人工智慧作業報告簡報

信息工程學院 2021 級 干皓丞 2101212850

DAGs with No Fears: A Closer Look at Continuous Optimization for Learning Bayesian Networks

無所畏懼的 DAG：深入了解學習貝葉斯網絡的持續優化

https://arxiv.org/abs/2010.09133

Neural Information Processing Systems (NeurIPS 2020)

Subjects: Machine Learning (cs.LG); Machine Learning (stat.ML)

Motivation 动机

[33] Xun Zheng, Bryon Aragam, Pradeep K Ravikumar, and Eric P Xing. DAGs with NO TEARS: Continuous optimization for structure learning. In *Advances in Neural Information Processing Systems*, pages 9472–9483, December 2018.

arXiv | Cornell University

Statistics > Machine Learning

arXiv:1803.01422 (stat)

[Submitted on 4 Mar 2018 (v1), last revised 3 Nov 2018 (this version, v2)]

DAGs with NO TEARS: Continuous Optimization for Structure Learning

Xun Zheng, Bryon Aragam, Pradeep Ravikumar, Eric P. Xing

Download PDF

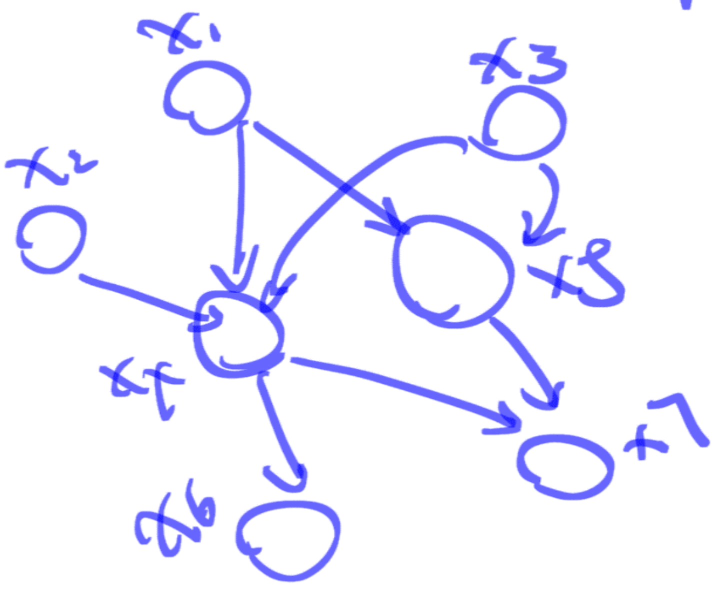
Estimating the structure of directed acyclic graphs (DAGs, also known as Bayesian networks) is a challenging problem since the search space of DAGs is combinatorial and scales superexponentially with the number of nodes. Existing approaches rely on various local heuristics for enforcing the acyclicity constraint. In this paper, we introduce a fundamentally different strategy: We formulate the structure learning problem as a purely \emph{continuous} optimization problem over real matrices that avoids this combinatorial constraint entirely. This is achieved by a novel characterization of acyclicity that is not only smooth but also exact. The resulting problem can be efficiently solved by standard numerical algorithms, which also makes implementation effortless. The proposed method outperforms existing ones, without imposing any structural assumptions on the graph such as bounded treewidth or in-degree. Code implementing the proposed algorithm is open-source and publicly available at this https URL.

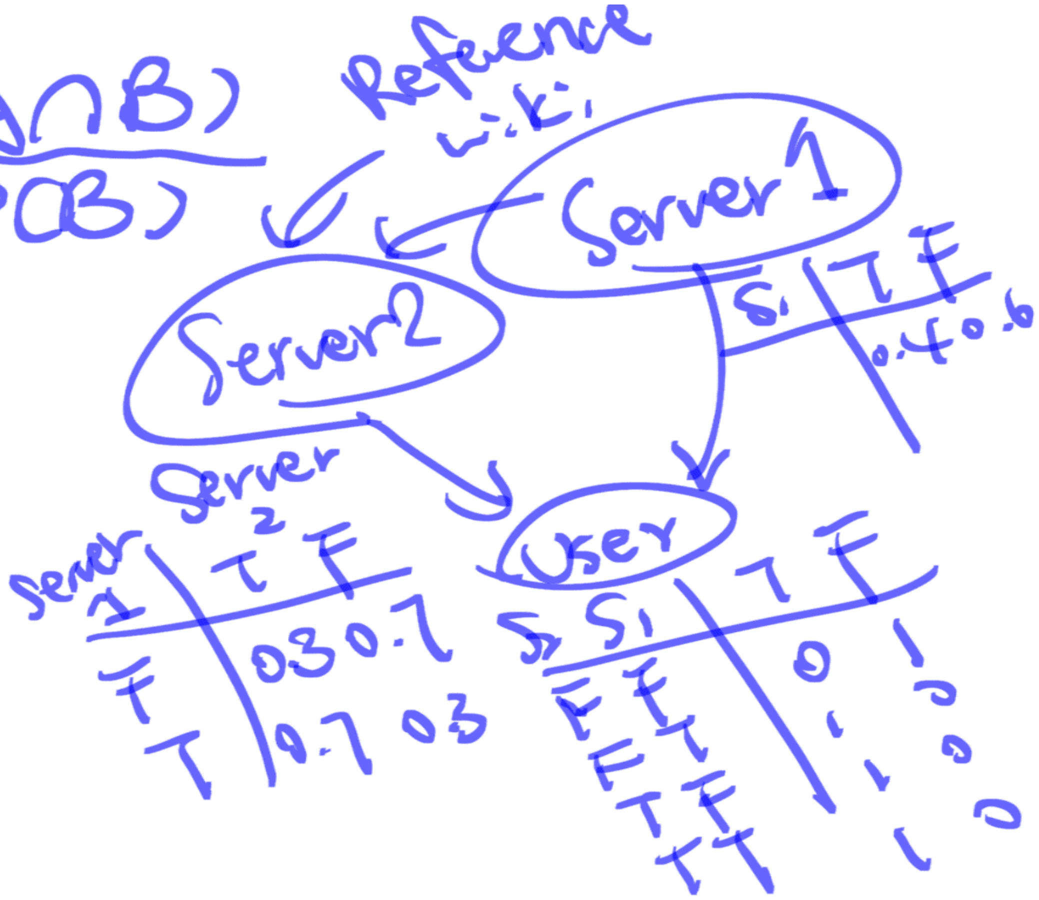
Comments: 22 pages, 8 figures, accepted to NIPS 2018
Subjects: **Machine Learning (stat.ML)**; Artificial Intelligence (cs.AI); Machine Learning (cs.LG); Methodology (stat.ME)
Cite as: arXiv:1803.01422 **[stat.ML]**
(or arXiv:1803.01422v2 **[stat.ML]** for this version)

該研究重新審視了一個名為 NOTEARS 的連續優化框架，想將其使用在貝葉斯網絡學習上。研究者首先將非循環性的現有代數特徵(existing algebraic characterizations of acyclicity)推廣到一類矩陣多項式(a class of matrix polynomials)上面，接著重點關注每邊一個參數的設置，結果表明，排除在不影響整體地的細節情況下，不能滿足 NOTEARS 公式的 Karush-Kuhn-Tucker (KKT) 最優性條件，並解釋該演算法的行為。

Intuition 直觉

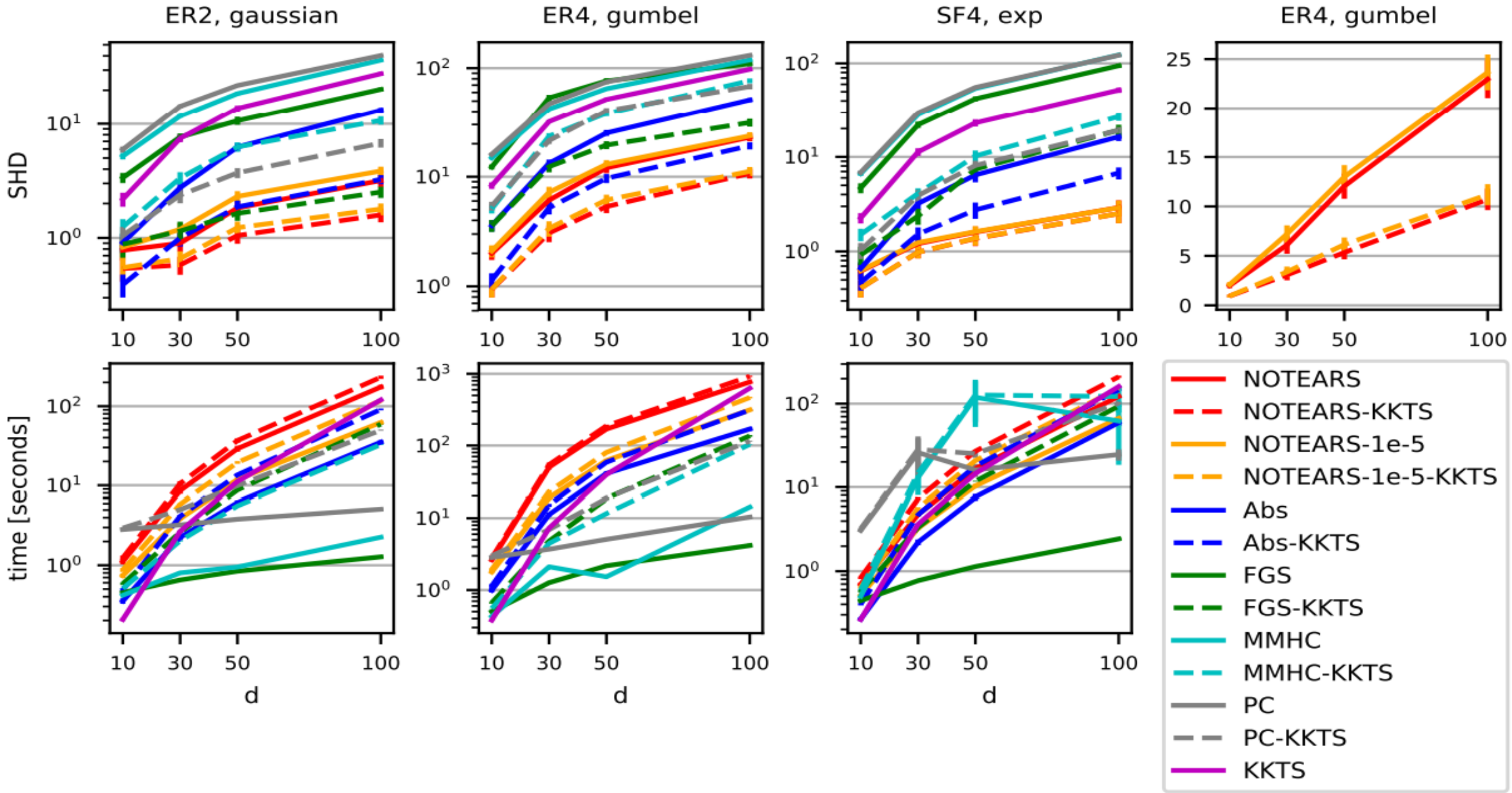
$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$





1. 貝葉斯定理；條件機率 (又稱後驗機率) 就是事件A在另外一個事件B已經發生條件下的發生機率。其條件機率表示為P(A|B)，讀作“在B條件下所發生A的機率”。P(A|B)=P(A∩B)/P(B)
2. 貝葉斯網絡(Bayesian network)，又稱信念網絡(Belief Network)，或有向無環圖模型(directed acyclic graphical model)，是一種機率圖模型。它是一種模擬人類推理過程中因果關係的不確定性處理模型，其網絡拓樸結構是一個有向無環圖(DAG)。
3. 假設有兩個伺服器 (S1,S2)，會傳送封包到使用者端（以U表示之）P(U, S1, S2) = P(U|S1, S2) * P(S2|S1)*P(S1)

Justification 理由



可以從該研究結果呈現的圖 1結構漢明距離 (SHD; Structural Hamming distances) 相對於真實圖和 n = 1000 的求解時間。誤差條表示 100 次試驗的標準誤差。在 SF4 SHD 圖中，紅線與橙色重疊。右上角的面板側重於使用線性垂直刻度與 NOTEARS 的組合。可以看到該研究首先關注基本演算法（實線），其中 NOTEARS 在 SHD 方面顯然是最好的。

Framework 框架

The development in this subsection suggests the meta-algorithm in Algorithm 1, which we refer to as KKT-informed local search. An instantiation is described in Section 4.2.

Algorithm 1 KKT-informed local search (KKTS)

Require: Initial set $\mathcal{Z}$ of edge absence constraints. Solve (10).

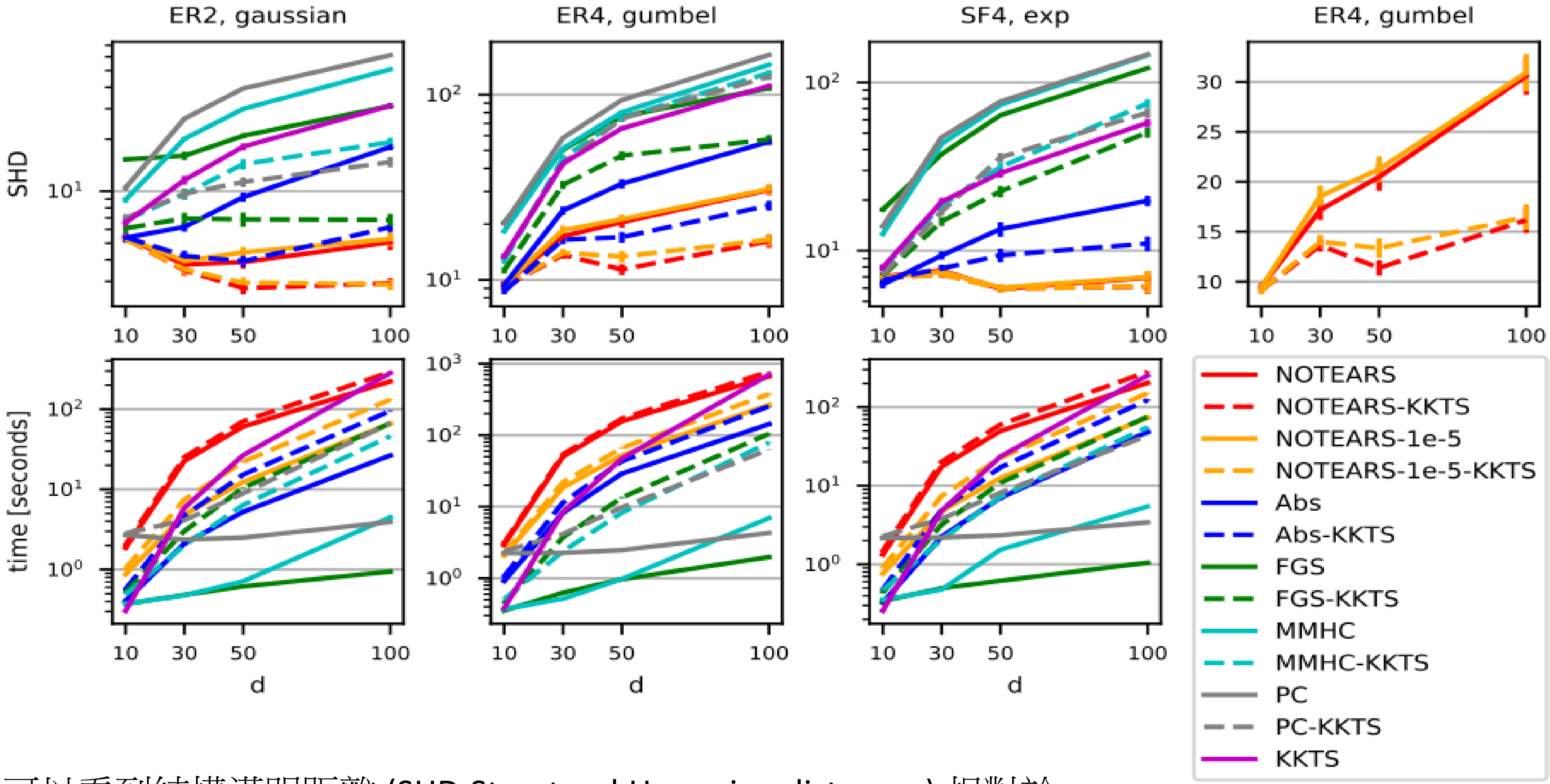
while $W^*(\mathcal{Z})$ infeasible **do**
    Select edge(s) in cycle $((W^*(\mathcal{Z}))_{ij} \neq 0, \left(\nabla h(A^*(\mathcal{Z}))\right)_{ij} > 0)$. Add to $\mathcal{Z}$. Re-solve (10).
end while
while $\mathcal{Z}$ reducible **do**
    Remove one or more unnecessary constraints $(i, j) \in \mathcal{Z}$ for which $\left(\nabla h(A^*(\mathcal{Z}))\right)_{ij} = 0$ (see Lemma 8). Re-solve (10).
end while

Theorem 9. *If $F(W)$ is separable, KKT-informed local search yields a solution satisfying the KKT conditions* (9).

元算法(the meta-algorithm in Algorithm 1)，研究者將其稱為 KKT 知情本地搜索(KKT-informed local search)。
可以從該研究第 4.2 節所描述的實例看到名為 KKT-informed local search (KKTS)演算法
定理 9。如果 F(W) 是可分離的，則 KKT 通知的局部搜索會產生滿足 KKT 條件 (9) 的解，
而研究的定理 7 和凸 F(W) 指出，定理 9 保證 KKT 通知的局部搜索將導致局部最小值。

When combined with Theorem 7 and a convex F(W), Theorem 9 guarantees that KKT-informed local search will result in local minima.

Result 结果



從研究結果結論的圖 2可以看到結構漢明距離 (SHD;Structural Hamming distances) 相對於真實圖和 n = 2d 的求解時間，在 SF4 SHD 圖中，紅線與橙色重疊，且可以看到在圖二右上角的面板側重於使用線性垂直刻度與 NOTEARS 的組合。