Accelerated Low-Rank Tensor Online Learning For Multi-Model Ensemble

Rose Yu, Dehua Cheng, and Yan Liu
University of Southern California
{qiyu,dehua.cheng,yanliu.cs}@usc.edu

UNIVERSITY OF SOUTHERN CALIFORNIA

1. Abstract

Motivation

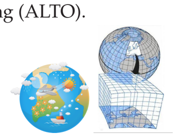
Multivariate spatio-temporal data can be represented as a three-mode tensor. Low-rank tensor corresponds to low complexity model. In climate data analysis, we are confronted with large-scale tensor streams. Batch learning suffers from computational bottleneck.

Goal

Online Low-Rank Tensor Learning: update a model tensor while preserving the low-rank structure.

Solution

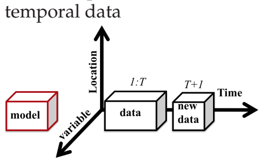
A simple and efficient algorithm: Accelerated Low-rank Online Tensor Learning (ALTO).



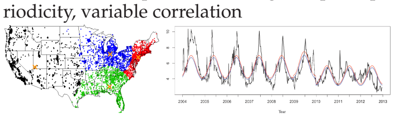
2. Introduction

Background

- Tensor representation of multivariate spatio-temporal data



- Low-rankness: spatial clustering, temporal periodicity, variable correlation

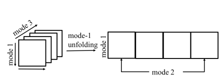


Challenges

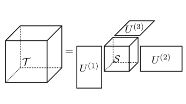
- Inherent complexity of tensor analysis [Hillar 2013].
- Most works are on online low-rank matrix learning. Local solution (e.g. streaming tensor analysis [Sun 2008]) lacks theoretical understandings.
- Using nuclear norm as a convex surrogate for the rank (e.g. Stochastic ADMM [Ouyang 2013]) may lead to sub-optimal solutions.
- Existing multi-model ensemble methods such as supermodeling [Wiegerinck2011] are computationally expensive.

Preliminary

- Tensor Unfolding:



- Tucker Decomposition:



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

3. Methodology

Tensor 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 $\mathcal{W} \in \mathbb{R}^{P\times Q\times M}$

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

Two-Step Procedure

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.

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

$$\Downarrow$$

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

An ordinary linear regression problem, can be updated with two possible strategies:

- Exact update:

$$\mathbf{W}^{(k)} = \mathbf{X}_{1:T+b}\mathbf{Z}^{\dagger}_{1:T+b}.$$

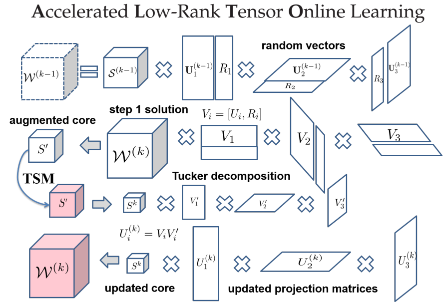
- Increment update:

$$\mathbf{W}^{(k)} = (1-\alpha)\mathbf{W}^{(k-1)} + \alpha\mathbf{X}_{T+1:T+b}\mathbf{Z}^{\dagger}_{T+1:T+b}.$$

2. Online Low-Rank Tensor Approximation

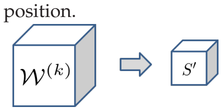
- Update the solution by low-rank projection.
- Perform low-rank projection at each iteration is computationally expensive.

Accelerated Low-Rank Tensor Online Learning

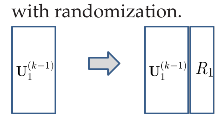


Theoretical Analysis

- Dimension reduction based on previous decomposition.



- Jumping out of the same low-rank subspace with randomization.



4. Application

Description

Multi-model ensemble: combining multiple simulation model forecasts into more accurate predictions.

Design Principal

- Global consistency: the data in the common structure are likely to be similar.
- Local consistency: the data in close neighborhood locations are likely to be similar.

Formulation

$\mathbf{Y}_{t,m} = [\mathcal{Y}^{\top}_{:,t,m,1}, \ldots, \mathcal{Y}^{\top}_{:,t,m,S}]^{\top}$ denotes the concatenation of S model outputs at time t for variable m, $\mathcal{Y} \in \mathbb{R}^{P\times T\times M\times S}$. $\mathcal{X} \in \mathbb{R}^{P\times T\times M}$ denotes observations.

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

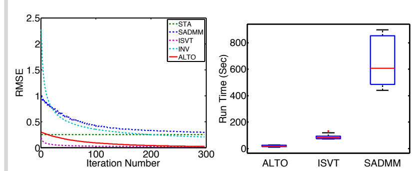
$$\text{s.t.} \ \widehat{\mathcal{X}}_{:,t,m} = \mathcal{W}_{:,:,m}\mathbf{Y}_{t,m}, \ \sum_{n=1}^{N} \text{rank}(\mathcal{W}_{(n)}) \le R$$

where L is the Laplacian matrix constructed from the location information and $\mu, \rho > 0$ are the local and global consistency tradeoff parameters.

5. Experiments

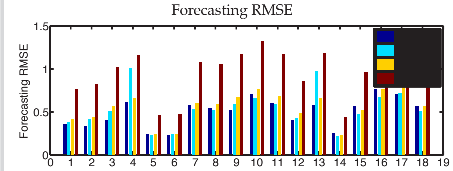
Synthetic Experiments

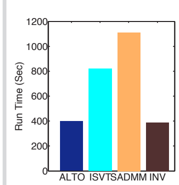
Baselines: simple VAR model (INV), stochastic ADMM [Ouyang 2013] (SADMM), iterative singular value thresholding (ISVT), greedy algorithm [Bahadori 2013] (GREEDY).
Setting: 30000 time stamps generated from VAR(2) model. Parameter tensor $\mathcal{W} \in \mathbb{R}^{30\times 60\times 20}$. Initial batch size 200, mini-batch size 100.

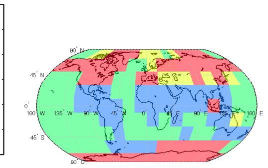


Multi-model Ensemble

Observation: monthly measurements from NCEP-DOE Reanalysis 2. 7 different model outputs: simulation data from the World Climate Research Programme's (WCRP's) CMIP3 multi-model dataset. 19 variables are selected with 252 time points from 1979 to 1999.







Overall run time   Climate models influential areas

References

1 R. Yu[†], M. T. Bahadori[†], and Y. Liu, "Fast Multivariate Spatio-temporal Analysis via Low Rank Tensor Learning," Accepted as spotlight in NIPS 2014. ([†] Equal contributions)
2 Sun, Jimeng and Tao, Dacheng and Papadimitriou, Spiros and Yu, Philip S and Faloutsos, Christos, "Incremental tensor analysis: Theory and applications," TKDD, 2008.
3 Wiegerinck, W and Selten, F, "Supermodeling: Combining imperfect models through learning," NIPS workshop,2011.
4 Ouyang, Hua and He, Niao and Tran, Long and Gray, Alexander, "Stochastic alternating direction method of multipliers," ICML, 2013.
5 Hillar, Christopher J and Lim, Lek-Heng, "Most tensor problems are NP-hard," Journal of the ACM (JACM), 2013.