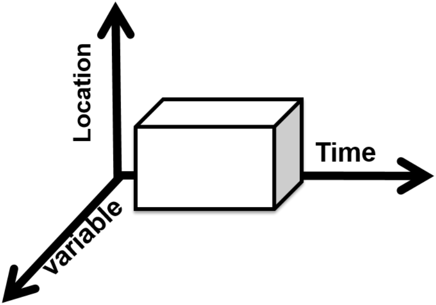
Accelerated Online Low-Rank Tensor Learning for Multivariate Spatio-Temporal Streams

Rose Yu, Dehua Cheng, Yan Liu

University of Southern California

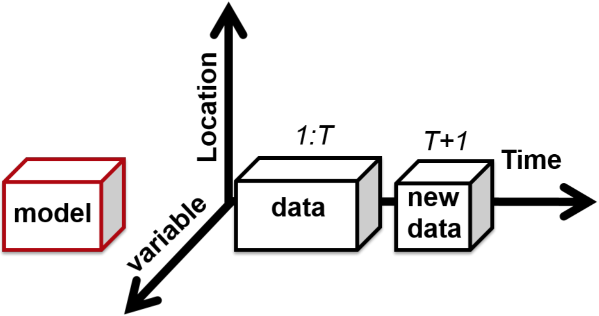
Introduction



- Multivariate spatio-temporal data can be represented as tensors.
- Low-rank structure corresponds to properties such as spatial clustering, temporal periodicity, etc.
- Low-rank tensor learning framework for the multivariate spatio-temporal analysis.

Introduction

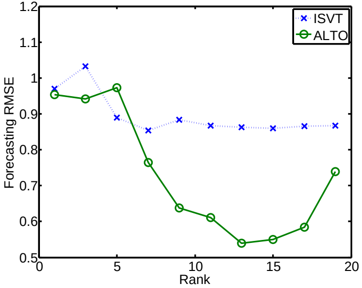
- Large-scale spatio-temporal data come in streams.
- Learning low-rank tensor in batch settings suffers from computational bottleneck, especially short response time.
- Goal: Online Low-Rank Tensor Learning
    - Efficiently update model tensor as data come in.
    - Preserve the low-rank structure.



Challenges

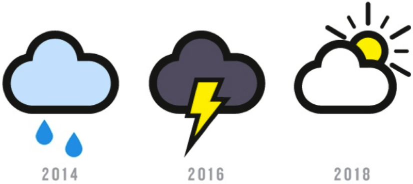
Inherent complexity of tensor analysis[HL13]:

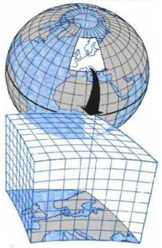
- Most works are on online low-rank matrix learning.
- Local solution (e.g. streaming tensor analysis [STP$^+$08]) lacks theoretical justification.
- Using nuclear norm as a convex surrogate for the rank (e.g. Stochastic ADMM [OHTG13]) is computationally expensive, may lead to sub-optimal solutions.



Contribution

- **A**ccelerated **L**ow-rank **T**ensor **O**nline Learning (ALTO) algorithm.
  - Acceleration by keeping track of the low-rank components.
  - Overcome the local optima issue via randomization.
- Applications in two spatio-temporal stream analysis tasks.
  - **Online forecasting**: n-step ahead prediction from historical observations.
  - **Multi-model ensemble**: combining multiple simulation model forecasts for more accurate predictions.





Online Low Rank Tensor Learning

Low-rank tensor learning problem for regression

- Predictor tensor $\mathcal{Z} \in \mathbb{R}^{Q \times T \times M}$.
- Response tensor $\mathcal{X} \in \mathbb{R}^{P \times T \times M}$.
- Model tensor $\mathcal{W} \in \mathbb{R}^{P \times Q \times M}$.

Problem Definition

$$\widehat{\mathcal{W}} = \operatorname{argmin}_{\mathcal{W}} \left\{ \sum_{t,m} \|\mathcal{W}_{:,:,m} \mathcal{Z}_{:,t,m} - \mathcal{X}_{:,t,m}\|_F^2 \right\}$$
$$\text{s.t.} \quad \text{rank}(\mathcal{W}) \leq R$$

- Two Stage Procedure
  1. Solve unconstrained optimization problem.
  2. Update solution to satisfy low-rank constraint.

Step 1: Tensor Stream in Online Setting

At time T, given a new data batch of size b.
Denote $\mathbf{W}_m = \mathcal{W}_{:,:,m}$, omit the variable index m for simplicity.

$$\widehat{\mathcal{W}} = \underset{\mathcal{W}}{\operatorname{argmin}} \left\{ \sum_{t,m} \|\mathcal{W}_{:,:,m}\mathcal{Z}_{:,t,m} - \mathcal{X}_{:,t,m}\|_F^2 \right\}$$

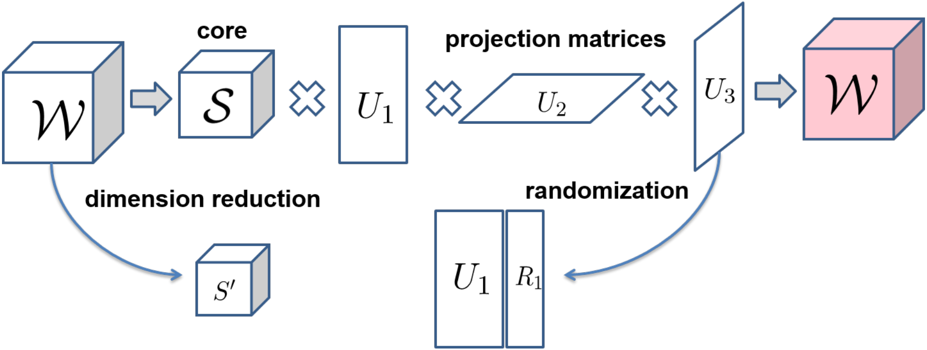
$$\Downarrow$$

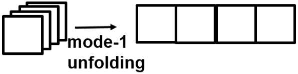
$$\min_{\mathbf{W}} \|\mathbf{W}\mathbf{Z}_{1:T} - \mathbf{X}_{1:T}\|_{\mathrm{F}}^2$$

- An ordinary linear regression problem.
- Two possible updating strategies.
    - Exact update:
    $$\mathbf{W}^{(k)} = \mathbf{X}_{1:T+b}\mathbf{Z}_{1:T+b}^{\dagger}.$$
    - Increment update:
    $$\mathbf{W}^{(k)} = (1-\alpha)\mathbf{W}^{(k-1)} + \alpha\mathbf{X}_{T+1:T+b}\mathbf{Z}_{T+1:T+b}^{\dagger}.$$

Step 2: Online Low-Rank Tensor Approximation

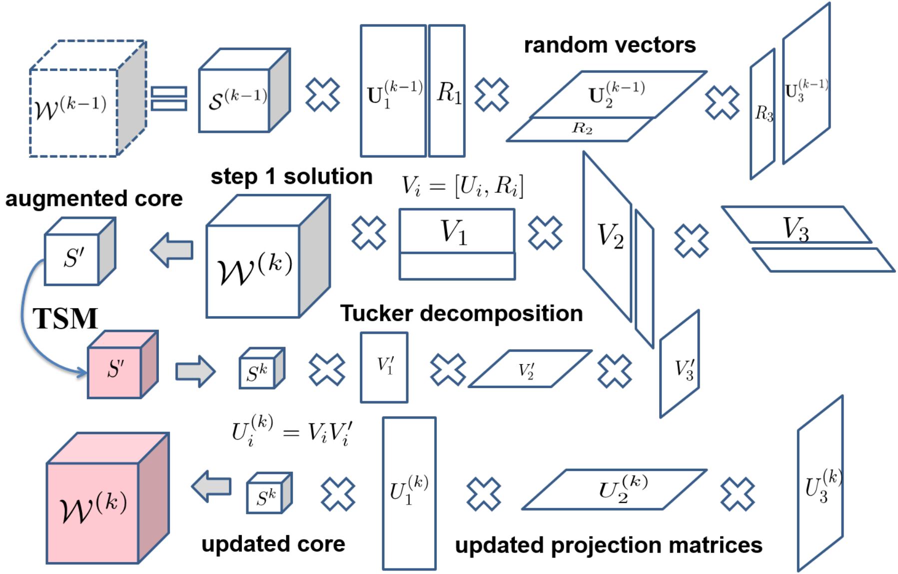
- Update the solution by low-rank projection.
- Perform low-rank projection at each iteration is computationally expensive.





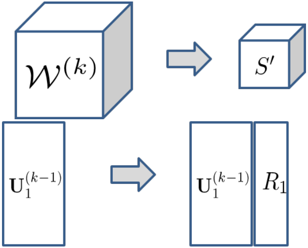
- Tensor sum-n-rank: $\sum_{n=1}^{N} \text{rank}(\mathcal{W}_{(n)})$

Step 2: Online Low-Rank Tensor Approximation



Theoretical Analysis

- Dimension reduction based on previous decomposition.
- Jumping out of the same low-rank subspace with randomization.



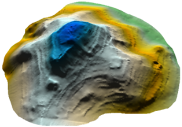
- **Low-Rank Guarantee**: The solution is guaranteed to be low-sum-n-rank after the tensor sequential mapping (TSM) procedure.
- **Approximation Guarantee**: When the target tensor is low-rank, then in its neighborhood, we can conduct low-rank mapping and expect the error to be reduced. The approximation error of the low-rank mapping is upper bounded by a factor of 8 in the worst case scenario.

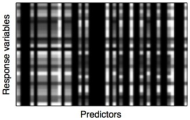
Applications I: Online Forecasting

Description: Predict the value of $(\mathcal{X}_{p,t+1,m}, \mathcal{X}_{p,t+1,m}, \cdots ,)$ for all M variables and P locations given their historical measurements.
Formulation: Vector Auto-regressive (VAR) model:
$\mathcal{X}_{:,t,m} = \mathcal{W}_{:,:,m}\mathbf{X}_{t,m} + \mathcal{E}_{:,t,m}$, where $\mathbf{X}_{t,m} = [\mathcal{X}^\top_{:,t-1,m}, \ldots, \mathcal{X}^\top_{:,t-L,m}]^\top$

$$\widehat{\mathcal{W}} = \underset{\mathcal{W}}{\operatorname{argmin}} \left\{ \|\widehat{\mathcal{X}} - \mathcal{X}\|_F^2 + \mu \sum_{m=1}^{M} \mathrm{tr}(\widehat{\mathcal{X}}^\top_{:,:,m} \mathbf{L} \widehat{\mathcal{X}}_{:,:,m}) \right\}$$

$$\text{s.t.} \quad \widehat{\mathcal{X}}_{:,t,m} = \mathcal{W}_{:,:,m}\mathbf{X}_{t,m} \qquad \sum_{n=1}^{N} \mathrm{rank}(\mathcal{W}_{(n)}) \leq R,$$





Applications II: Multimodel Ensemble

Description: Combine S climate model outputs of M climate variables in P locations over time period T into a more accurate description of the observations.
Formulation: : Linear model between observations and model outputs. $\mathbf{Y}_{t,m} = [\mathcal{Y}_{:,t,m,1}^\top, \ldots, \mathcal{Y}_{:,t,m,S}^\top]^\top$ denotes the concatenation of S model outputs at time t for variable m, $\mathcal{Y} \in \mathbb{R}^{P \times T \times M \times S}$.

$$\widehat{\mathcal{W}} = \underset{\mathcal{W}}{\operatorname{argmin}} \left\{ \|\widehat{\mathcal{X}} - \mathcal{X}\|_F^2 + \mu \sum_{m=1}^{M} \mathsf{tr}(\widehat{\mathcal{X}}_{:,:,m}^\top \mathbf{L} \widehat{\mathcal{X}}_{:,:,m}) \right\}$$

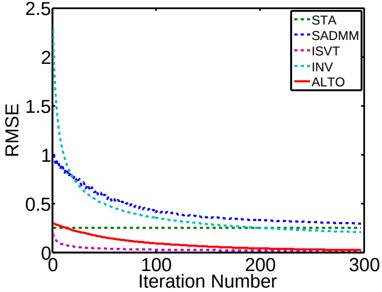
$$\text{s.t.} \quad \widehat{\mathcal{X}}_{:,t,m} = \mathcal{W}_{:,:,m} \mathbf{Y}_{t,m} \qquad \sum_{n=1}^{N} \mathsf{rank}(\mathcal{W}_{(n)}) \leq R,$$

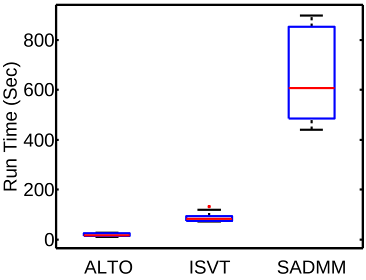
Synthetic Experiments

Baselines:

- simple VAR model (INV),
- streaming tensor analysis [STP+08] (STA)
- stochastic ADMM [OHTG13] (SADMM),
- iterative singular value thresholding [JMD10] (ISVT),
- greedy algorithm [BYL14] (GREEDY).

Setting: 30000 time stamps with VAR(2) model, parameter tensor $\mathcal{W} \in \mathbb{R}^{30\times 60\times 20}$. Initial batch size 200, mini-batch size 100.





Online Forecasting

Datasets:

- *Foursquare:* 121 user check-ins in 15 categories of business venues over 1200 time intervals.

- *AWS:* 153 weather stations measurements of 4 climate variables over 76 time stamps.

Setting: 90 % training data on both datasets for VAR model with different lags and average run time.

Forecasting RMSE

Lag	ALTO	ISVT	SADMM	INV	Greedy
		Foursquare			
1	0.1239	0.1285	0.1240	0.1394	0.1246
2	0.1244	0.1244	0.1234	0.1357	0.1225
3	0.1241	0.1240	0.1242	0.1362	0.1223
		AWS			
1	0.9318	1.0055	0.9441	1.4707	0.8951
2	0.9285	0.9182	0.9447	1.0853	0.9131
3	0.9303	0.9297	0.9485	0.9840	0.9166

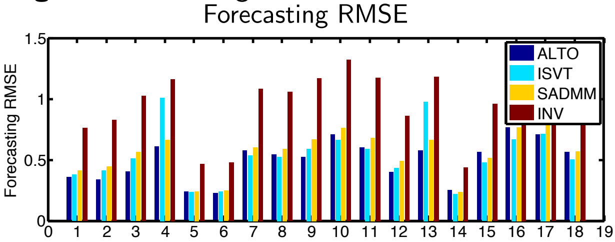
Overall run time

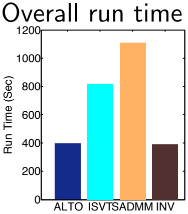
Data set	ALTO	ISVT	SADMM
Foursquare	16 (s)	65 (s)	119 (s)
AWS	20 (s)	64 (s)	126 (s)

Multimodel Ensemble

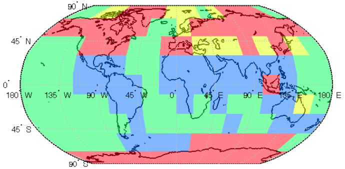
Dataset:7 model outputs, 19 climate variables, 252 time points. variables of observation series are aligned with the model output series.

Settings: 90 % training data, 5-fold cross validation





- Japan Center for Climate System Research (Red) has a dominating area in Asia.
- Norway Bjerknes Centre for Climate Research (Yellow) is most influential in Europe.



Conclusion and Future Work

Conclusion

- A simple and efficient algorithm, ALTO, to accelerate the process of online low-rank tensor learning.
- Randomization technique to overcome the local optimal issue.
- Accurate predictions and reduced computational costs for online forecasting and multi-model ensemble tasks.

Future Work

- Examining broader applications of online low-rank tensor learning
- Relaxing the assumptions of ALTO for better theoretical properties.

References I

Mohammad Taha Bahadori, Qi Rose Yu, and Yan Liu, *Fast multivariate spatio-temporal analysis via low rank tensor learning*, Advances in Neural Information Processing Systems, 2014, pp. 3491–3499.

Christopher J Hillar and Lek-Heng Lim, *Most tensor problems are np-hard*, Journal of the ACM (JACM) **60** (2013), no. 6, 45.

Prateek Jain, Raghu Meka, and Inderjit S Dhillon, *Guaranteed rank minimization via singular value projection*, Advances in Neural Information Processing Systems, 2010, pp. 937–945.

Hua Ouyang, Niao He, Long Tran, and Alexander Gray, *Stochastic alternating direction method of multipliers*, Proceedings of the 30th International Conference on Machine Learning, 2013, pp. 80–88.

Jimeng Sun, Dacheng Tao, Spiros Papadimitriou, Philip S Yu, and Christos Faloutsos, *Incremental tensor analysis: Theory and applications*, ACM Transactions on Knowledge Discovery from Data (TKDD) **2** (2008), no. 3, 11.