LEARNING TIME VARYING GRAPHS

Vassilis Kalofolias, Andreas Loukas, Dorina Thanou, Pascal Frossard
Signal Processing Laboratory (LTS2/LTS4), École Polytechnique Fédérale de Lausanne

Abstract

We consider the problem of inferring the hidden structure of high-dimensional time-varying data. We aim at capturing the dynamic relationships by representing data as valued nodes in a sequence of graphs. Our approach is motivated by the observation that imposing a meaningful graph topology can help solving the generally ill-posed and challenging problem of structure inference.

We introduce a new prior that asserts that the graph edges change smoothly in time. We propose a primal-dual optimization algorithm that scales linearly with the number of allowed edges and can be easily parallelized. Our new algorithm is shown to outperform standard graph learning and other baseline methods both on a synthetic and a real dataset.

Why graph learning?

Unless the number of samples is large, learning a graph “explains” the data structure (variance) better.

Figure 1. Pixelwise structure of the MNIST image dataset obtained by a) graph learning [Kalofolias], b) empirical covariance estimation, and c) the more computationally expensive sparse inverse covariance estimator also known as graphical LASSO [Banerjee]. The vertical axis intuitively quantifies how efficiently the learned structure captured the data variance, whereas the horizontal axis corresponds to the number of images used for training.

Learning time-varying graphs

We consider the case of a graph that changes slowly over time.

Optimization problem. Discretize time in $K$ windows and denote by $W(k)$ for $k = 1, \ldots, K$ the adjacency matrix of the $k$-th window. Solve:

\[
\min_{W(1)\ldots W(K)} \sum_{k=2}^{K} \left( \| W(k) - Z(k) \|_{1,1} + f(W(k)) + \ldots \right)
\]

\[
\times \sum_{k=2}^{K} \text{f_{loss}} \left( W(k), W(k-1) \right)
\]

where $f_{loss} = \| W(k) - W(k-1) \|_{1,1}$ enforces that the graph edges change smoothly over time.

Complexity. $O(N^2K)$ per iteration, and the number of iterations is typically within the hundreds.

Measuring the quality of a graph

Definition 1. Total cumulative energy residual (TCER): Given a data distribution $p(X)$ with mean $\mu$ and covariance $C$, and a sorted orthogonal basis $Q = [q_1, \ldots, q_N]$, we have

\[
TCER(p(X), Q) = 1 - \frac{\sum_{n=1}^{N} w_n \|q_n^T X\|^2}{\max_{Q \subset \mathbb{R}^d} \frac{\sum_{n=1}^{N} w_n \|q_n^T A\|^2}{\sum_{n=1}^{N} w_n \|q_n\|^2}}
\]

where $w_n = N+1-n$ and the denominator is simply $\sum_{n=1}^{N} w_n \|A\|^2$, for $\sigma_i$ the $i$th singular value of $C + \mu I$.

Dynamic point-cloud de-noising

Goal is to overcome coordinate noise and point registration error, even if the geometry evolves over time.

Gaussian Markov chain graph estimation

Signals sampled from Gaussian Markov chain:

\[
p(X_t | X_{t-1}) = N(x_t | x_{t-1}, \alpha^2 I - \beta) N(x_{t-1} | x_{t-2}, \beta)
\]

Random walktype graph model

Figure 2. Learning a graph from each window independently ($\gamma = 0$), performs better than the covariance matrix with fewer samples. The new prior helps when the number of samples are few, achieving the best result for $\gamma = 200$ (6 samples per window).

Conclusions

New framework for learning dynamically changing graphs from smooth data observations:

- A time smoothness prior imposes that the graphs learnt in successive windows change smoothly over time.
- We achieve a good trade off between temporal resolution and computation cost.
- Our experiments show that the new model can outperform classical graphical learning and other baseline methods.

References

1. Vassilis Kalofolias, "How to learn a graph from smooth signals", Conference on Artificial Intelligence and Statistics (AISTATS), 2016