Spatiotemporal Covariance Neural Networks

Andrea Cavallo, Mohammad Sabbaqi, and Elvin Isufi

Delft University of Technology, Delft, Netherlands
{a.cavallo, m.sabbaqi, e.isufi-1}@tudelft.nl

1 Introduction

In this work, we introduce the SpatioTemporal coVariance Neural Network (STVNN), a temporal convolutional graph neural network for multivariate time series that employs the sample covariance matrix of the data as a graph. We show that STVNN draws similarities with temporal PCA, as it manipulates the data according to eigencomponents of their covariance. To account for streaming data and potential distribution shifts, we adopt an online learning strategy, thus updating the model when new data becomes available. Observing data in a stream introduces uncertainties, as the estimate of the covariance matrix might be imprecise and the model parameters are optimized only on the observed data. In this context, we theoretically show that STVNN is stable to errors in both covariance matrix and parameters estimation, thus improving over temporal PCA approaches which are unstable to ill-defined covariance matrix eigenvalues (i.e., when two covariance matrix eigenvalues are close, a slight change in data points might correspond to a large change in the principal directions [2]). This abstract summarizes the work in [1].

2 Online Time coVariance Neural Networks

We assume to have M observations of a multivariate time series, each of size N, in a matrix $\mathbf{X} \in \mathbb{R}^{N \times M}$, where samples are temporally ordered. The time series has time-varying mean $\boldsymbol{\mu}_t$ and covariance $\mathbf{C}_t$ (constant in case of stationarity). Since we do not have access to the true mean and covariance, we compute estimates $\hat{\boldsymbol{\mu}}_t, \hat{\mathbf{C}}_t$ from the t observed samples.

We define the SpatioTemporal coVariance graph Filter (STVF) as

$$\mathbf{z}_t = \sum_{t'=0}^{T-1} \sum_{k=0}^{K} h_{kt'} \hat{\mathbf{C}}_t^k \mathbf{x}_{t-t'} = \mathbf{H}(\hat{\mathbf{C}}_t, \mathbf{h}_t, \mathbf{x}_{t-T+1:t}) \quad (1)$$

where T is the time window size, K is the order of the convolutional filter and $h_{kt'}$ are coefficients. This filter performs graph signal shift operations up to the K-th neighborhood to model spatial relationships and aggregates information over the previous T temporal samples to account for temporal interactions. We define a STVNN layer by using a bank of STVF of size F_{out} whose filters process the input in parallel and by applying a pointwise non-linearity σ to the output. Then, we define a STVNN by stacking L consecutive STVNN layers and we denote it as $\Phi(\hat{\mathbf{C}}, \mathbf{h}, \mathbf{x}_t)$, where $\mathbf{h}$ contains all the learnable coefficients.

3 Stability analysis and experiments

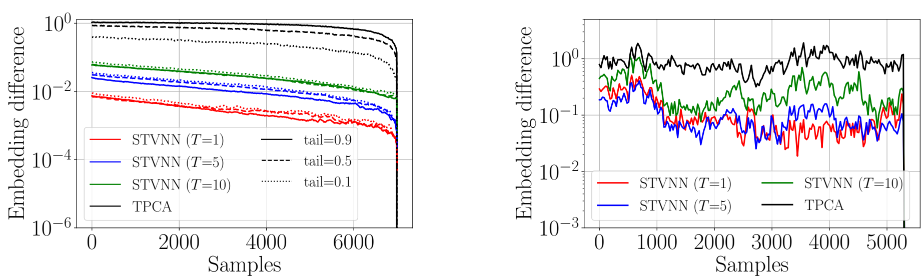
We analyze the stability of the STVF with respect to to two sources of errors due to streaming data: the uncertainties in the estimation of the covariance matrix and the suboptimal model parameters.

Theorem 1. *Consider a random variable* $\mathbf{x}_t \in \mathbb{R}^N$ *with covariance matrix* $\mathbf{C}$ *and such that* $\|\mathbf{x}\|_2 \leq G\sqrt{\mathbb{E}[\|\mathbf{x}\|_2^2]}$ *for a constant* $G \geq 1$*. Given a sample covariance matrix* $\hat{\mathbf{C}}_t$ *estimated from* t *samples* $\mathbf{x}_t$ *such that* $\|\mathbf{x}_t\| \leq 1$*, an integral Lipschitz STVF with temporal dimension* T*, Lipschitz constant* P *and two sets of coefficients,* $\mathbf{h}^*$ *optimized over the complete dataset, and* $\mathbf{h}_t$ *optimized over* t *samples using online gradient descent with learning rate* η*, the following holds with probability at least* $(1-e^{-\epsilon})(1-2e^{-u})$*:*

$$
\begin{aligned}
&\|\mathbf{H}(\hat{\mathbf{C}}_t, \mathbf{h}_t, \mathbf{x}_t) - \mathbf{H}(\mathbf{C}, \mathbf{h}^*, \mathbf{x}_t)\| \leq \mathcal{O}\left(\frac{1}{t}\right) + \\
&\frac{1}{\sqrt{t}} PT\sqrt{N}\left(k_{max}e^{\epsilon/2} + QG\|\mathbf{C}\|\sqrt{N(\log N + u)}\right) + \frac{\|\mathbf{h}^*\|_2^2}{2\eta t}
\end{aligned} \tag{2}
$$

where Q *is an absolute constant,* $k_{max} = \max_j k_j$ *and* $k_j = \left(\mathbb{E}[\|\mathbf{X}\mathbf{X}^\mathsf{T}\mathbf{v}_j\|_2^2] - \lambda_j^2\right)^{\frac{1}{2}}$ *is a related to the kurtosis of the data.*

The bound decreases with the number of time samples with rate $\mathcal{O}\left(1/\sqrt{t}\right)$, since the sample covariance becomes closer to the true one and the parameters approach their optimal values. The covariance-related errors (first term) dominate the parameters-related ones (second term) asymptotically. Larger window sizes T consider more temporal information but lead to lower stability. The Lipschitz constant P shows that the more discriminative the filter is (higher P), the less stable it is, as common for GNNs.



(a) Synthetic datasets

(b) Exchange rate [3]

Fig. 1: Stability of STVNN and TPCA models.

The numerical results in Figure 1 support the theoretical findings as STVNN variants are more stable than TPCA. We refer to [1] for more details.

References

1. Andrea Cavallo, Mohammad Sabbaqi, and Elvin Isufi. Spatiotemporal covariance neural networks. *ECML PKDD*, 2024.
2. Ian T. Jolliffe. *Principal component analysis.* Springer Verlag, New York, 2002.
3. Zonghan Wu, Shirui Pan, Guodong Long, Jing Jiang, Xiaojun Chang, and Chengqi Zhang. Connecting the dots: Multivariate time series forecasting with graph neural networks. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 753–763, 2020.