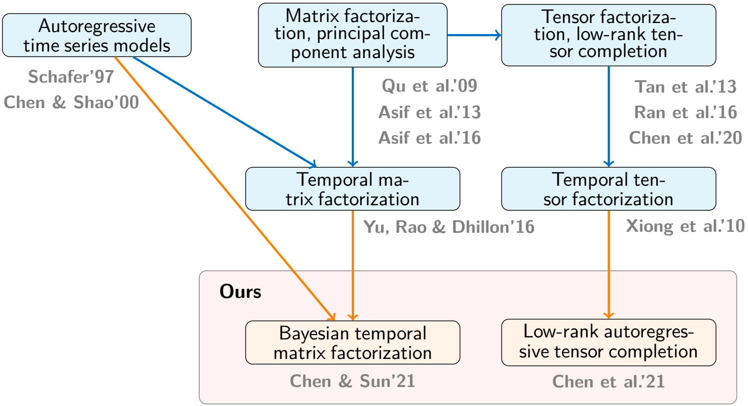
- **NYC** movement speed data (2019)
  - **98,210** road segments & 8,760 time steps (hours)
  - Overall missing rate: **64.43%**



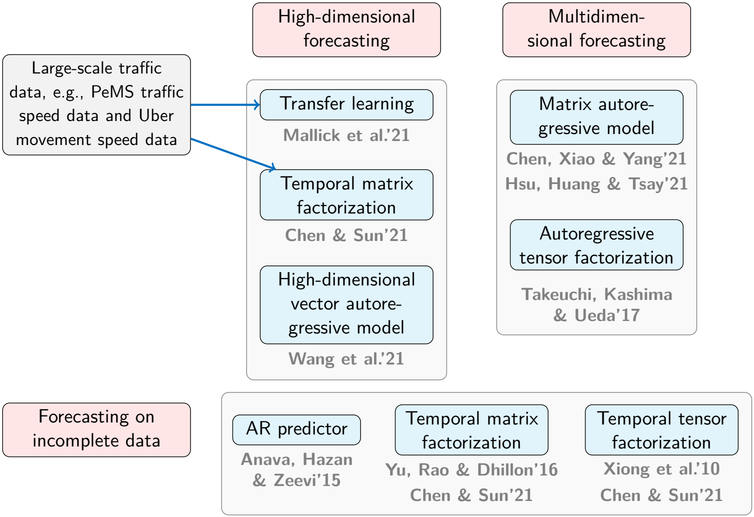
- **Seattle** movement speed data (2019)
  - **63,490** road segments & 8,760 time steps (hours)
  - Overall missing rate: **84.95%**



Spatiotemporal Traffic Data Imputation



Spatiotemporal Traffic Forecasting

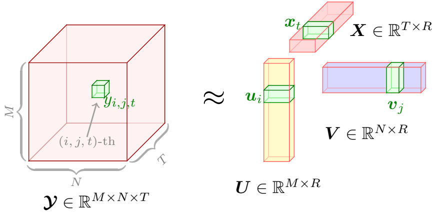


Low-Rank Tensor Learning

From data to model

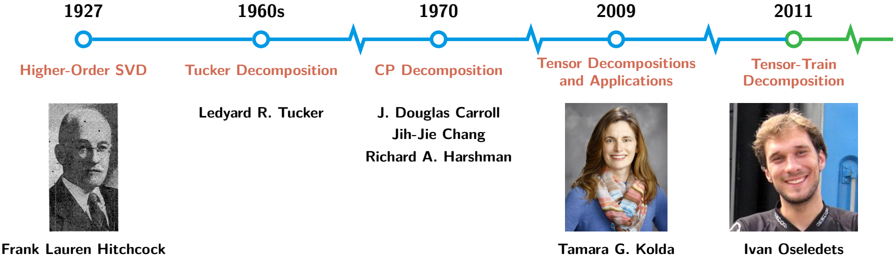
- Matrix factorization:

$$\min_{\boldsymbol{W},\boldsymbol{X}} \frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{Y})-\mathcal{P}_{\Omega}(\boldsymbol{W}\boldsymbol{X})\|_F^2+\frac{\lambda}{2}(\|\boldsymbol{W}\|_F^2+\boldsymbol{X}\|_F^2)$$



- Tensor factorization:

$$\min_{\boldsymbol{U},\boldsymbol{V},\boldsymbol{X}} \frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{\mathcal{Y}})-\mathcal{P}_{\Omega}(\sum_{r=1}^{R}\boldsymbol{u}_r\circ\boldsymbol{v}_r\circ\boldsymbol{x}_r)\|_F^2+\frac{\lambda}{2}(\|\boldsymbol{U}\|_F^2+\|\boldsymbol{V}\|_F^2+\boldsymbol{X}\|_F^2)$$

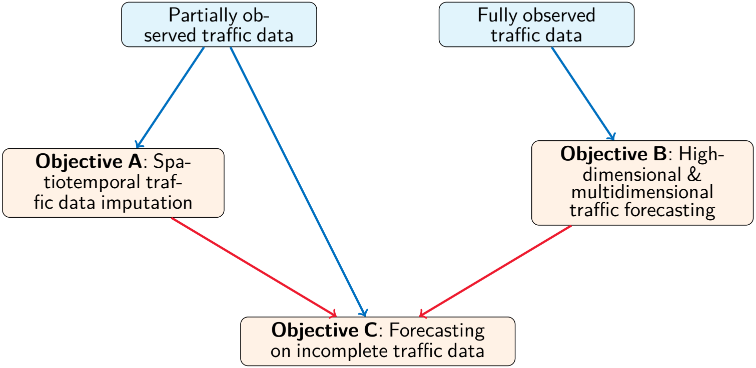

1927
Higher-Order SVD
Frank Lauren Hitchcock
1960s
Tucker Decomposition
Ledyard R. Tucker
1970
CP Decomposition
J. Douglas Carroll
Jih-Jie Chang
Richard A. Harshman
2009
Tensor Decompositions and Applications
Tamara G. Kolda
2011
Tensor-Train Decomposition
Ivan Oseledets

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.

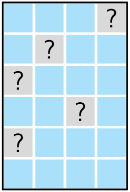
Objectives: A Whole Structure

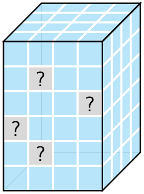
We are working on **spatiotemporal traffic data modeling**.



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 learn and 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 matrix completion

Given the multivariate traffic time series data $\boldsymbol{Y}\in\mathbb{R}^{N\times T}$ with the observed index set Ω, the low-rank matrix completion combines both nuclear norm and univariate autoregressive process:

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

Low-rank autoregressive tensor completion

If T time steps can be separated into I time steps per day and J days, i.e., $T=IJ$, then we can define an operator $\mathcal{Q}(\cdot)$ to convert multivariate data to third-order tensor:

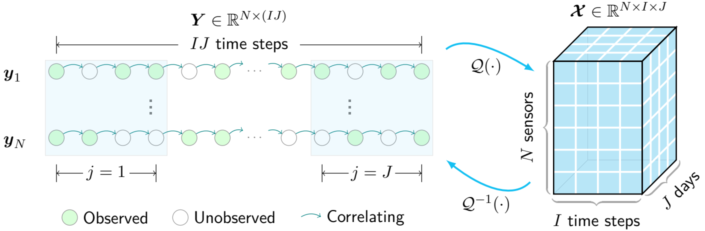
$$\begin{aligned}
&\min_{\boldsymbol{\mathcal{X}}}\ \|\boldsymbol{\mathcal{X}}\|_*+\frac{\lambda}{2}\sum_{n=1}^{N}\sum_{t=d+1}^{T}\Big(z_{n,t}-\sum_{k=1}^{d}a_{n,k}z_{n,t-k}\Big)^2 \\
&\text{s.t.}\ \begin{cases}\boldsymbol{\mathcal{X}}=\mathcal{Q}(\boldsymbol{Z}), \\ \mathcal{P}_{\Omega}(\boldsymbol{Z})=\mathcal{P}_{\Omega}(\boldsymbol{Y}).\end{cases}
\end{aligned} \tag{4}$$

Spatiotemporal Traffic Data Imputation

Low-rank autoregressive tensor completion

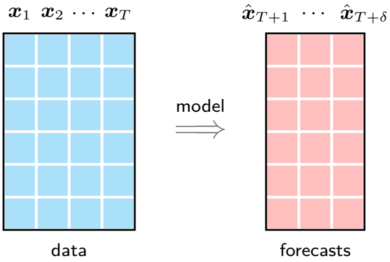
$$\begin{aligned}&\min_{\boldsymbol{\mathcal{X}}}\ \|\boldsymbol{\mathcal{X}}\|_*+\frac{\lambda}{2}\sum_{n=1}^{N}\sum_{t=d+1}^{T}\Big(z_{n,t}-\sum_{k=1}^{d}a_{n,k}z_{n,t-k}\Big)^2\\&\text{s.t. }\begin{cases}\boldsymbol{\mathcal{X}}=\mathcal{Q}(\boldsymbol{Z}),\\ \mathcal{P}_\Omega(\boldsymbol{Z})=\mathcal{P}_\Omega(\boldsymbol{Y}).\end{cases}\end{aligned}\tag{5}$$

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



High-Dimensional Traffic Forecasting

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

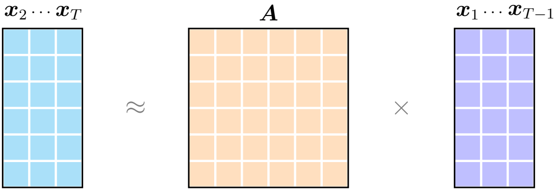


- **Solution:** For time series $\boldsymbol{x}_1, \dots, \boldsymbol{x}_T \in \mathbb{R}^N$, the dth-order vector autoregressive (VAR(d)) model: $\boldsymbol{x}_t = \sum_{k=1}^{d} \boldsymbol{A}_k \boldsymbol{x}_{t-k} + \boldsymbol{\epsilon}_t$.
- **Advantage:** Co-evolution patterns
- **Limitation:**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

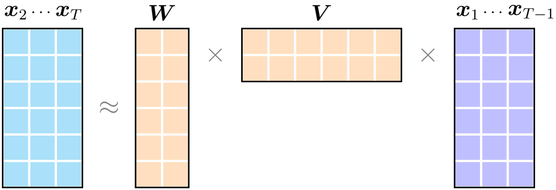
High-Dimensional Traffic Forecasting

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

Multivariate reduced-rank regression

- Data: $\boldsymbol{Z}\in\mathbb{R}^{N\times T}$ (input) & $\boldsymbol{Y}\in\mathbb{R}^{M\times T}$ (output)
- Component matrices $\boldsymbol{W}\in\mathbb{R}^{M\times R},\boldsymbol{V}\in\mathbb{R}^{R\times N}$ such that $\boldsymbol{A}=\boldsymbol{W}\boldsymbol{V}$
- Optimization problem:

$$\boldsymbol{W}^{\star},\boldsymbol{V}^{\star}\triangleq\arg\min_{\boldsymbol{W},\boldsymbol{V}}\frac{1}{2}\|\boldsymbol{Y}-\boldsymbol{W}\boldsymbol{V}\boldsymbol{Z}\|_F^2 \tag{6}$$

Theorem 1 (Izenman'75)

Suppose the optimization problem of multivariate reduced-rank regression trained on the centered data matrices $\boldsymbol{Z}$ and $\boldsymbol{Y}$, then for any positive definite matrix $\boldsymbol{\Gamma}$, the solutions are

$$\boldsymbol{W}^{\star}=\boldsymbol{\Gamma}^{-1/2}\boldsymbol{\xi},\,\boldsymbol{V}^{\star}=\boldsymbol{\xi}^{\top}\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}_{yz}\boldsymbol{\Sigma}_{zz}^{-1}, \tag{7}$$

where $\boldsymbol{\xi}\in\mathbb{R}^{N\times R}$ consists of the R eigenvectors corresponding to the R largest eigenvalues of the matrix $\boldsymbol{\Gamma}^{1/2}\boldsymbol{\Sigma}_{yz}\boldsymbol{\Sigma}_{zz}^{-1}\boldsymbol{\Sigma}_{zy}\boldsymbol{\Gamma}^{1/2}$.

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 $\boldsymbol{A}_k \in \mathbb{R}^{N\times N}, k = 1, \dots, d$ 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 $\boldsymbol{x}_t \in \mathbb{R}^N, t = 1, \dots, T$, the reduced-rank VAR via Tucker decomposition is given by

$$\min_{\boldsymbol{\mathcal{G}},\boldsymbol{U}_1,\boldsymbol{U}_2,\boldsymbol{U}_3} \frac{1}{2}\sum_{t=d+1}^{T} \|\boldsymbol{x}_t - (\boldsymbol{\mathcal{G}} \times_1 \boldsymbol{U}_1 \times_2 \boldsymbol{U}_2 \times_3 \boldsymbol{U}_3)_{(1)}\boldsymbol{z}_t\|_2^2 \tag{8}$$

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 $\boldsymbol{U}_1 \in \mathbb{R}^{N\times R_1}$, $\boldsymbol{U}_2 \in \mathbb{R}^{N\times R_2}$, and $\boldsymbol{U}_3 \in \mathbb{R}^{d\times R_3}$ are the component matrices.

Advantage: High compression rate.

Limitations: Nonconvex optimization; not scalable to large problems.

High-Dimensional Traffic Forecasting

High-dimensional VAR(d)

Given the (high-dimensional) multivariate time series $\boldsymbol{x}_t \in \mathbb{R}^N, t = 1, \ldots, T$, for the data $\boldsymbol{x}_t \in \mathbb{R}^N, \boldsymbol{z}_t \in \mathbb{R}^{dN}, t = d+1, \ldots, T$, we have a lower-dimensional problem:

$$\min_{\{\boldsymbol{U},\boldsymbol{W},\boldsymbol{V} \mid \boldsymbol{U}^\top\boldsymbol{U}=\boldsymbol{V}^\top\boldsymbol{V}=\boldsymbol{I}_R\}} \frac{1}{2}\sum_{t=d+1}^{T} \|\boldsymbol{x}_t - \boldsymbol{U}\boldsymbol{W}\boldsymbol{V}^\top\boldsymbol{z}_t\|_2^2 \tag{9}$$

or equivalently, we have a canonical form:

$$\min_{\{\boldsymbol{U},\boldsymbol{W},\boldsymbol{V} \mid \boldsymbol{U}^\top\boldsymbol{U}=\boldsymbol{V}^\top\boldsymbol{V}=\boldsymbol{I}_R\}} \frac{1}{2}\sum_{t=d+1}^{T} \|\boldsymbol{U}^\top\boldsymbol{x}_t - \boldsymbol{W}\boldsymbol{V}^\top\boldsymbol{z}_t\|_2^2 \tag{10}$$

where $\boldsymbol{W} \in \mathbb{R}^{R\times R}$ is the coefficient matrix.

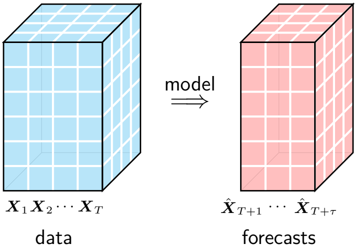
VARMA(d, q)

$$\boldsymbol{x}_t = \sum_{k=1}^{d} \boldsymbol{A}_k\boldsymbol{x}_{t-k} + \boldsymbol{\epsilon}_t + \sum_{p=1}^{q} \boldsymbol{B}_p\boldsymbol{\epsilon}_{t-p} \tag{11}$$

Moving average (MA) processes smooth out the data noises.

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 $\boldsymbol{X}_{T+\tau}, \tau \in \mathbb{N}_+$.



[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

$$\boldsymbol{X}_t = \sum_{k=1}^{d} \boldsymbol{A}_k \boldsymbol{X}_{t-k} \boldsymbol{B}_k^\top + \boldsymbol{E}_t \tag{12}$$

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: Not scalable to large problems.

Multidimensional Traffic Forecasting

Borrowing the idea of Sylvester equation.

Sylvester-type MAR(d)

Given matrix-variate time series $\boldsymbol{X}_t \in \mathbb{R}^{M\times N}, t=1,\ldots,T$, then we define the MAR(d) as follows,

$$\boldsymbol{X}_t = \sum_{k=1}^{d}\left(\boldsymbol{A}_k\boldsymbol{X}_{t-k} + \boldsymbol{X}_{t-k}\boldsymbol{B}_k^\top\right) + \boldsymbol{E}_t \tag{13}$$

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.

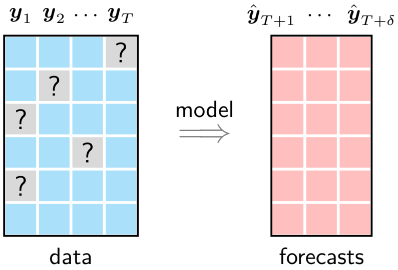
Advantage:

- Instead of the bilinear form of MAR(d), solving Sylvester-type MAR(d) is computationally economic.

Traffic Forecasting on Sparse Data

Multivariate traffic time series

- **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?

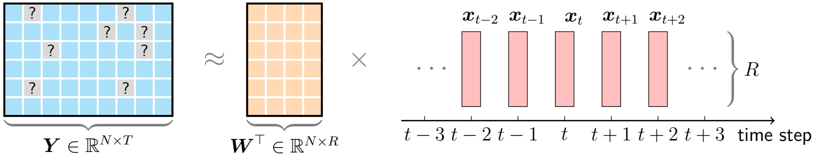
Traffic Forecasting on Sparse Data

Multivariate traffic time series

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:

$$\begin{aligned}\min_{\boldsymbol{W},\boldsymbol{X},\{\boldsymbol{A}_k\}_{k=1}^{d}} \quad & \frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^\top\boldsymbol{X})\|_F^2+\frac{\rho}{2}(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2) \\ & +\frac{\lambda}{2}\sum_{t=d+1}^{T}\|\boldsymbol{x}_t-\sum_{k=1}^{d}\boldsymbol{A}_k\boldsymbol{x}_{t-k}\|_2^2\end{aligned} \tag{14}$$



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

Traffic Forecasting on Sparse Data

Multivariate traffic time series

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:

$$\begin{aligned}\min_{\boldsymbol{W},\boldsymbol{X},\{\boldsymbol{A}_k\}_{k=1}^{d}}\;&\frac{1}{2}\|\mathcal{P}_{\Omega}(\boldsymbol{Y}-\boldsymbol{W}^\top\boldsymbol{X})\|_F^2+\frac{\rho}{2}(\|\boldsymbol{W}\|_F^2+\|\boldsymbol{X}\|_F^2)\\&+\frac{\lambda}{2}\sum_{t=d+m+1}^{T}\|(\boldsymbol{x}_t-\boldsymbol{x}_{t-m})-\sum_{k=1}^{d}\boldsymbol{A}_k(\boldsymbol{x}_{t-k}-\boldsymbol{x}_{t-m-k})\|_2^2\end{aligned}\tag{15}$$

- 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 $\boldsymbol{x}_t''=(\boldsymbol{x}_t-\boldsymbol{x}_{t-m})-(\boldsymbol{x}_{t-1}-\boldsymbol{x}_{t-m-1})$.

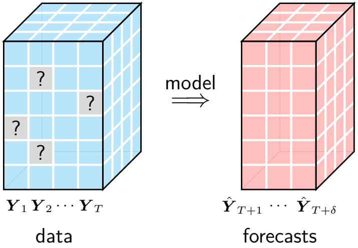
🙂 Stationarizing a time series with differencing can improve the prediction.[4]

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

Traffic Forecasting on Sparse Data

Multidimensional traffic time series

- **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., CP factorization + VAR (on latent temporal factors).

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.

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. *IEEE Transactions on Knowledge and Data Engineering*.
- **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. (19 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