Online Learning of Nonlinear and Dynamic Graphs

Seminar on GraphsData@TUDelft

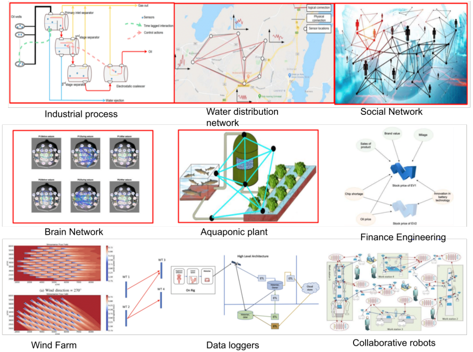
Rohan T. Money

SIGIPRO Department, Simula Metropolitan Center for Digital Engineering, Norway

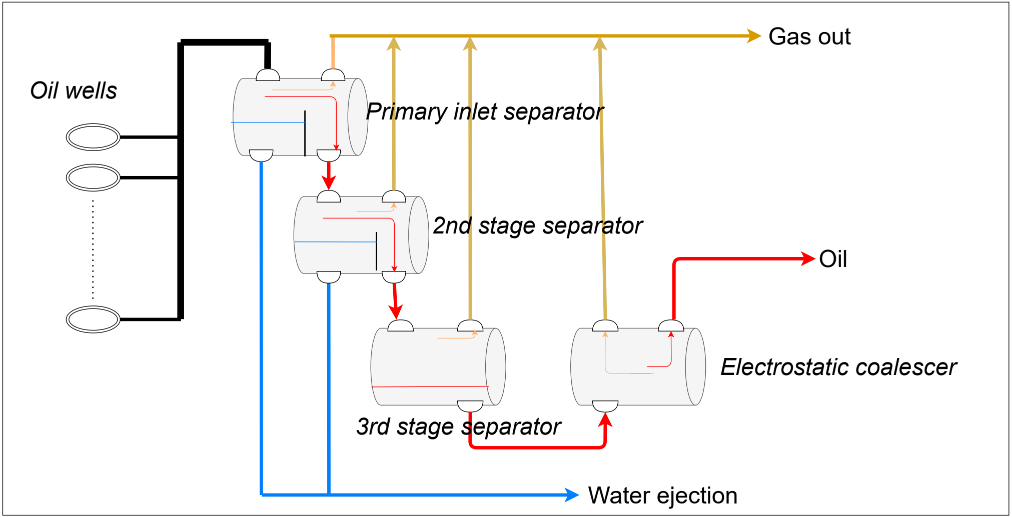
11 July 2024

Motivation

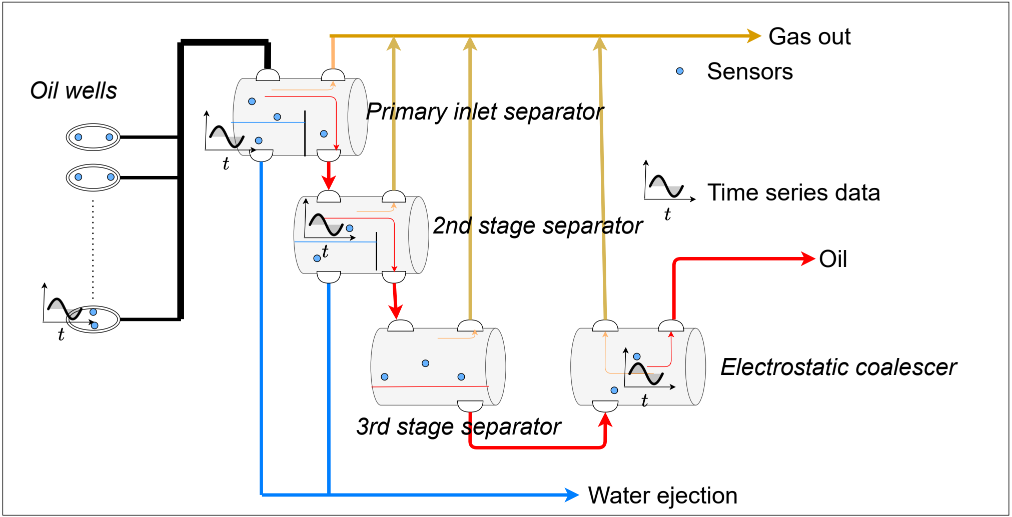
- Spatial and temporal inter-dependency of data
- Different possible machine learning tasks



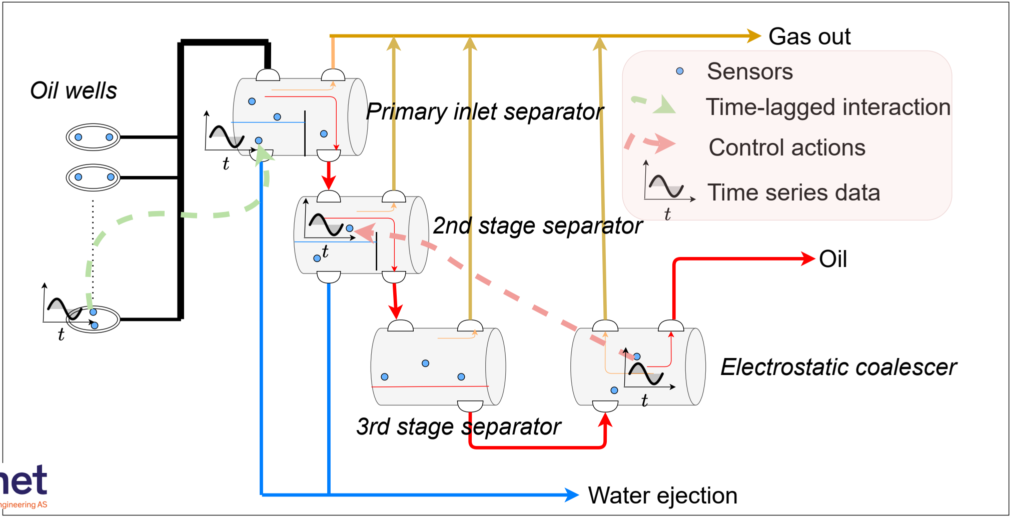
Schematic of oil and gas plant



Schematic of oil and gas plant



Schematic of oil and gas plant



simulamet
Simula Metropolitan Center for Digital Engineering AS

Nonlinear Vector Auto-regressive(VAR) model

- A P-th order non-linear VAR model with N number of nodes

$$y_n[t] = \sum_{n'=1}^{N} \sum_{p=1}^{P} f_{n,n'}^{(p)}(y_{n'}[t-p]) + u_n[t] \tag{1}$$

Nonlinear Vector Auto-regressive(VAR) model

- A P-th order non-linear VAR model with N number of nodes

$$y_n[t] = \sum_{n'=1}^{N} \sum_{p=1}^{P} f_{n,n'}^{(p)}(y_{n'}[t-p]) + u_n[t] \qquad (1)$$

- Estimate $f_{n,n'}^{(p)}(.)$

Reproducing kernel Hilbert space (RKHS)

- Assume functions $f_{n,n'}^{(p)}(.)$ in (1) belong to RKHS:

$$\mathcal{H}_{n'}^{(p)} := \left\{ f_{n,n'}^{(p)} | f_{n,n'}^{(p)}(y) = \sum_{t=0}^{\infty} \beta_{n,n',t}^{(p)} \, \kappa_{n'}^{(p)}(y, y_{n'}[t-p]) \right\}, \tag{2}$$

- $\kappa_{n'}^{(p)}(.,.) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is the Hilbert space basis function, often known as the kernel

Reproducing kernel Hilbert space (RKHS)

- Assume functions $f_{n,n'}^{(p)}(.)$ in (1) belong to RKHS:

$$\mathcal{H}_{n'}^{(p)} := \left\{ f_{n,n'}^{(p)} | f_{n,n'}^{(p)}(y) = \sum_{t=0}^{\infty} \beta_{n,n',t}^{(p)} \ \kappa_{n'}^{(p)}(y, y_{n'}[t-p]) \right\}, \tag{2}$$

- $\kappa_{n'}^{(p)}(.,.) : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is the Hilbert space basis function, often known as the kernel
- Hilbert space is characterized by the inner product $\langle \kappa_{n'}^{(p)}(y, x_1), \kappa_{n'}^{(p)}(y, x_2) \rangle = \sum_{t=0}^{\infty} \kappa_{n'}^{(p)}(y[t], x_1) \ \kappa_{n'}^{(p)}(y[t], x_2)$

Non-parametric optimization

- For a node n, the least-squares (LS) estimates of $\left\{f_{n,n'}^{(p)} \in \mathcal{H}_{n'}^{(p)}; n' = 1, \ldots, N,\ p = 1, \ldots, P\right\}$ are obtained by solving,

$$\left\{\widehat{f}_{n,n'}^{(p)}\right\}_{n',p} = \arg \min_{\left\{f_{n,n'}^{(p)} \in \mathcal{H}_{n'}^{(p)}\right\}} \frac{1}{2} \sum_{\tau=P}^{T-1} \left[y_n[\tau] - \sum_{n'=1}^{N} \sum_{p=1}^{P} f_{n,n'}^{(p)}(y_{n'}[\tau - p])\right]^2 + \lambda \sum_{n'=1}^{N} \sum_{p=1}^{P} \Omega(||f_{n,n'}^{(p)}||_{\mathcal{H}_{n'}^{(p)}}). \quad (3)$$

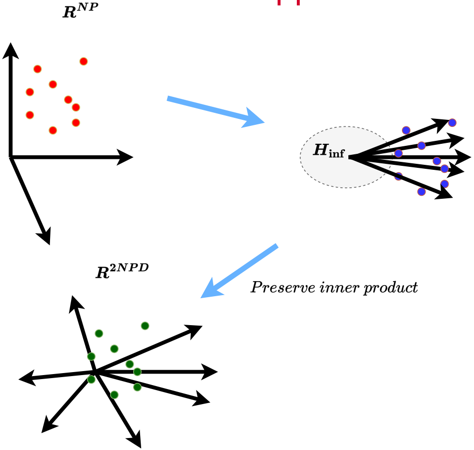
Representer Theorem

- The solution of (3) can be written using a finite number of data samples:

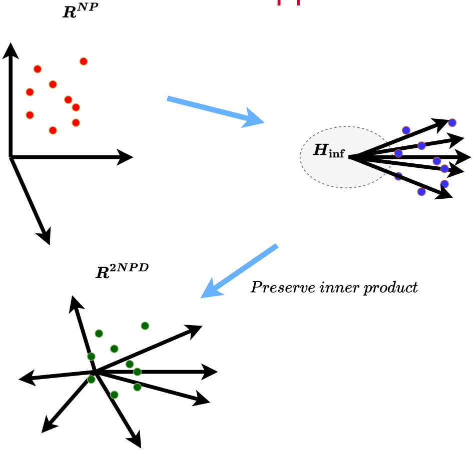
$$\widehat{f}^{(p)}_{n,n'}\left(y_{n'}[\tau-p]\right) \quad = \sum_{t=p}^{p+T-1} \beta^{(p)}_{n,n',(t-p)} \kappa^{(p)}_{n'}\left(y_{n'}[\tau-p]), y_{n'}[t-p]\right) \tag{4}$$

- Solution becomes prohibitive as number of data points increases

Random feature approximation

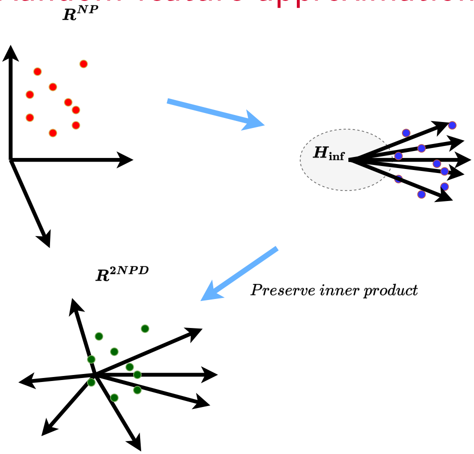


Random feature approximation



- Inner product preserving map

Random feature approximation



- Inner product preserving map
- A. Rahimi and B. Recht, “Random features for large-scale kernel machines,” NIPS’07

- Obtain a fixed dimension ($2D$ terms) approximation of the function $\hat{f}^{(p)}_{n,n'}$:

$$\hat{\check{f}}^{(p)}_{n,n'}\left(y_{n'}[\tau - p]\right) = \sum_{t=p}^{p+T-1} \beta^{(p)}_{n,n',(t-p)} \boldsymbol{z}_{\boldsymbol{v}}\left(y_{n'}[\tau - p]\right)^{\top} \boldsymbol{z}_{\boldsymbol{v}}\left(y_{n'}[t-p]\right)$$

$$= \boldsymbol{\alpha}^{(p)}_{n,n'}{}^{\top} \boldsymbol{z}_{\boldsymbol{v}}\left(y_{n'}[\tau - p]\right), \quad (5)$$

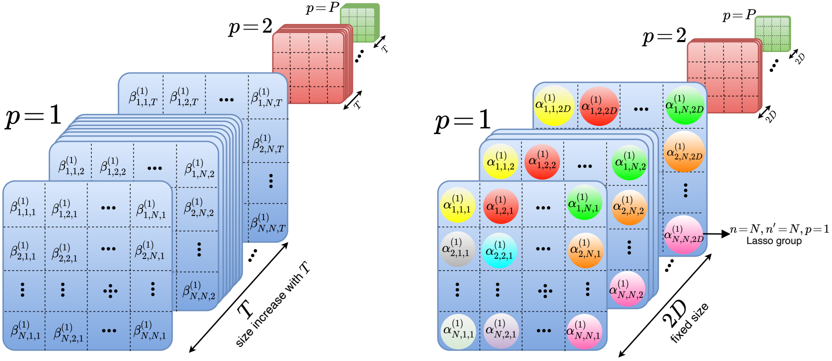
$$\boldsymbol{z}_{\boldsymbol{v}}(x) = \frac{1}{\sqrt{D}}[\sin v_1 x, \ldots, \sin v_D x, \cos v_1 x, \ldots, \cos v_D x]^{\top}. \quad (6)$$

- Stack the entries of $\boldsymbol{\alpha}_{n,n'}^{(p)}$ and $\boldsymbol{z}_{n',d}^{(p)}(\tau)$ to obtain the vectors $\boldsymbol{\alpha}_n \in \mathbb{R}^{2PND}$ and $\boldsymbol{z}_\tau \in \mathbb{R}^{2PND}$

$$\widehat{\boldsymbol{\alpha}}_n = \arg\min_{\boldsymbol{\alpha}_n} \mathcal{L}^n(\boldsymbol{\alpha}_n) + \lambda \sum_{n'=1}^{N} \sum_{p=1}^{P} \|\boldsymbol{\alpha}_{n,n'}^{(p)}\|_2, \tag{7}$$

$$\mathcal{L}^n(\boldsymbol{\alpha}_n) = \frac{1}{2} \sum_{\tau=P}^{T-1} \left[y_n[\tau] - \boldsymbol{\alpha}_n^\top \boldsymbol{z}_\tau \right]^2 \tag{8}$$

- $\lambda \geq 0$ is the regularization parameter


$p=1$
$p=2$
$p=P$
T
size increase with T
$p=1$
$p=2$
$p=P$
$2D$
fixed size
$n=N, n'=N, p=1$
Lasso group

Online optimization

- Replace the original loss function $\mathcal{L}^n(\boldsymbol{\alpha}_n)$ in (7) with a running average loss function:

$$\tilde{\ell}_t^n(\boldsymbol{\alpha}_n) = \mu \sum_{\tau=P}^{t} \gamma^{t-\tau} \ell_\tau^n(\boldsymbol{\alpha}_n) \tag{9}$$

where $l_\tau^n(\boldsymbol{\alpha}_n) = \frac{1}{2}[y_n[\tau] - \boldsymbol{\alpha}_n^\top \boldsymbol{\kappa}_\tau]^2$.

- convex loss and non differentiable regularizer

$$\widehat{\boldsymbol{\alpha}}_n = \arg\min_{\boldsymbol{\alpha}_n} \tilde{\ell}_t^n\left(\boldsymbol{\alpha}_n\right) + \lambda \sum_{n'=1}^{N} \sum_{p=1}^{P} \|\boldsymbol{\alpha}_{n,n'}^{(p)}\|_2. \tag{10}$$

simulamet
Simula Metropolitan Center for Digital Engineering AS

- convex loss and non differentiable regularizer

$$\widehat{\boldsymbol{\alpha}}_n = \arg\min_{\boldsymbol{\alpha}_n} \tilde{\ell}_t^n\left(\boldsymbol{\alpha}_n\right) + \lambda \sum_{n'=1}^{N} \sum_{p=1}^{P} \|\boldsymbol{\alpha}_{n,n'}^{(p)}\|_2. \quad (10)$$

- Closed form solution

$$\boldsymbol{\alpha}_{n,n'}^{(p)}[t+1] = \left(\boldsymbol{\alpha}_{n,n'}^{(p)}[t] - \gamma_t \mathbf{v}_{n,n'}^{(p)}[t]\right) \times \left[1 - \frac{\gamma_t \lambda \, 1\{n \neq n'\}}{\|\boldsymbol{\alpha}_{n,n'}^{(p)}[t] - \gamma_t \mathbf{v}_{n,n'}^{(p)}[t]\|_2}\right]_+, \quad (11)$$

where $[x]_+ = \max\{0, x\}$ and

$$1\{n \neq n'\} = \begin{cases} 1, & \text{if } n \neq n' \\ 0, & n = n' \end{cases}$$

simulamet
Simula Metropolitan Center for Digital Engineering AS

Theoretical Analysis: Dynamic Regret

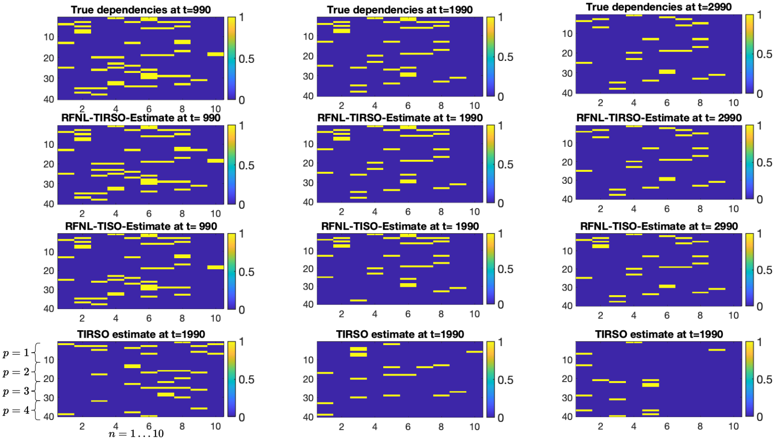
- Dynamic Regret: Test the capability of an online algorithm in a dynamic environment.

$$\boldsymbol{R}_n[t] = \sum_{t=P}^{T} [h_t^{(n)}(\boldsymbol{f}_n[t]) - h_t^{(n)}(\boldsymbol{f}_n^*[t])] \tag{12}$$

- Sub-linear dynamic regret by suitably choosing ϵ as long as $\boldsymbol{W}_T^n = \sum_{t=P}^{T} \|\boldsymbol{\alpha}_n^*[t] - \boldsymbol{\alpha}_n^*[t-1]\|_2$ is sub-linear.

Experiment 1: synthetic data

- N=5,P=4,T=3000, equation (1)(VAR)
- Adjacency matrix generated with edge probability .3
- Non linearity in (1) is induced by Gaussian kernel
- 30% edges disappears
- $\left\{\boldsymbol{\alpha}_{n,n'}^{(p)}[t]\right\}$ are estimated $\widehat{b}_{n,n'}^{(p)} = \|\boldsymbol{\alpha}_{n,n'}^{(p)}[t]\|_2$ at $t = T$ and find pseudo adjacency matrix

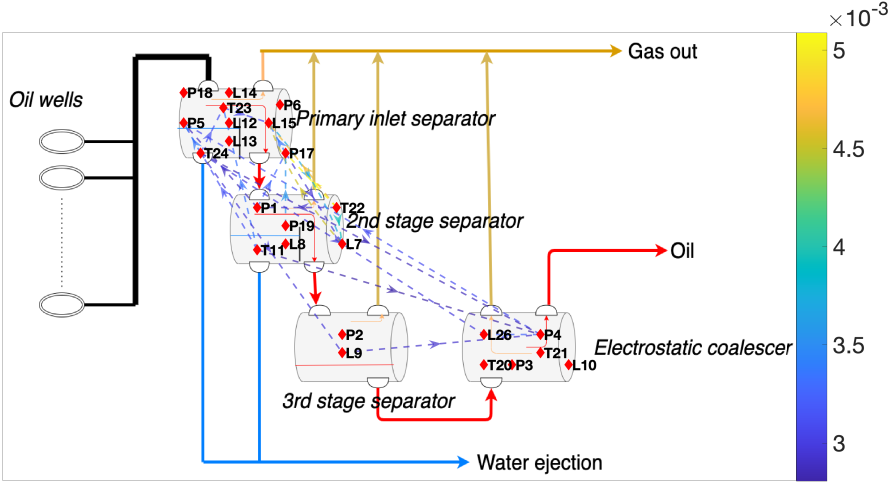


Simula Metropolitan Center for Digital Engineering AS

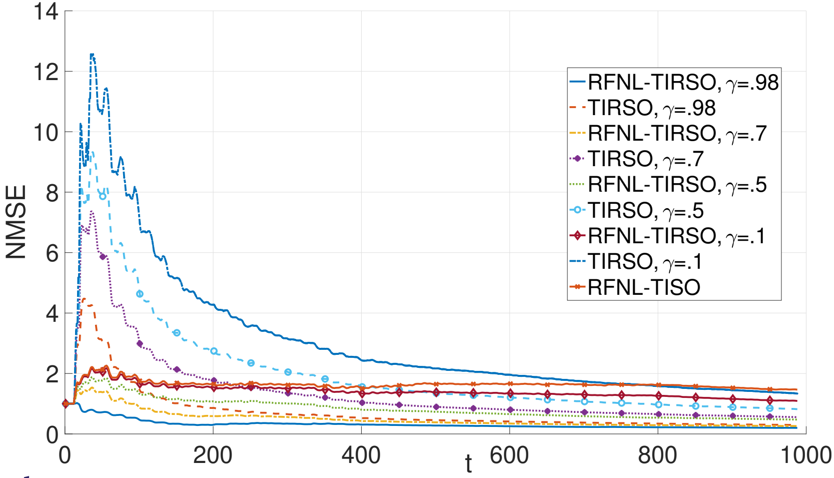
Experiment 2: Real data

- Real data from Lundin's offshore oil and gas (O&G) platform Edvard-Grieg[1]
- Temperature (T), pressure (P), or oil-level (L) sensors placed in separators.
- The causal dependencies among the 24 time series obtained by averaging the RFNL-TIRSO estimates for one hour

[1] https://www.lundin-energy.com/

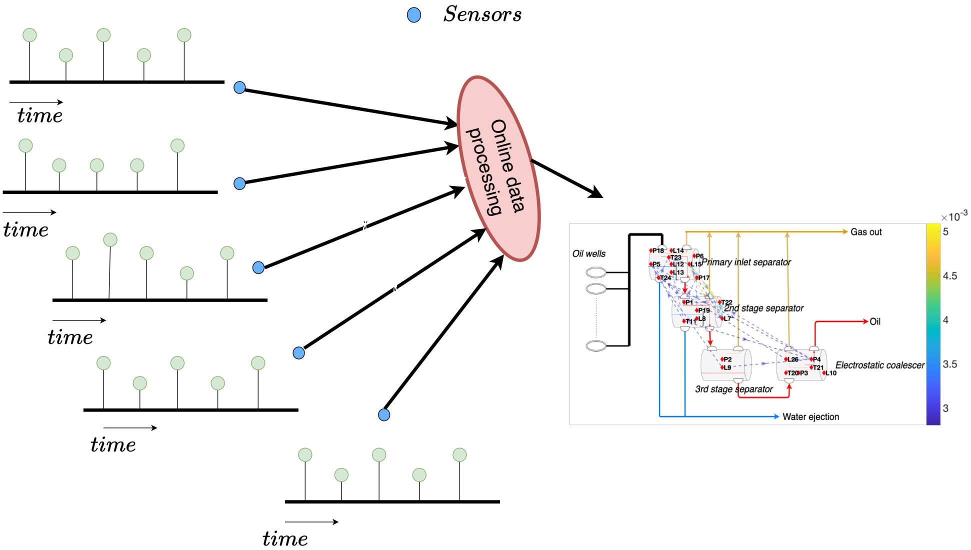

Oil wells
Gas out
Primary inlet separator
2nd stage separator
Oil
Electrostatic coalescer
3rd stage separator
Water ejection
P18
L14
T23
P6
P5
L12
L15
L13
T24
P17
P1
T22
P19
L8
L7
T11
P2
L9
L26
P4
T21
T20
P3
L10
×10-3
5
4.5
4
3.5
3

NMSE

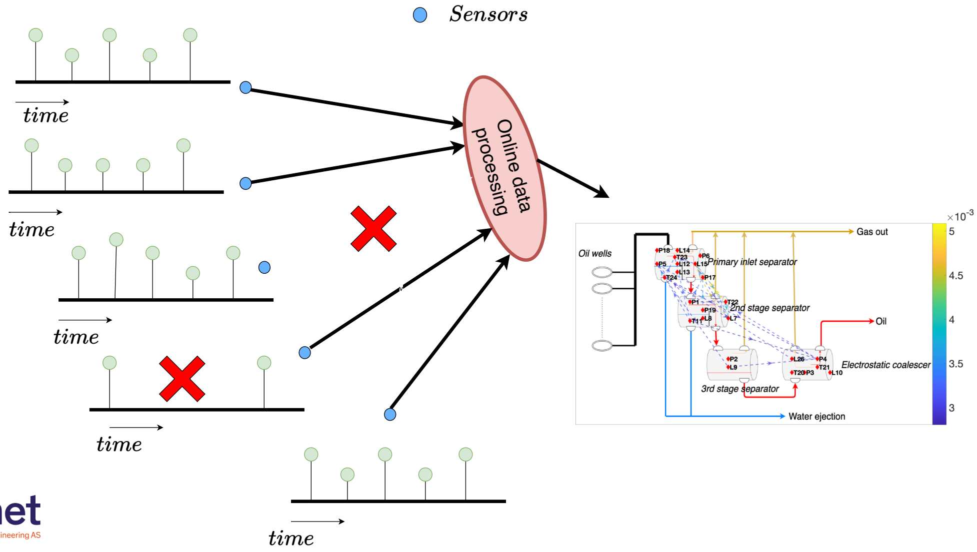


NMSE
t
0
2
4
6
8
10
12
14
0
200
400
600
800
1000
RFNL-TIRSO, γ=.98
TIRSO, γ=.98
RFNL-TIRSO, γ=.7
TIRSO, γ=.7
RFNL-TIRSO, γ=.5
TIRSO, γ=.5
RFNL-TIRSO, γ=.1
TIRSO, γ=.1
RFNL-TISO

Missing data



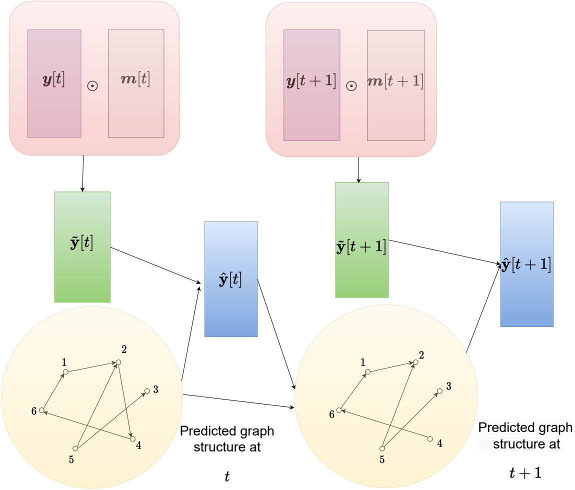
Missing data



Problem formulation

- Masking vector $\mathbf{m}[t] \in R^N$
- Observed vector signal $\tilde{\mathbf{y}}[t]$
- $\mathbf{y}[t] = [y_1[t], ..., y_n[\tau]]\top \in \mathbb{R}^N$

$$\tilde{\mathbf{y}}[t] = \mathbf{m}[t] \odot (\mathbf{y}[t] + \mathbf{e}[t]) \tag{13}$$


$\boldsymbol{y}[t]$ ⊙ $\boldsymbol{m}[t]$
$\boldsymbol{y}[t+1]$ ⊙ $\boldsymbol{m}[t+1]$
$\tilde{\mathbf{y}}[t]$
$\hat{\mathbf{y}}[t]$
$\tilde{\mathbf{y}}[t+1]$
$\hat{\mathbf{y}}[t+1]$
1
2
3
4
5
6
Predicted graph structure at
t
Predicted graph structure at
$t+1$

Signal reconstruction

$$\hat{y}_n[t] = \arg\min_{y_n[t]} \ell_t^n \left(\boldsymbol{\alpha}_n, y_n[t]\right) \tag{14}$$

$$\hat{y}_n[t] = \frac{\nu m_n[t]\tilde{y}_n[t]}{M_t + \nu m_n[t]} + \frac{k_n[t] M_t}{\nu m_n[t] + M_t} \tag{15}$$

Online Topology Identification

- $\ell_t^n(\boldsymbol{\alpha}_n) = \frac{1}{2}[\hat{y}_n[t] - \boldsymbol{\alpha}_n^\top \boldsymbol{z}_{\boldsymbol{v}}[t]]^2$

$$\widehat{\boldsymbol{\alpha}}_n = \arg \min_{\boldsymbol{\alpha}_n} \ell_t^n(\boldsymbol{\alpha}_n) + \lambda \sum_{n'=1}^{N} \sum_{p=1}^{P} \|\boldsymbol{\alpha}_{n,n'}^{(p)}\|_2 \tag{16}$$

Online Topology Identification

- $\ell_t^n(\boldsymbol{\alpha}_n) = \frac{1}{2}[\hat{y}_n[t] - \boldsymbol{\alpha}_n^\top \boldsymbol{z}_{\boldsymbol{v}}[t]]^2$

$$\widehat{\boldsymbol{\alpha}}_n = \arg \min_{\boldsymbol{\alpha}_n} \ell_t^n(\boldsymbol{\alpha}_n) + \lambda \sum_{n'=1}^{N} \sum_{p=1}^{P} \|\boldsymbol{\alpha}_{n,n'}^{(p)}\|_2 \tag{16}$$

- Closed form solution

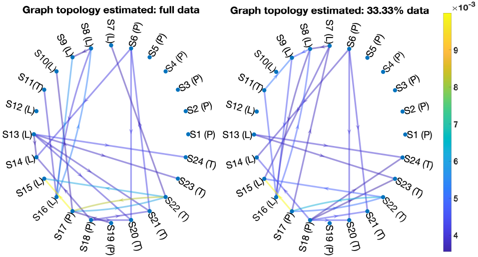
$$\boldsymbol{\alpha}_{n,n'}^{(p)}[t+1] = \left(\boldsymbol{\alpha}_{n,n'}^{(p)}[t] - \gamma_t \mathbf{v}_{n,n'}^{(p)}[t]\right) \times \left[1 - \frac{\gamma_t \lambda \, 1\{n \neq n'\}}{\|\boldsymbol{\alpha}_{n,n'}^{(p)}[t] - \gamma_t \mathbf{v}_{n,n'}^{(p)}[t]\|_2}\right]_+, \tag{17}$$

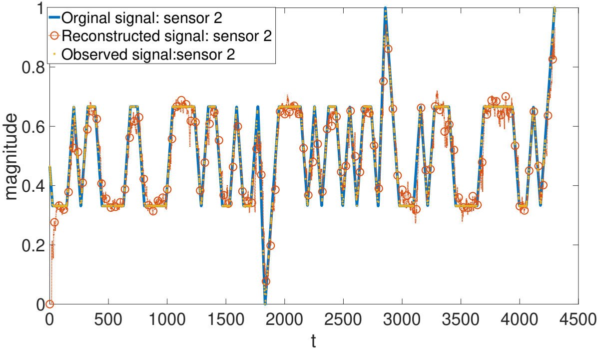
where $[x]_+ = \max\{0, x\}$ and

$$1\{n \neq n'\} = \begin{cases} 1, & \text{if } n \neq n' \\ 0, & n = n' \end{cases}$$

Experiment (a): Real data

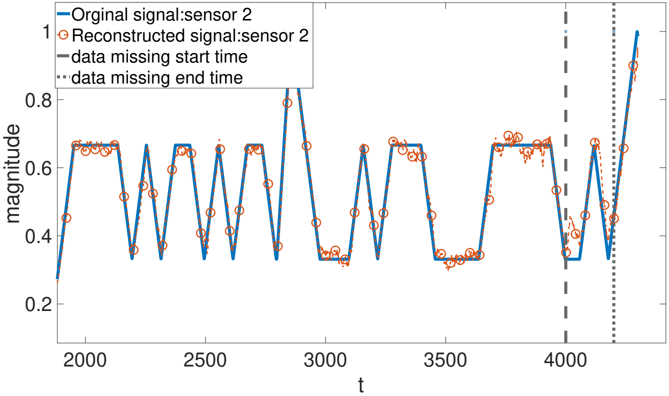
- $N = 24,\ P = 12,\ T = 4300$
- Data from 8 sensors available at a time





Experiment (b): Real data

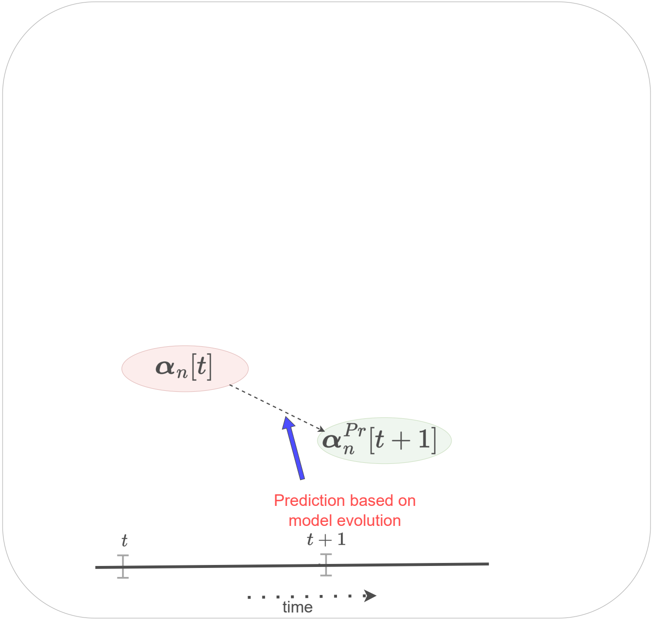
- Sensor data missing from $t = 4000$ to $t = 4200$



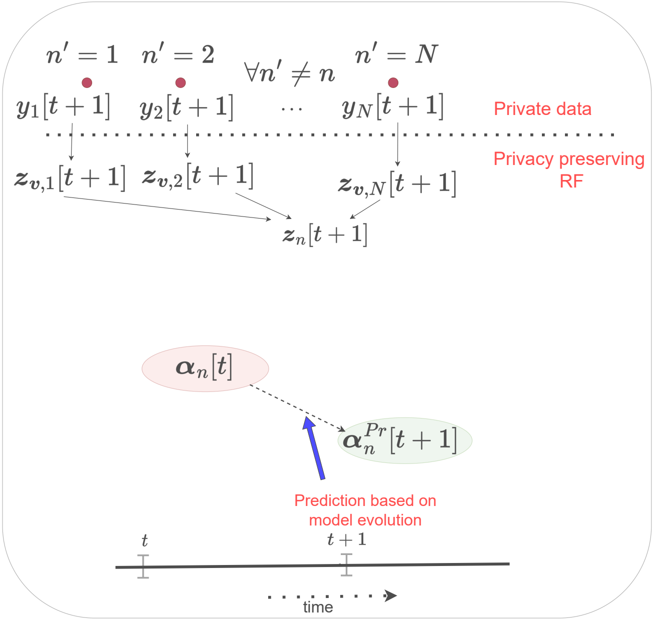
simulamet
Simula Metropolitan Center for Digital Engineering AS

Time structured approach

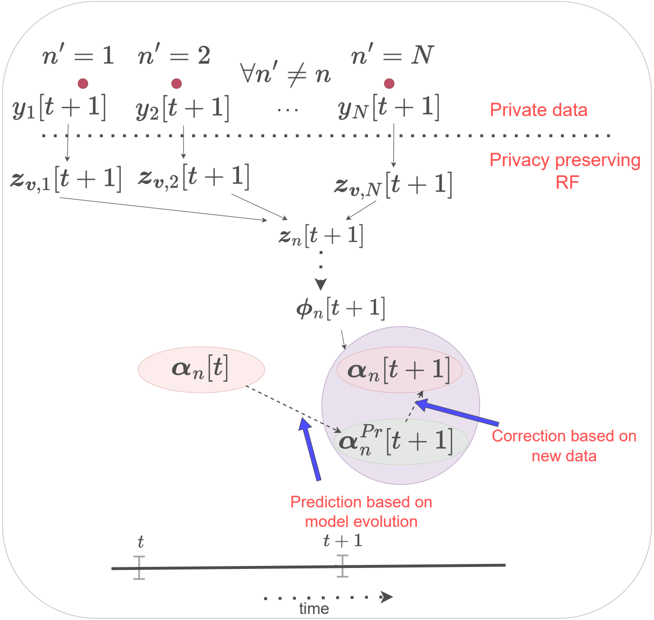
- Predict the model based on its evolution and then correct the prediction when the new data sample is available


$\boldsymbol{\alpha}_n[t]$
$\boldsymbol{\alpha}_n^{Pr}[t+1]$
Prediction based on model evolution
t
$t+1$
time

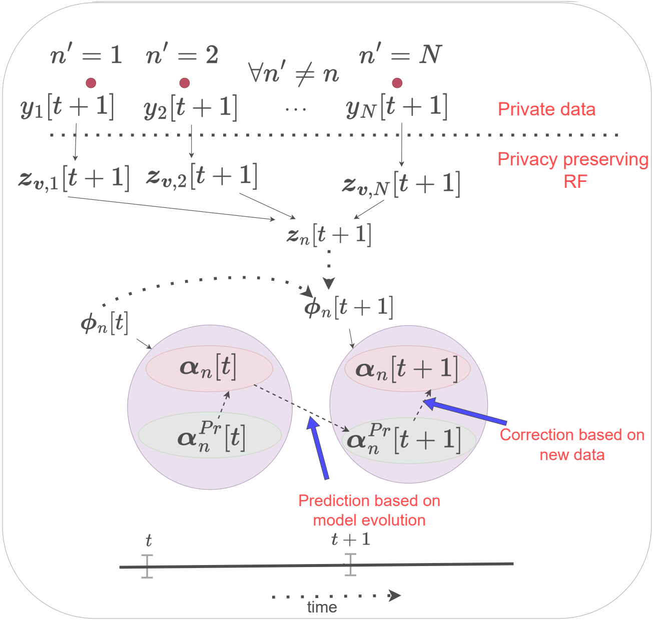
SIGIPRO



$n' = 1$
$n' = 2$
$\forall n' \neq n$
$n' = N$
$y_1[t+1]$
$y_2[t+1]$
$\cdots$
$y_N[t+1]$
Private data
Privacy preserving RF
$z_{v,1}[t+1]$
$z_{v,2}[t+1]$
$z_{v,N}[t+1]$
$z_n[t+1]$
$\alpha_n[t]$
$\alpha_n^{Pr}[t+1]$
Prediction based on model evolution
t
$t+1$
time



$n' = 1$
$n' = 2$
$\forall n' \neq n$
$n' = N$
$y_1[t+1]$
$y_2[t+1]$
$\cdots$
$y_N[t+1]$
Private data
$z_{v,1}[t+1]$
$z_{v,2}[t+1]$
$z_{v,N}[t+1]$
Privacy preserving RF
$z_n[t+1]$
$\phi_n[t+1]$
$\alpha_n[t]$
$\alpha_n[t+1]$
$\alpha_n^{Pr}[t+1]$
Correction based on new data
Prediction based on model evolution
t
$t+1$
time


$n' = 1$
$n' = 2$
$\forall n' \neq n$
$n' = N$
$y_1[t+1]$
$y_2[t+1]$
$\cdots$
$y_N[t+1]$
Private data
$z_{v,1}[t+1]$
$z_{v,2}[t+1]$
$z_{v,N}[t+1]$
Privacy preserving RF
$z_n[t+1]$
$\phi_n[t+1]$
$\phi_n[t]$
$\alpha_n[t]$
$\alpha_n^{Pr}[t]$
$\alpha_n[t+1]$
$\alpha_n^{Pr}[t+1]$
Correction based on new data
Prediction based on model evolution
t
$t+1$
time

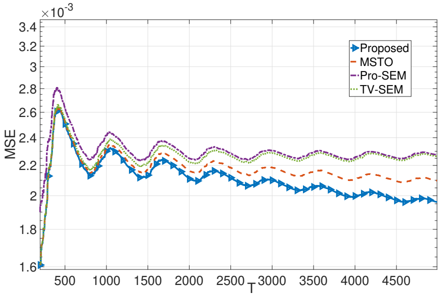
Experiment synthetic data

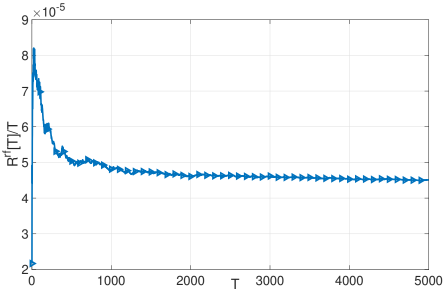
- N=5, T= 5000
- $\mathbf{W}[0] \in \mathbb{R}^{5\times 5}$ is constructed using an Erdős-Rényi random graph with diagonal entries zero

$$\mathbf{y}[t] = 0.1(\mathbf{I} - \mathbf{W}[t])^{-1}\mathbf{u}[t] + 0.1\sin((\mathbf{I} - \mathbf{W}[t])^{-1}\mathbf{u}[t]) \tag{18}$$

$$\mathbf{W}[t+1] = \mathbf{W}[t] + 0.001\sin(0.01t)\mathbf{W}[t] \tag{19}$$

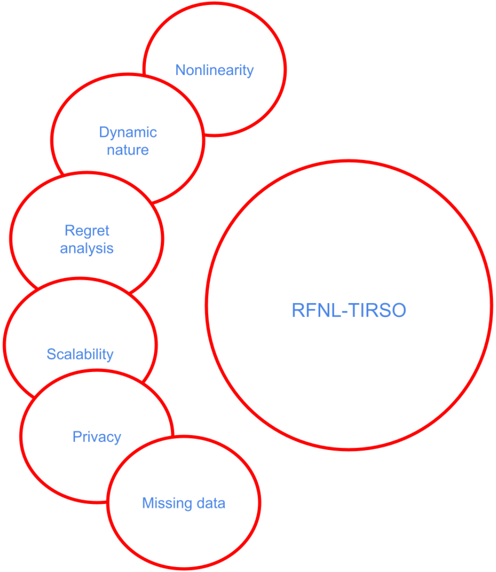
MSE comparison and convergence in terms of dynamic regret





simulamet
Simula Metropolitan Center for Digital Engineering AS

Conclusion


Nonlinearity
Dynamic nature
Regret analysis
Scalability
Privacy
Missing data
RFNL-TIRSO

Do check out

- R. Money, J. Krishnan and B. Beferull-Lozano, ”Sparse online learning with kernels using random features for estimating nonlinear dynamic graphs,” in IEEE Transactions on Signal Processing 2023
- R. Money, J. Krishnan and B. Beferull-Lozano, ”Random feature approximation for online nonlinear graph topology identification,” European Signal Processing Conference (EUSIPCO) 2022
- R. Money, J. Krishnan, B. Beferull-Lozano and E. Isufi, ”Scalable and privacy-aware online learning of nonlinear structural equation models,” in IEEE Open Journal on Signal Processing 2023

Thank you!