Sparse Causal Temporal Modeling to Inform Power System Defense

Michail Misyrlis, Rajgopal Kannan, Charalampos Chelmis, Viktor K. Prasanna

Presented by:
Michail Misyrlis

Overview

Smart Grid Defense and Major Influencers

Sparse Causal Temporal Methods

- **Lasso Granger**
- **Graphical Lasso**

Modelling Influence

Data Set

Experimental Set Up

Results

- **N50 per Node**
- **Single Feeder Single Week**
- **Predictive Accuracy**

Discussion

Smart Grid Defense and Major Influencers

Defense Against False Data Injection Attack

- Man –in-the-middle attack bypasses bad data detection in SCADA
- Undetectably manipulate state estimates: control electricity price, cascading blackouts
- Defense: Start from an initial set of nodes and gradually expand until entire set is observable

Partial Data Problem

- Need predictions in real time
- Some sensors systematically only provide partial data
- Need other sensors' data to perform prediction
- Which other sensors to use? Major Influencers

Protect Major Influencers

- Budget constraints allow securing only a fraction of nodes
- Incentivized to protect Major Influencers

Sparse Causal Temporal Methods

Regularized Regression Setting

$X \in \mathbb{R}^{m \times n}, y \in \mathbb{R}^n, \beta \in \mathbb{R}^d$

Objective: $\min_{\beta} f(\beta, X, y) + \lambda J(\beta)$

Lasso: $J(\beta) = \|\beta\|_1$

Squared error: $f(\beta, X, y) = \|y - X\beta\|_2^2$

Inherently Sparsity Inducing Regularization Term

Lasso Granger

- Combines Granger Causality and ℓ_1-norm regularization
- A time series $\boldsymbol{x}^i$ “Granger causes” another time series feature $\boldsymbol{x}^j$ if and only if using both $\boldsymbol{x}^i$’s and $\boldsymbol{x}^j$’s data leads to statistically significantly higher accuracy than predicting based solely on past values of $\boldsymbol{x}^j$
- Model dependencies between multiple time series
- Exhaustively trying pairs entails $O(n^2)$ tests, reduce the cost to by using Lasso regression: Lasso Granger

Sparse Causal Temporal Methods

Graphical Lasso

Graphical Models describe the topology of components in a complex probability model

- Components: Variables
- Topology: dependencies

Gaussian Graphical Models: Markov Random Fields of jointly Gaussian variables

$X \in \mathbb{R}^{m \times n}$: design matrix

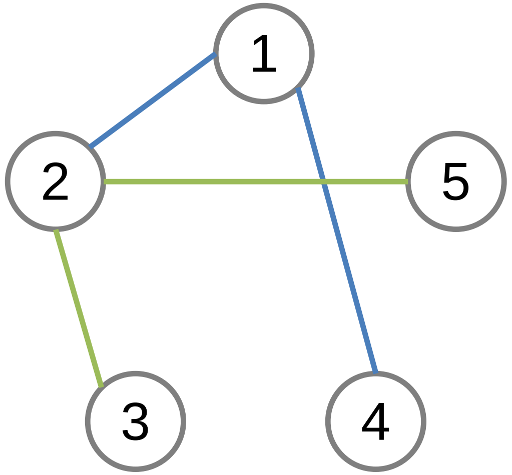
$\Sigma \in R^{nxn}$: covariance matrix

$\Omega = \Sigma^{-1}$: Inverse Covariance / Precision matrix

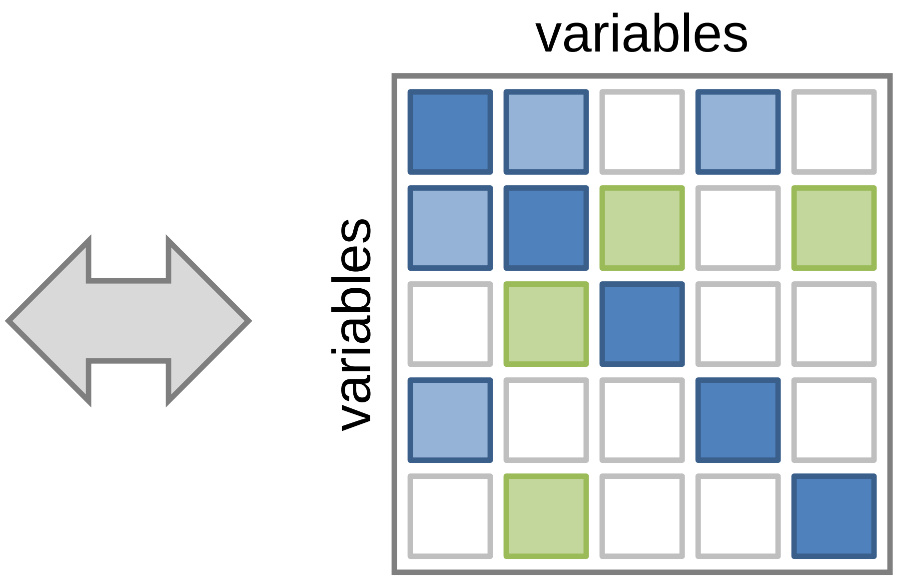
$\max \log det\Omega - tr(\Sigma\Omega) - \lambda||\Omega||_1$ (1)

Sparsity inducing regularization term

Graphical Lasso solves (1) by fitting a modified Lasso regression to each variable in a cyclical Fashion



Graph



Σ^{-1}: Precision matrix (Inverse covariance)

Modelling Influence

Dependency Matrix

- $S = \{s_1, \ldots, s_n\}$: Set of sensors which collect real time data
- Need to make predictions for each s_i
- $M \in \mathbb{R}^{n \times n}$, $M[i, j]$ reflects dependency of sensor s_i's time series data on sensor s_j's time series data
- Can be used as a look-up table to select most useful features for predicting s_i's time series data by choosing non-zero elements in the i-th row of M

Influence

- Use Dependency Matrix M to rank sensors by their degree of influence on other features
- Influence I^k of time series $\boldsymbol{x}^k$ defined as: $\mathrm{I}^{\mathrm{k}} = \Sigma_{j=1}^{n} M[k, j]$
- Obtain a ranking of time series features based on their Influence

Lasso Granger & Graphical Lasso both provide a Dependency Matrix

Data Set & Experimental Set-Up

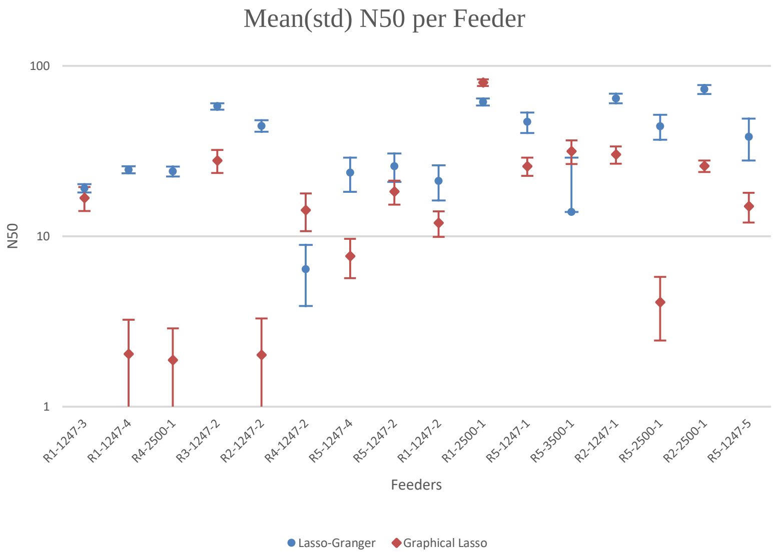
Feeders Data Set

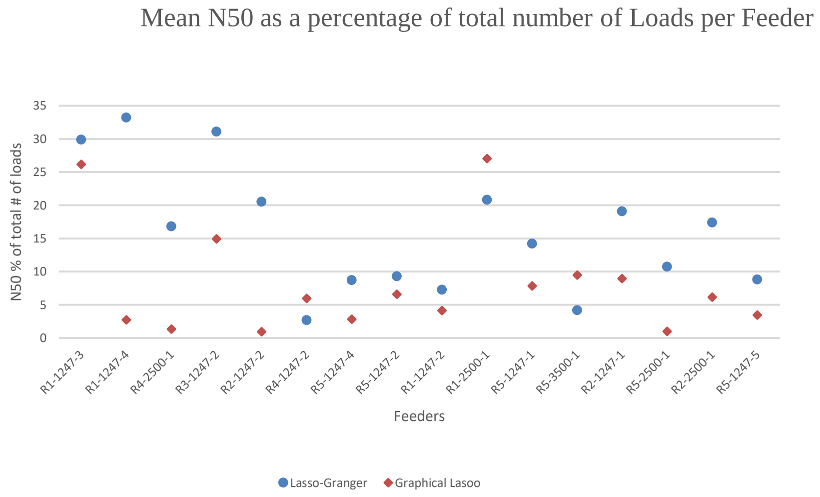
- Distribution nodes that connect substation to consumers
- Simulated with GridLab-D to study Photovoltaic Penetration
- Comprise 5 Geographical Regions
- 16 Feeders
- Each Feeder has Residential and Commercial loads, ranging from 64 to 436
- Averaged hourly values over one year – 8760 values per residential / commercial load

Experimental Set-Up

- For each Feeder: Use Data from first half of the week to learn two Dependency Matrices: Lasso-Granger, Graphical Lasso
- Determine globally most influential features using the Dependency Matrices using N50 statistic (number of features that contain 50% of the total weights)
- Use influential nodes' data to train Regression Trees on the first half of the week
- Use 24h data for training and a temporal horizon varying from 1h to 24h
- Use the learned Dependency Matrices and the Regression Tree models to test on the second half of the week
- Assess predictive accuracy by MAPE (mean absolute percentage error)
- Compare with ART, which is solely based on a node's own past time series data

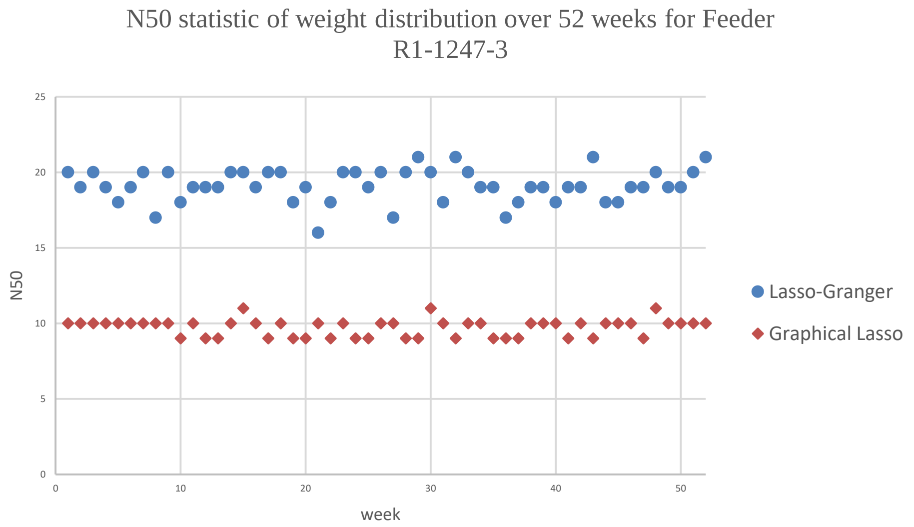
Results

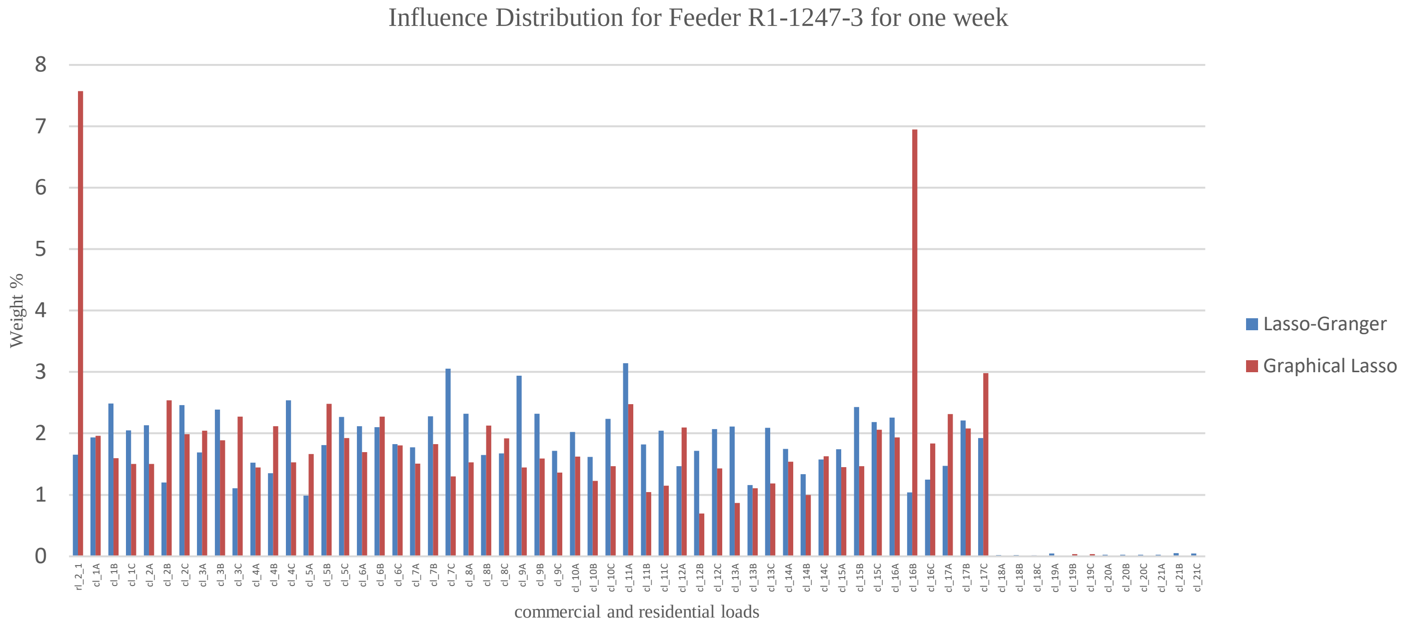




N50 calculated for Lasso Granger learned structure deviates less from the mean

Results





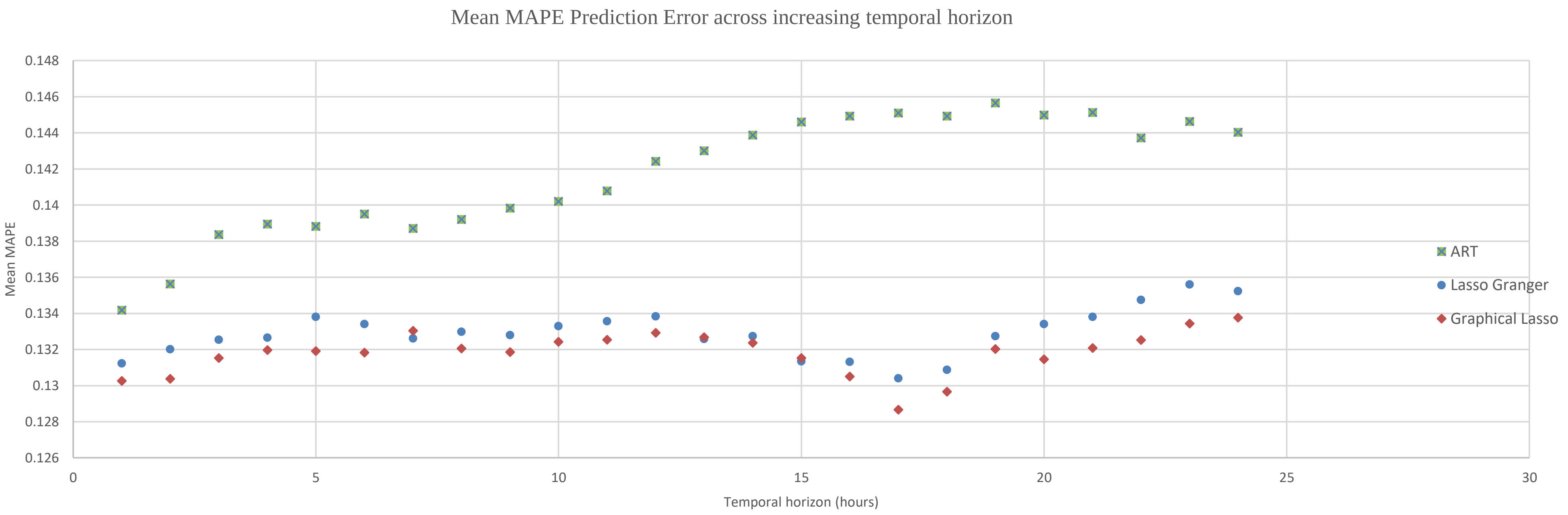
Lasso-Granger returns a N50 that amounts to larger set of nodes than that of Graphical Lasso

7-th week, Graphical Lasso zeros-out last 12 influence weights. Lasso-Granger weights more evenly distributed.

USC Viterbi School of Engineering

University of Southern California

Results



Even from the first hour, the latter two models lead to lower MAPE and the difference becomes more pronounced as the temporal horizon increases to 17 hours look-ahead.

Discussion

- Other feature's data lead to better short term and long term Prediction
- Graphical Lasso – Probabilistic models incorporation into SmartGrid prediction
- Fast methods – Real Time Prediction – Large Datasets
- Comparison of the methods i. Optimality ii. Predictive Accuracy
- Budget Restrictions – Calibrating cutoff percentage of weights

End of Presentation

Thank you

Portland Crime Dataset

- 213 days worth of data (latest uploaded data)
- Clustering location points based on Euclidean distance
- Limitations of clustering based on location: ignoring other similarities between locations
- Modelling Dependencies between clusters using Graphical Models

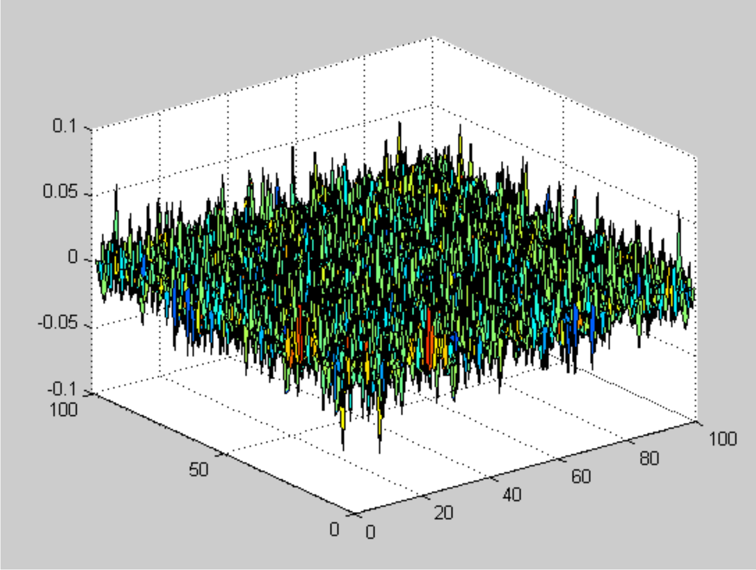
- Graphical Models Used: Graphical Lasso, Variable Selection Structure Learning, Multitask Structure Learning

$$\max_{\boldsymbol{\Omega} \succ 0} \left(\log\det \boldsymbol{\Omega} - \langle \widehat{\boldsymbol{\Sigma}}, \boldsymbol{\Omega} \rangle - \rho \|\boldsymbol{\Omega}\|_1 - \tau \|\boldsymbol{\Omega}\|_{1,p} \right)$$

$$\max_{(\forall k)\boldsymbol{\Omega}^{(k)} \succ 0} \left(\sum_k T^{(k)} \ell_{\widehat{\boldsymbol{\Sigma}}^{(k)}}(\boldsymbol{\Omega}^{(k)}) - \rho \|\boldsymbol{\Omega}\|_{1,\infty} \right)$$

- Graphical Models allow measuring likelihood of Data according to the model
- Variable Selection Consistently leads to models with the highest likelihood
- **Pipeline:**
- Cluster locations into 100 clusters
- Learn a Structure for 2 weeks
- Train Regression models using major influencers to predict next two weeks
- In the training, hyperparameters are learned: number of influencers & regression parameters
- Test on the next month

Portland Crime Dataset



- Lasso Regression: Mean Pearson Correlation – 0.2165

- Elastic Net Regression: Mean Pearson Correlation – 0.4507

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}(\|y - X\beta\|^2 + \lambda_2\|\beta\|^2 + \lambda_1\|\beta\|_1).$$

- Support Vector Regression: Mean Pearson Correlation – 0.05