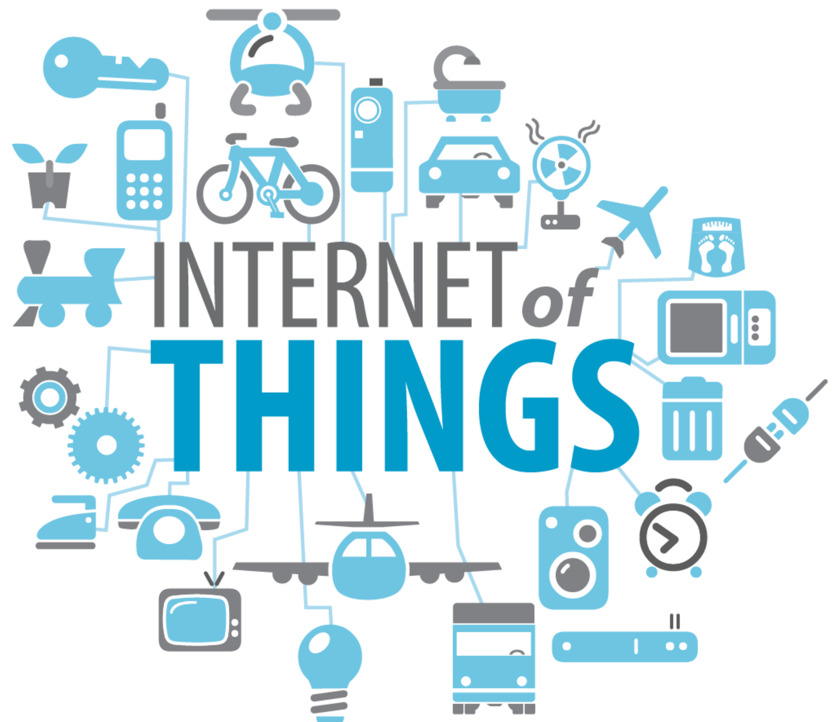
High-dimensional Time Series Clustering via Cross-Predictability

Dezhi Hong, Quanquan Gu, Kamin Whitehouse
Computer Science, University of Virginia

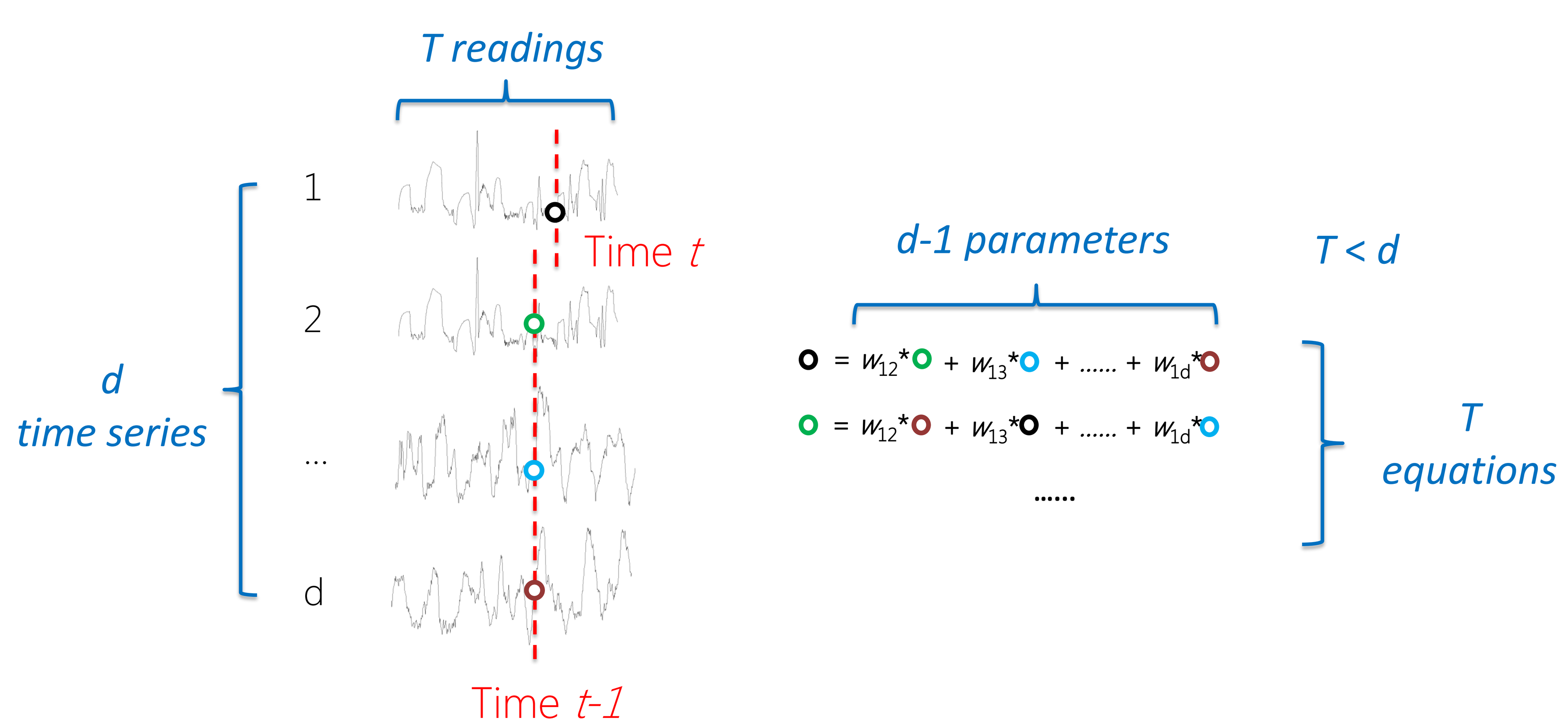
A Connected World in Future



The Internet of Things are expected to have over 25 billion devices by 2020.

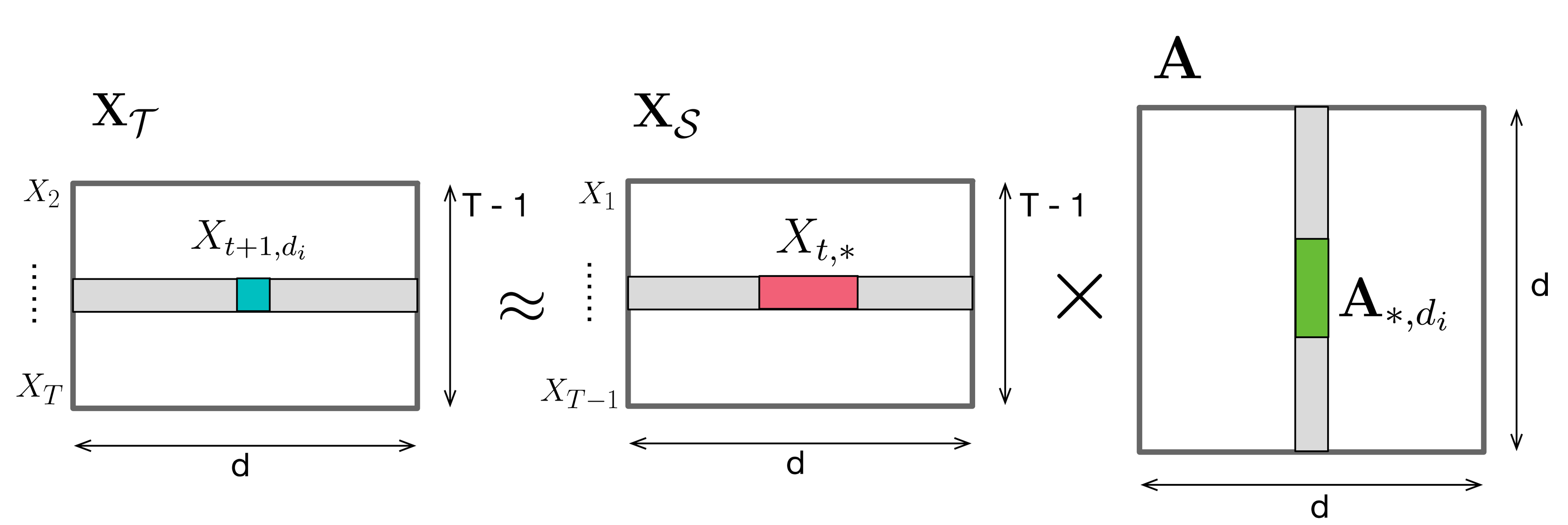
Clustering is an important primitive step, and time series data will often be high-dimensional: the number of time series *d* (i.e., number of sensors) will be **much larger** than the length of each time series *T*.

Challenge of High-dimensional Data



High-dimensional data result in an under-constrained problem: there are T equations for estimating d parameters, where T << d.

Clustering High-dimensional Data from VAR



We assume the high-dimensional data follow a vector autoregressive model (VAR) and do the clustering based on the degree to which a future value in each time series is predicted by past values of the others. We call this "cross-predictability".

Methodology

The input:

$$\mathbf{X} = [\boldsymbol{X}_1, \ldots \boldsymbol{X}_T]^\top \in \mathbb{R}^{T\times d} \quad (d \text{ time series of length } T)$$

$$\mathbf{X}_{\mathcal{S}} = [\boldsymbol{X}_1, \ldots, \boldsymbol{X}_{T-1}]^\top \in \mathbb{R}^{(T-1)\times d}, \ \ \mathbf{X}_{\mathcal{T}} = [\boldsymbol{X}_2, \ldots, \boldsymbol{X}_T]^\top \in \mathbb{R}^{(T-1)\times d}$$

$$\hat{\boldsymbol{\Sigma}} = \mathbf{X}_{\mathcal{S}}^\top \mathbf{X}_{\mathcal{S}}/(T-1), \text{ and } \hat{\boldsymbol{\gamma}}_i = \mathbf{X}_{\mathcal{S}}^\top (\mathbf{X}_{\mathcal{T}})_{*i}/(T-1)$$

The VAR model:

$$\boldsymbol{X}_{t+1} = \mathbf{A}\boldsymbol{X}_t + \boldsymbol{Z}_t, \ \text{ for } t = 1, 2, \ldots, T-1.$$

The estimator:

$$\hat{\boldsymbol{\beta}}_i = \arg\min_{\boldsymbol{\beta}_i} \ \lambda \|\hat{\boldsymbol{\Sigma}}\boldsymbol{\beta}_i - \hat{\boldsymbol{\gamma}}_i\|_{\infty,\infty} + \|\boldsymbol{\beta}_i\|_1$$

Set $\hat{\mathbf{A}} = \left[\hat{\boldsymbol{\beta}}_1, \ldots, \hat{\boldsymbol{\beta}}_d\right]^\top$, and then apply a spectral clustering algorithm.

Main Results

Main Theorem *Under the assumption of VAR model with a block diagonal transition matrix, we compactly denote* $\mathcal{P}_0^l = \mathcal{P}(\boldsymbol{\Sigma}_{\mathcal{S}_l,\mathcal{S}_l})$, $\mathcal{P}_1^l = \mathcal{P}((\boldsymbol{\Sigma}_1)_{\mathcal{S}_l,\mathcal{S}_l})$, $r_0^l = r(\mathcal{P}_0^l)$, $r_1^l = r(\mathcal{P}_1^l)$, *and* $r_0 r_1 = \min_l r_0^l r_1^l$ *for* $l = 1, 2, \ldots, k$, *and let*

$$\rho = \frac{16\|\boldsymbol{\Sigma}\|_2 \max_j \Sigma_{jj}}{\min_j \Sigma_{jj}(1 - \|\mathbf{A}\|_2)} \sqrt{\frac{6\log d + 4}{T}}.$$

Furthermore, if

$$r_0 r_1 > \frac{\|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\mathcal{S}_l}\|_{\infty,\infty} + 2\rho}{\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty} - 2\rho},$$

where $\boldsymbol{\gamma}_{\mathcal{S}_l} \in \mathbb{R}^{d_l}$ *is a column of* $\boldsymbol{\Sigma}_1$, *then with probability at least* $1 - 6d^{-1}$ *the cluster recovery property holds for all the values of the regularization parameter* λ *in the range:*

$$\frac{1}{r_0 r_1(\|\boldsymbol{\gamma}_{\mathcal{S}_l}\|_{\infty,\infty} - 2\rho) - \rho} < \lambda < \frac{1}{\rho + \|\boldsymbol{\Sigma}_{\mathcal{S}_l^c,\mathcal{S}_l}\|_{\infty,\infty}},$$

which is guaranteed to be non-empty.