Multiresolution Tensor Learning for Efficient and Interpretable Spatiotemporal Analysis

Raechel Walker, Rose Yu
University of California, San Diego
La Jolla, CA
r6walker@ucsd.edu, roseyu@eng.ucsd.edu

Abstract

We introduce the Spatiotemporal Multiresolution Tensor Learning (ST-MRTL) model, an extension of MRTL [11] to spatiotemporal data. ST-MRTL offers greater time-efficiency and produces interpretable latent factors by training on both spatial and temporal data in different resolutions. We apply ST-MRTL to sea salinity and temperature data. ST-MRTL predicts the precipitation's variation from the mean in the Midwest region of the United States and generates interpretable latent factors. Additionally, ST-MRTL converges 9-20x faster than the original MRTL solution and depicts historic causes of precipitation in the Midwest, so we can understand patterns in rainfall over time.

1 Introduction

Different aspects of the Earth's climate system vary at different space and time frequencies. For example, the Madden-Julian Oscillation (MJO) and the El Nino-Southern Oscillation (ENSO) increase the chances of an environment receiving above average rainfall within a period of 30-60 days [16], or 4 years [4] respectively. Each of these large sea-atmosphere climate interactions contribute to our understanding of how spatiotemporal frequencies affect rainfall. Since the MRTL model does not train on discrete temporal resolutions, it can not detect the variance of these aforementioned frequencies. The ST-MRTL algorithm can detect these changes in long-term global warming trends that may happen every 100 years and shorter term frequencies (such as MJO and ENSO) because it trains on finer spatial and temporal resolutions as the runtime increases. Spatiotemporal analysis, which is a form of geographical data analysis, discovers and explains patterns in large-scale spatiotemporal data. Oftentimes however, existing spatiotemporal analysis methods that don't train on different temporal and spatial resolutions converge at a slow rate or do not produce interpretable latent factors.

Several methods such as, geographically weighted regression [1], and Gaussian processes [2] are used to study large spatial areas, but have difficulty with scaling to high-resolutions. Past research demonstrated that multiresolution methods, such as latent factor models [10, 7], multiresolution matrix factorization, and deep learning [13, 14] converge at a much faster rate than fixed-resolution models [11]. The Multiresolution Tensor Learning (MRTL) algorithm more accurately predicts spatial events, and are not as computationally expensive as fixed-resolution models [11]. However, MRTL's design is unable to detect the variance of these temporal frequencies, because it does not train on different temporal resolutions.

We present the Spatiotemporal Multiresolution Tensor Learning (ST-MRTL) model. The ST-MRTL algorithm improves upon MRTL because it trains on datasets in different spatiotemporal resolutions. We will compare ST-MRTL to MRTL throughout this paper, in order to explain the purpose of incorporating multiresolution spatiotemporal tensors for climate prediction. In this research, we evaluate the ST-MRTL model on climate data and discover that: (1) ST-MRTL can be applied to

Submitted to the AI for Earth Sciences Workshop at NeurIPS 2020. Do not distribute.

Algorithm 1 Multiresolution Tensor Learning:

Input: initialization W_0, data X, Y.
Output: latent factors $F^{(r)}$
 # full rank tensor model
for each resolution $s \in 1, \ldots, s_0$ **do**
  for each resolution $r \in 1, \ldots, r_0$ **do**
    Initialize $t \leftarrow 0$
    Get a mini-batch B from training set
    while stopping criterion not true **do**
      $t \leftarrow t + 1$
      $W_{t+1}^{(s,r)} \leftarrow \texttt{Opt}\left(W_t^{(s,r+1)} \mid B \right)$
    end while
    $W^{(s,r+1)} = \texttt{Finegrain}\left(W^{(s,r)} \right)$
  end for
  $W^{(s+1,r_1)} = \texttt{Finegrain}\left(W^{(s,r)} \right)$
end for
 # tensor decomposition
$F^{(s_0,r_0)} \leftarrow W^{(s_0,r_0)}$
 # low rank tensor model
for each resolution $s \in s_0, \ldots, S$ **do**
  for each resolution $r \in r_0, \ldots, R$ **do**
    Initialize $t \leftarrow 0$
    Get a mini-batch B from training set
    while stopping criterion not true **do**
      $t \leftarrow t + 1$
      $F_{t+1}^{(s,r)} \leftarrow \texttt{Opt}\left(F_t^{(s,r)} \mid B \right)$
    end while
    $F^{(s,r+1)} = \texttt{Finegrain}\left(F^{(s,r)} \right)$
  end for
  $F^{(s+1,r_1)} = \texttt{Finegrain}\left(F^{(s,r)} \right)$
end for

large-scale datasets, (2) ST-MRTL converges faster than the original MRTL model, and (3) ST-MRTL is able to produce interpretable and accurate spatial and temporal latent factors.

2 Multiresolution Tensor Learning for Spatiotemporal Data

Full Rank Tensor Model Initially, ST-MRTL consists of a full-rank tensor model that can be applied to data with spatial and temporal features. The number of temporal features are s = 1,..., S resolutions with corresponding dimensions $S_1, ..., S_R$. The number of spatial features are r = 1,..., R resolutions with corresponding dimensions $D_1, ..., D_R$.The weight tensor, $\mathcal{W}^{(r,s)} \in \mathbb{R}^{I,T_s,F,D_r}$, is parameterized by tensor learning. i is the number of outputs and F is the number of non-spatial features. The form of the input data is $X^{(r,s)} \in \mathbb{R}^{I,T_s,F,D_r}$. The input features and the model depend on the resolution. At resolution r and s, the full rank tensor learning model can be written as:

$$y_i = a\left(\sum_{t=1}^{T_s} \sum_{f=1}^{F} \sum_{d=1}^{D_r} \mathcal{W}_{i,t,f,d}^{(r,s)} \mathcal{X}_{i,t,f,d}^{(r,s)} + b_i \right), \tag{1}$$

where a is the identity activation function, and b_i is the bias for the output i.

Low Rank Tensor Model A low-rank tensor model is created by using CANDE-COMP/PARAFAC (CP) decomposition. Low-rank models contain less variables, which remove noise in the results. CANDE-COMP/PARAFAC (CP) decomposition decomposes the weight tensor, W, which is the gradient. The sum of rank-1 tensors represents W [5]. The weight tensor factorizes into multiple

factor matrices:

$$\mathcal{W}_{i,f,d}^{(r,s)} = \sum_{k=1}^{K} A_{i,k} B_{f,k} C_{d,k}^{(r,s)} \tag{2}$$

$$y_i = a\left(\sum_{f=1}^{F} \sum_{t=1}^{T_s} \sum_{d=1}^{D_r} \sum_{k=1}^{K} A_{i,k} B_{f,k} C_{d,k}^{(r,s)} \mathcal{X}_{i,t,f,d}^{(r,s)} + b_i \right) \tag{3}$$

where the columns of A, B, $C^{(r,s)}$ are latent factors for each mode of $\mathcal{W}$ and $C^{(r,s)}$ depend on the resolution.

3 Experiments

To demonstrate the potential of ST-MRTL, we predict precipitation in the U.S. Midwest. Further explanation of the datasets, implementation details and finegraining criteria are in the Appendix.

Tensor Regression on Climate Data
We produce Eqn. (4) when Eqn. (1) is applied to the climate dataset. We train ST-MRTL on sea surface salinity (SSS) and sea surface temperature (SST) data in order to predict the variability in rainfall in the Midwest by using the previous 12 months. Additionally, we are using the previous 1 year, 4 seasons, or 12 months to predict the variability in rainfall. Stochastic gradient descent and tensor linear regression train the ST-MRTL model. Let X be the historical oceanic data, which contains SSS and SST across D_r locations and T_s time resolutions. The different resolutions for time consist of year, seasons, and months. The weight tensor, W, is the gradient. F is number of non-spatial features.

$$\mathcal{Y} = \sum_{t=1}^{T_s} \sum_{f=1}^{F} \sum_{d=1}^{D_r} \mathcal{W}_{t,f,d}^{(r,s)} \mathcal{X}_{t,f,d}^{(r,s)} + b \tag{4}$$

4 Results

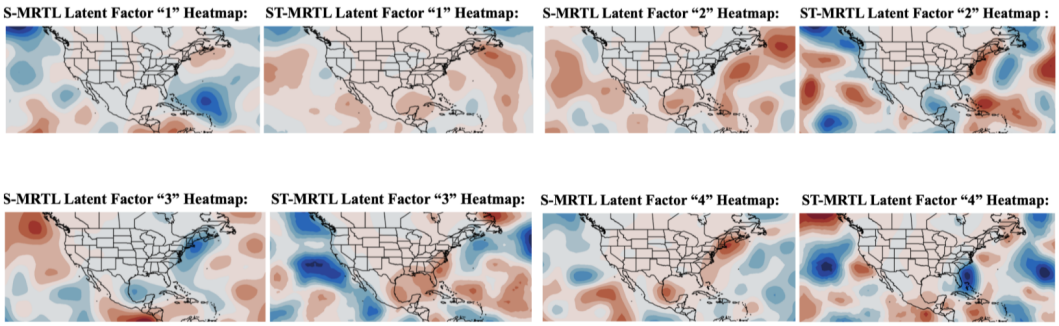

Figure 1: This latent factor heat map depicts the sea surface latent factors. The red and blue areas in the northwest Atlantic region (east of North America and Gulf of Mexico) indicate the soil moisture feedback that contributes to precipitation in the U.S. Midwest.

Accuracy and Convergence We evaluate ST-MRTL with validation loss, training loss, and runtime. Fig. 2 displays the logistic regression curve of the training loss over time of the MRTL model vs. the ST-MRTL model. We utilize ten trials to generate the results. Table 1 displays that the ST-MRTL full-rank model converges approximately 9-20x faster than the MRTL full-rank model, and that the ST-MRTL low-rank model converges approximately 9-13x faster than the MRTL low-rank model. Additionally, the low-rank ST-MRTL model produces more accurate results than the full-rank ST-MRTL model. Since the low-rank ST-MRTL model's validation loss converges to a smaller value

Table 1: Validation Loss and Runtime of S. MRTL models vs. S.T. MRTL models

Model	Validation Loss	Validation Loss Time(s)
S.T. Full	0.0453 ±0.0036	**0.0730±0.0104**
S.T. Low	0.0414±0.0023	**0.4591 ±0.1048**
S. Full	**0.0315±0.0004**	1.2431±0.3788
S. Low	**0.0270±0.0003**	5.3702±2.0393

than the full-rank ST-MRTL model's validation loss, this proves the importance of a full-rank model initializing the low-rank model. The ST-MRTL model's validation loss is 1.3-1.6x greater than the original MRTL algorithm. The ST-MRTL's model's accuracy is relevant because the percentage difference between the ST-MRTL loss and the MRTL loss is minor.

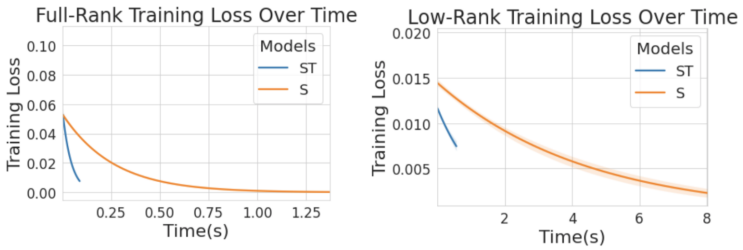

Figure 2: Logistic Regression Curve of Training Loss Over Time of S. MRTL vs. S.T. MRTL for the Full-Rank(left) and Low-Rank(right) case based on 10 Trials

Interpretability Interpretability is one aspect of evaluating ST-MRTL. By training the full-rank model before the low-rank model, we produce spatial and temporal latent factors that are more understandable. Fig. 3 is a series of line graphs that depict a temporal latent factor. The line plots for the monthly and seasonal resolutions tend to increase and decrease during similar time periods.
Fig. 1 contains heatmaps that compare the MRTL vs ST-MRTL model's spatial latent factors to each other. We normalize the latent factors in Fig. 1 to (-1,1). The red portion represents positive values, and the blue portion corresponds to negative values. The red and blue areas near the Gulf of Mexico and the Northwest Atlantic ocean in Fig. 1 prove that the ST-MRTL and S-MRTL predictions align with the findings from [8, 9] because these aforementioned regions influence the precipitation in the Midwest due to the soil moisture feedback from these areas. Both the MRTL and ST-MRTL algorithms are equally as interpretable because aggregations of colors are present in different areas on the heat map.

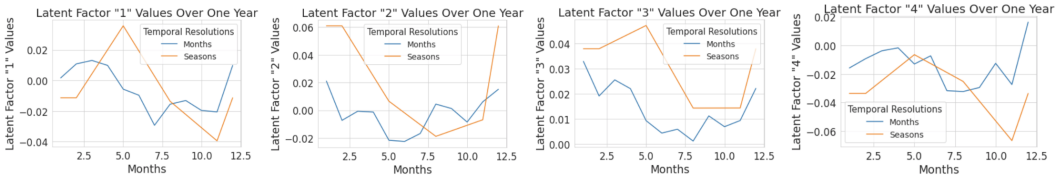

Figure 3: Line plots for the temporal latent factor, A

5 Conclusion and Future Work

The ST-MRTL model's ability to accurately predict the variation in rainfall further proved that this model is less computationally expensive than MRTL and can produce explainable latent factors. In the future, we hope to 1) make predictions in real-time so we can apply it to real-time data, 2) incorporate additional metrics to improve predictions, such as using climatology and persistence, 3) apply ST-MRTL to additional spatiotemporal datasets, 4) experiment with different stopping criteria and patience factors, so we increase the accuracy of ST-MRTL.

References

[1] Brunsdon, C., Fotheringham, S., and Charlton, M. Geo- graphically weighted regression. Journal of the Royal Statistical Society: Series D (The Statistician), 47(3): 431–443, 1998.

[2] Cressie, N. Statistics for spatial data. Terra Nova, 4(5): 613–617, 1992.

[3] Good, S., Martin, M., and Rayner, N. En4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. Journal of Geophysical Research: Oceans, 118:6704–6716, 2013. Version EN4.2.1, https://www.metoffice.
gov.uk/hadobs/en4/download-en4-2-1.html, accessed 06/23/19.

[4] Huang, J., Higuchi, K., Shabbar, A. (1998). The relationship between the North Atlantic Oscillation and El Niño-Southern Oscillation. Geophysical Research Letters, 25(14), 2707-2710.

[5] Hitchcock, F. L. The expression of a tensor or a polyadic as a sum of products. Journal of Mathematics and Physics, 6(1-4):164–189, 1927.

[6] Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[7] Kondor, R., Teneva, N., and Garg, V. Multiresolution matrix factorization. In Proceedings of the 31st International Conference on Machine Learning (ICML-14), pp. 1620– 1628, 2014.

[8] Li, L., Schmitt, R., Ummenhofer, C., and Karnauskas, K. Im- plications of north atlantic sea surface salinity for summer precipitation over the us midwest: Mechanisms and pre- dictive value. Journal of Climate, 29:3143–3159, 2016a.

[9] Li, L., Schmitt, R. W., and Ummenhofer, C. C. The role of the subtropical north atlantic water cycle in recent us extreme precipitation events. Climate dynamics, 50(3-4): 1291–1305, 2018.

[10] Ozdemir, A., Iwen, M. A., and Aviyente, S. Multi- scale analysis for higher-order tensors. arXiv preprint arXiv:1704.08578, 2017.

[11] Park, J. Y., Carr, K. T., Zhang, S., Yue, Y., Yu, R. (2020). Multiresolution Tensor Learning for Efficient and Interpretable Spatial Analysis. arXiv preprint arXiv:2002.05578.

[12] PRISM Climate Group. Gridded climate data for the contiguous usa., 2013. http://prism.oregonstate. edu, accessed 07/08/19.

[13] Reed, S., Oord, A., Kalchbrenner, N., Colmenarejo, S. G., Wang, Z., Chen, Y., Belov, D., and Freitas, N. Parallel multiscale autoregressive density estimation. In Interna- tional Conference on Machine Learning, pp. 2912–2921, 2017

[14] Serban, I. V., Klinger, T., Tesauro, G., Talamadupula, K., Zhou, B., Bengio, Y., and Courville, A. C. Multiresolu- tion recurrent neural networks: An application to dialogue response generation. In AAAI, pp. 3288–3294, 2017.

[15] Yue, Y., Lucey, P., Carr, P., Bialkowski, A., and Matthews, I. Learning Fine-Grained Spatial Models for Dynamic Sports Play Prediction. In IEEE International Conference on Data Mining (ICDM), 2014.

[16] Zhang, C. (2005). Madden-julian oscillation. Reviews of Geophysics, 43(2).

Multiresolution Tensor Learning for Efficient and Interpretable Spatiotemporal Analysis

Raechel Walker, Rose Yu
University of California, San Diego
La Jolla, CA
r6walker@ucsd.edu, roseyu@eng.ucsd.edu

A Experiment Details

A.1 Climate Dataset

The monthly precipitation in the U.S. dataset from 1895-2019 is from the PRISM dataset. The monthly sea salinity and temperature datasets from 1900-2018 are from the EN4 reanalysis datasets. The spatial data contains North America and parts of South America. The coordinates are [-180W, 0W] and [-20S, 60N].

Moreover, we deseasonalize and detrend the datasets before they are split into the train, test, and validation sets. The historical oceanic and precipitation datasets contain monthly measurements of sea salinity, temperature, and precipitation respectively. Bilinear interpolation finegrains the model weights. We calculate the ratio for the number of gridpoints and the ratio for the number of months that we aggregate together in the next temporal resolution each time the spatiotemporal resolution changes. These ratios preserve the magnitude of the predictions.

We map each feature in the data to different dimensions in the tensor. Time is the first dimension. The variable is the second dimension(SSS and SST). The third and fourth dimensions are the latitude and longitude respectively. The shaded blue portion in Fig. 1 represents data we select given the lead value.

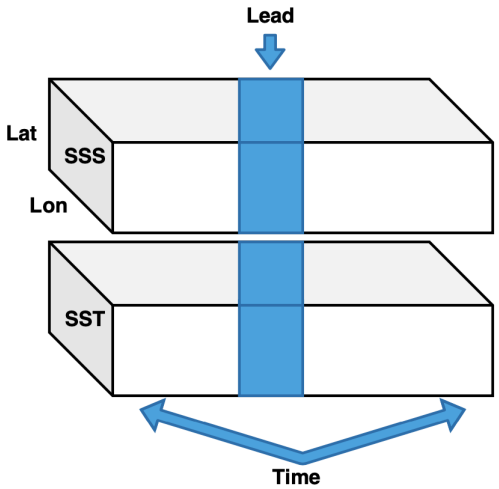

Figure 1: This is a diagram of the datasets that train the spatiotemporal MRTL model

Submitted to the AI for Earth Sciences Workshop at NeurIPS 2020. Do not distribute.

We utilize supervised tensor linear regression, and an order-3 tensor, but our model can use higher order cases. The three dimensions in this dataset include: spatial, temporal, and variable(SSS and SST).

A.2 Implementation Details

The datasets are split into train, validation, and test sets accordingly: 60-20-20. We utilize the Adam optimizer algorithm along with the stepwise learning rate decay. The stepsize is 1, and $\gamma = .95$. We use ten trials to generate the results. The mean squared error is the loss function. Stochastic gradient descent and tensor linear regression train the ST- MRTL model. We utilize spatial regularization in the loss function. Ridge regression and spatial regularization produce the most accurate results and prevent overfitting.

A.3 Finegraining Criteria and Resolutions

The original resolution of the climate data that ST-MRTL trains on is monthly. We aggregate these monthly data sets at different time resolutions. The full-rank ST-MRTL model trains on the following temporal resolutions: 1, and 4. The low-rank ST-MRTL model trains on the following temporal resolutions: 4, and 12. Furthermore, we train the full-rank ST-MRTL model on resolutions 4x9, 8x18, and 12x27. The low-rank model trains on resolutions 12x27, 24x54, 40x90, 60x135, and 80x180. In regards to the finegraining criteria, we terminate training the model when the training loss is greater than all of the past five training losses.

A.4 Hyperparameters

Table 1 displays the range of ST-MRTL hyperparameters that we experimented with. We experiment with hyperparameters to produce the most accurate model.

Table 1: Range of ST-MRTL model Hyperameters

Hyperparameter	Climate
Batch size	8-128
Full-rank learning rate η	$10^{-2} - 10^{-5}$
Full-rank regularization λ	$10^{-4} - 10^{-1}$
Low-rank learning rate η	$10^{-2} - 10^{-5}$
Low-rank regularization λ	$10^{-4} - 0$
Spatial regularization σ	$.003 - .3$
Learning rate decay γ	$0.70 - 0.95$