IDYNO: Learning Nonparametric DAGs from Interventional Dynamic Data

Tian Gao, Debarun Bhattacharjya, Elliot Nelson, Miao Liu, Yue Yu

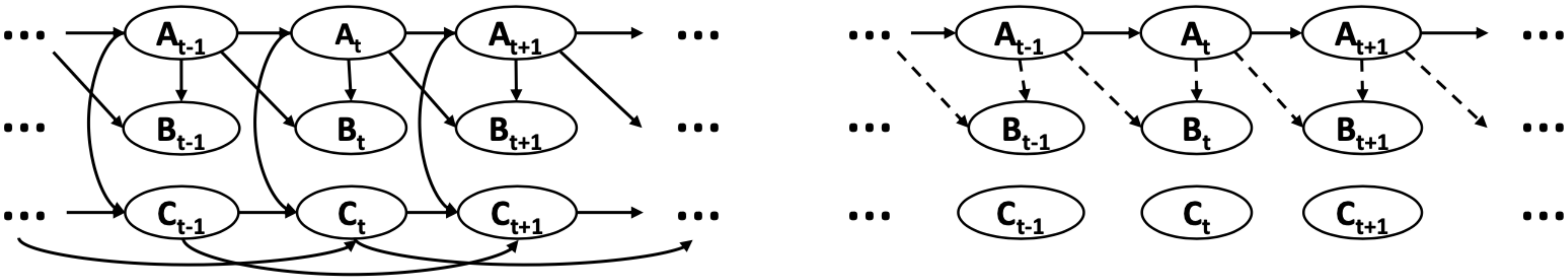
IBM Research, Lehigh University

tgao@us.ibm.com

ICML 2022 @ Baltimore, MD

Overview

- Directed acyclic graph (DAG) structure learning for dynamic Bayesian networks (DBN)
- IDYNO: capable to handle both observational and *(hard/soft)* *interventional* time series data
  - Optimization with potentially nonlinear objectives and continuous DAG constraints



(a) (b)

Related Works: DAG Learning

I.I.D. Data (BN)

- Score-based and constraint-based approaches [Spirtes et al, ’01; Chickering, ’02; Tsamardinos et al, ‘06]

- NOTEARS: Continuous algebraic characterization of the DAG [Zheng et al, ’18]

$$\min_{\theta, \mathbf{W}} L_\theta(\mathbf{X}; \mathbf{W}) - \lambda \Omega(\theta, \mathbf{W}) \quad s.t. \quad h(\mathbf{W}) = 0,$$

$$h(\mathbf{W}) = \mathrm{Tr}(e^{\mathbf{W}}) - d$$

Time Series Data (DBN)

- (Structured) vector auto-regressive (SVAR) [Swanson & Granger, ‘97; Reale & Wilson, ‘01]

- Constraint-based [Malinsky, ’19]

- DYNOTEARS: Continuous constraint [Pamfil, ‘20]

Time Series Data

Observation

- SVAR Model with Structure Equation Models (SEM)

$$\min_{\theta, \mathbf{W}, \mathbf{A}} L_\theta(\mathbf{X}; \mathbf{W}, \mathbf{A}) - \lambda\Omega(\theta, \mathbf{W}) \;\; s.t. \;\; h(\mathbf{W}) = 0,$$

$$X_t = X_t\mathbf{W} + X_{t-1}\mathbf{A_1} + ... + X_{t-p}\mathbf{A_p} + Z_t,$$

Intervention

- DCDI: Interventional DAG Learning [Brouillard et al, ‘20]

- For each interventional family k:

$$p^{(k)}(X) := \prod_{j \notin I_k} p_j^{(1)}(x_j | x_{\pi_j^G}) \prod_{j \in I_k} p_j^{(k)}(x_j | x_{\pi_j^G})$$

IDYNO: Formulation

- Linear SEM

$$\min_{\mathbf{W},\mathbf{A}} \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{d} \Big(|| \left(\mathbf{X} - \mathbf{X}\mathbf{W}_{(\mathbf{1})} - \mathbf{Y}\mathbf{A}_{(\mathbf{1})}\right)_j ||_2^{2(1-r_{kj}^{\mathcal{I}})} + || \left(\mathbf{X} - \mathbf{X}\mathbf{W}_{(\mathbf{k})} - \mathbf{Y}\mathbf{A}_{(\mathbf{k})}\right)_j ||_2^{2r_{kj}^{\mathcal{I}}} \Big) + \lambda\Omega(\theta)$$

$$s.t. \quad Tr(e^{\mathbf{W}}) - d = 0 \qquad (8)$$

Where $\left[r_{kj}^{I}\right] \in \{0,1\}^{k\times d} = 1$ when x_j is intervened on with interventional family k

IDYNO: Formulation

- Nonlinear Time Series Data

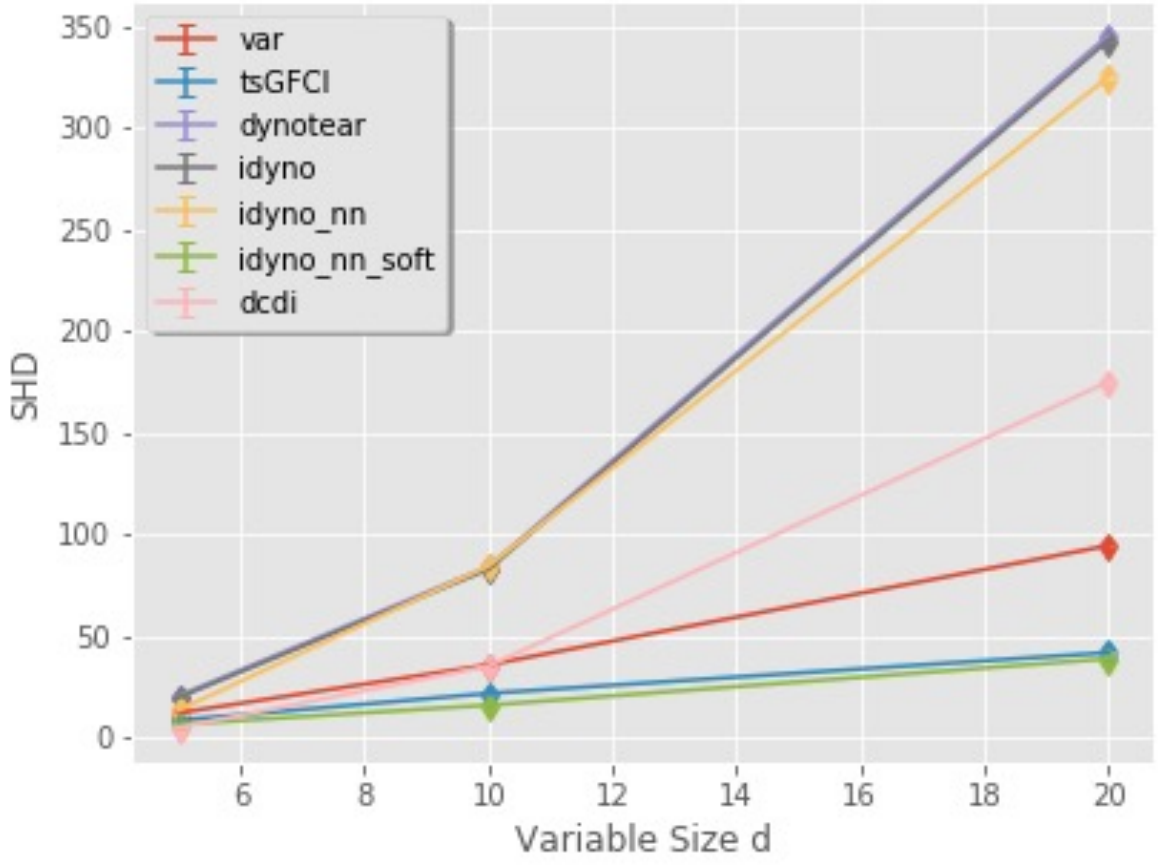
$$\min_{W,A,\theta} \frac{1}{n} \sum_{k=1}^{K} \sum_{j=1}^{d} L_j(\mathbf{X}, MLP(\mathbf{X}; \theta_j, \mathbf{A}_{(\mathbf{1})}, \mathbf{W}_{(\mathbf{1})}))^{1-r_{kj}^{\mathcal{I}}}$$
$$L_j(\mathbf{X}, MLP(\mathbf{X}; \theta_j, \mathbf{A}_{(\mathbf{k})}, \mathbf{W}_{(\mathbf{k})}))^{r_{kj}^{\mathcal{I}}}$$
$$+ \lambda_a ||A_j^{(1)}||_{1,1} + \lambda_w ||W_j^{(1)}||_{1,1} \qquad (12)$$
$$s.t. \quad Tr(e^{\mathbf{W}}) - d = 0 \qquad (13)$$

- **Identifiability**: Under some assumptions, the learned graph is (I, D_S)-Markov equivalent to the ground true graph G^*
  - D_S: DAGs that correspond to stationary dynamics with constant-in-time inter-slice and intra-slice conditional distributions
  - (I, D_S)-Markov equivalent: $D_S \subset DAG$ which are I-Markov equivalent to G^*

Empirical Evaluation

Table 1. SHD Results for Synthetic Linear Datasets

Dataset	DYNOTEARS	IDYNO
Observational	2.0± 0.0	2.0± 0.0
Interventional	32± 0.4	**19± 0.3**



Poster: Hall E #411

Tue 19 Jul 6:30 p.m. EDT — 8:30 p.m. EDT

Questions?

tgao@us.ibm.com