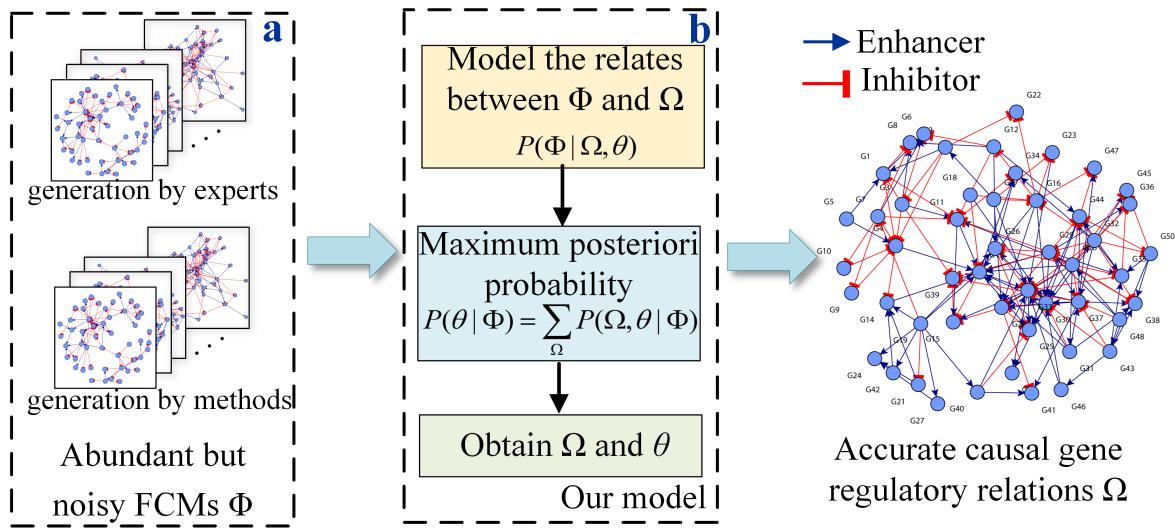


从噪声模糊认知图集中因果发现【公式详细解析】



Causal Discovery From Abundant but Noisy Fuzzy Cognitive Map Set

Yingzhi Teng, Kai Wu, and Jing Liu.

论文链接: <https://ieeexplore.ieee.org/abstract/document/10495183>

代码链接: <https://github.com/IngeTeng/Abundant-but-Noisy-FCMs>

一、研究动机:

- 近年来，涌现了大量基于模糊认知图（Fuzzy cognitive map, FCM）模型用于因果推断的方法（例如，推断基因调控网络）。然而，在一个真实的场景中，无法得知ground truth情况下，对于决策者而言，哪一种方法得到的FCM最能表示真实的因果关系？
- 受超参数的影响，单一方法也会产生不同性能的FCM。
- 另外，还有许多由不同领域专家，依靠专业知识构建的不同性能的FCM。

上述情况将会导致产生大量含噪声的候选因果关系集（FCMs）。

对于决策者而言，在没有先验知识的情况下，如何从上述大量含有噪声的数据集中，推断更真实的因果关系？这个问题首次得到关注，基于此，我们提出了一个基于贝叶斯和模糊理论的无参数模型，**CDFCM**。图1是CDFCM因果推断的示例。

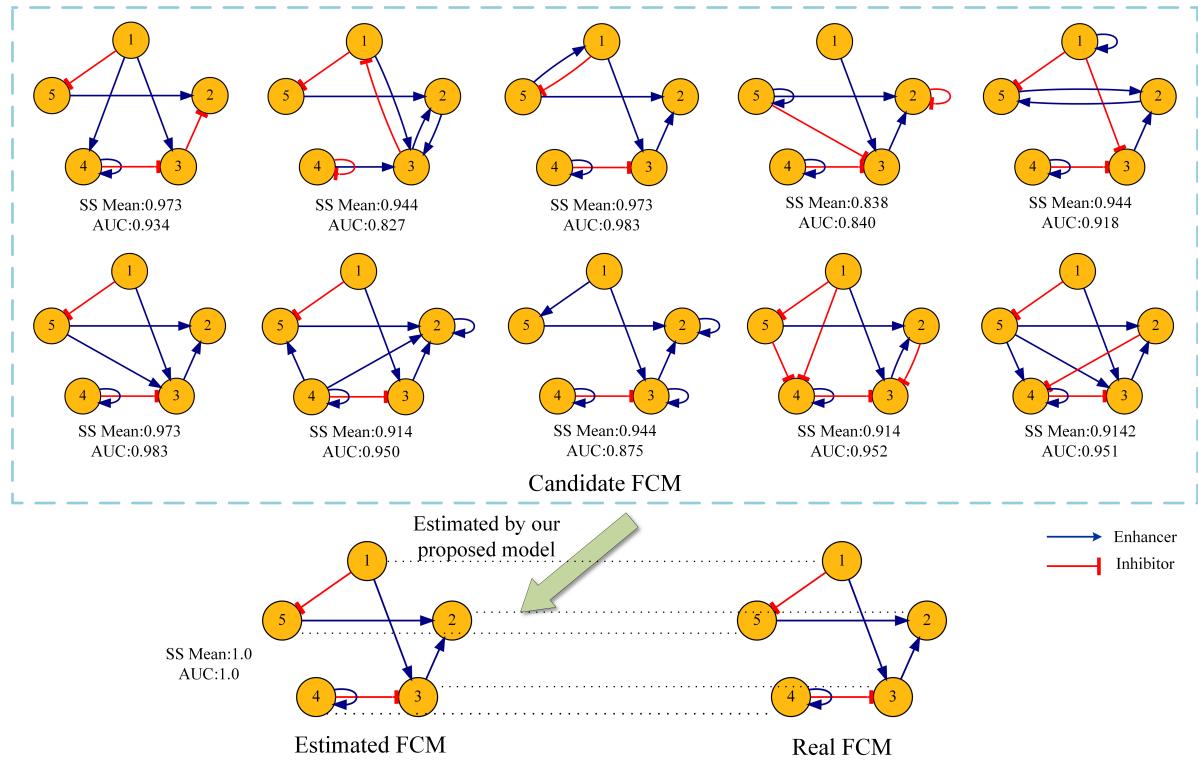


图1 CDFCM因果推断示例

关键词：模糊认知图，因果推断，贝叶斯理论，模糊理论

二、算法模型

1. 问题定义

我们假设观测数据（因果关系数据集） Φ 和真实因果（ground truth） Ω 之间可以用一个数据概率模型表示 $P(\Phi|\Omega, \theta)$ 。本工作旨在对候选FCM数据集 Φ 进行因果发现，以推断最优的因果关系 Ω 。我们算法的目的是通过计算最优的参数 θ 来确定这个概率模型。进而通过观测数据 Φ 与模型参数 θ 得到最接近真实（概率最大）的因果 Ω 。

2. 算法基本原理

首先根据贝叶斯公式，我们可以得到 $P(\Omega, \theta|\Phi)$ （从已知观测数据到待求参数的概率），

$$P(\Omega, \theta|\Phi) = \frac{P(\Phi|\Omega, \theta)P(\Omega)P(\theta)}{P(\Phi)} \quad (1)$$

总结所有可能的FCM真实因果结构 Φ ，我们得到 $P(\theta|\Phi) = \sum_{\Omega} P(\Omega, \theta|\Phi)$ ，我们通过寻找最大化的 $P(\theta|\Phi)$ ，即计算观测数据 Φ 下参数 θ 的最大后验概率，目的是为了得到最可能得参数 θ 。通过最大化 $P(\theta|\Phi)$ ，我们可以有效的建模ground truth Ω 和观测数据（即候选FCMs）之间的关联。最终，我们可以很容易的从候选FCMs中估计最优的因果关系。这个是我们文章的核心原理。

为了方便计算，我们最大化 $\log P(\theta|\Phi)$ ，利用Jensen不等式，

$$\log P(\theta|\Phi) = \log \sum_{\Omega} P(\Omega, \theta|\Phi) \geq \sum_{\Omega} q(\Omega) \log \frac{P(\Omega, \theta|\Phi)}{q(\Omega)} \quad (2)$$

其中 $q(\Omega)$ 表示任意 Ω 的分布概率，它满足 $\sum_{\Omega} q(\Omega) = 1$ ，即所有 Ω 的分布概率是为1的。当公式(2)中左侧与右侧相等时，右侧公式的值达到最大。此时

$$q(\Omega) = \frac{P(\Omega, \theta|\Phi)}{\sum_{\Omega} P(\Omega, \theta|\Phi)} \quad (3)$$

决定 $\log P(\theta|\Phi)$ 最大化的两个变量为 $q(\Omega)$ 和 $\log P(\Omega, \theta|\Phi)$ 。通过公式(3)可以最大化 $q(\Omega)$; 求 $\log P(\Omega, \theta|\Phi)$ 的最大值, 此时将 $q(\Omega)$ 看做常数, 即求 $\log P(\Omega, \theta|\Phi)$ 对于 θ 的偏导数。因此, 最大化 $\log P(\theta|\Phi)$ 时, θ 满足以下公式:

$$\sum_{\Omega} q(\Omega) \nabla_{\theta} \log P(\Omega, \theta|\Phi) = 0 \quad (4)$$

其中 $\nabla_{\theta} \log P(\Omega, \theta|\Phi)$ 表示 $\log P(\Omega, \theta|\Phi)$ 对 θ 的偏导数。

算法随机初始化 θ 后就在公式(3)和(4)之间迭代直至收敛。至此得到了模型性能最优时最大概率的 θ 值。一旦 θ 被确定, 就可以利用公式(3), 求得最优 Ω 的分布概率 $q(\Omega)$,

$$q(\Omega) = \frac{P(\Omega, \theta|\Phi)}{\sum_{\Omega} P(\Omega, \theta|\Phi)} = \frac{P(\Omega, \theta|\Phi)}{P(\theta|\Phi)} = P(\Omega|\Phi, \theta) \quad (5)$$

以上就是我们算法的基本原理。下面将介绍, 对应到FCM图中每个边是如何因果推断的。

3、算法细节

首先, 得到用以表示因果关系的FCM, 可能是带有权重的。在推断节点变量之间关系为“促进”或“抑制”之前, 我们需要对权重进行模糊化, 模糊为标签离散变量, 可以通过图2的成员函数, 本文中模糊为三种状态即“促进”、“抑制”和“不相关”。当然, 也可以直接通过数值截断的方式, 得到3个, 6个或更多的状态。

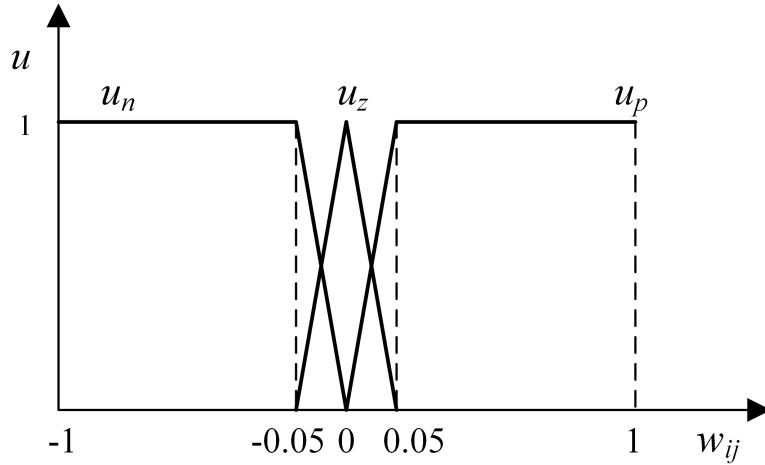


图2 3个文章中使用的成员函数

Ω 表示真实因果的邻接矩阵, 每个 Ω_{ij} 元素表示三种状态: 0-抑制、1-不相关和2-促进。参数 $\theta = \rho_0, \rho_1, \rho_2, \gamma_0, \gamma_1, \gamma_2$ 。其中 ρ_0, ρ_1 和 ρ_2 表示每条边三个状态的先验概率, 其中 $\rho_0 + \rho_1 + \rho_2 = 1$ 。 γ_0, γ_1 和 γ_2 表示每条边三个状态的观测概率。在 N 个候选FCM中, N_{ij} 为节点对 i, j 的总测量次数。 E_{ij} 表示观测到边存在的统计次数。假设对于每条边三个状态的观测概率 γ_0, γ_1 和 γ_2 都服从Bernoulli分布, 则

$$P(\Phi|\Omega, \theta) = \prod_{i,j} [\gamma_0^{E_{ij}} (1 - \gamma_0)^{(N_{ij} - E_{ij})}]^{I_{\Omega_{ij}=0}} \times [\gamma_1^{E_{ij}} (1 - \gamma_1)^{(N_{ij} - E_{ij})}]^{I_{\Omega_{ij}=1}} \times [\gamma_2^{E_{ij}} (1 - \gamma_2)^{(N_{ij} - E_{ij})}]^{I_{\Omega_{ij}=2}} \quad (6)$$

其中 $I(\cdot)$ 为指示函数

$$I_{conditions} = \begin{cases} 1 & \text{if conditions is True} \\ 0 & \text{if conditions is False} \end{cases} \quad (7)$$

根据每条边3个状态的先验概率 ρ_0, ρ_1, ρ_2 , Ω 的先验概率为

$$P(\Omega|\rho_0, \rho_1, \rho_2) = \prod_{i,j} \rho_0^{I_{\Omega_{ij}=0}} \rho_1^{I_{\Omega_{ij}=1}} \rho_2^{I_{\Omega_{ij}=2}} \quad (8)$$

我们假设 $\rho_0, \rho_1, \rho_2, \gamma_0, \gamma_1$ 和 γ_2 服从[0,1]均匀的先验概率分布。根据公式 (1) , 结合 (6) 和 (8) 可得

$$\begin{aligned} P(\Omega, \theta | \Phi) = & \frac{1}{\Phi} \prod_{i,j} [\rho_0 \gamma_0^{E_{ij}} (1 - \gamma_0)^{(N_{ij} - E_{ij})}]^{I_{\Omega_{ij}=0}} \\ & \times [\rho_1 \gamma_1^{E_{ij}} (1 - \gamma_1)^{(N_{ij} - E_{ij})}]^{I_{\Omega_{ij}=1}} \\ & \times [\rho_2 \gamma_2^{E_{ij}} (1 - \gamma_2)^{(N_{ij} - E_{ij})}]^{I_{\Omega_{ij}=2}} \end{aligned} \quad (9)$$

计算 $\log P(\Omega, \theta | \Phi)$

$$\begin{aligned} \log P(\Omega, \theta | \Phi) = & \sum_{i,j} \{ \\ & I_{\Omega_{ij}=0} [E_{ij} \log \gamma_0 + (N_{ij} - E_{ij}) \log(1 - \gamma_0)] \\ & + I_{\Omega_{ij}=1} [E_{ij} \log \gamma_1 + (N_{ij} - E_{ij}) \log(1 - \gamma_1)] \\ & + I_{\Omega_{ij}=2} [E_{ij} \log \gamma_2 + (N_{ij} - E_{ij}) \log(1 - \gamma_2)] \\ & + I_{\Omega_{ij}=0} \log \rho_0 + I_{\Omega_{ij}=1} \log \rho_1 + I_{\Omega_{ij}=2} \log \rho_2 \} \\ & - \log P(\Phi) \end{aligned} \quad (10)$$

我们根据基本原理部分的公式 (4) , 对 $\log P(\Omega, \theta | \Phi)$ 求关于 $\theta = \rho_0, \rho_1, \rho_2, \gamma_0, \gamma_1, \gamma_2$ 的偏导。以 γ_0 为例, 求公式 (10) 关于 γ_0 的偏导, 最大的后验概率 $\hat{\gamma}_0$ 满足

$$\sum_{\Omega} q(\Omega) \sum_{i,j} I_{\Omega_{ij}=0} \left(\frac{E_{ij}}{\hat{\gamma}_0} - \frac{N_{ij} - E_{ij}}{1 - \hat{\gamma}_0} \right) = 0 \quad (11)$$

计算求得 $\hat{\gamma}_0$

$$\hat{\gamma}_0 = \frac{\sum_{i,j} E_{ij} K_{ij}^{(0)}}{\sum_{i,j} N_{ij} K_{ij}^{(0)}} \quad (12)$$

其中 $K_{ij}^{(\tau)} = \sum_{\Omega} I_{\Omega_{ij}=\tau} q(\Omega)$ 是后验概率, $\Omega_{ij} = \tau, \tau = 0, 1, 2$ 。相似的, 计算所有参数 $\gamma_1, \gamma_2, \rho_0, \rho_1$ 和 ρ_2

$$\hat{\gamma}_{\tau} = \frac{\sum_{i,j} E_{ij} K_{ij}^{(\tau)}}{\sum_{i,j} N_{ij} K_{ij}^{(\tau)}} \quad (13)$$

$$\hat{\rho}_{\tau} = \frac{1}{\binom{n}{2}} \sum_{i,j} K_{ij}^{(\tau)} \quad (14)$$

至此, 我们得到了 θ 中包含的全部参数。回顾上述提到的基本原理, 我们可以根据 θ , 计算最接近ground truth的概率分布 $q(\Omega)$, 我们将公式 (9) 代入到 (5) 中

$$\begin{aligned} q(\Omega) = & \frac{\prod_{i,j} \prod_{\tau=0,1,2} [\hat{\rho}_{\tau} \hat{\gamma}_{\tau}^{E_{ij}} (1 - \hat{\gamma}_{\tau})^{N_{ij} - E_{ij}}]^{I_{\Omega_{ij}=\tau}}}{\sum_{\Omega} \prod_{i,j} \prod_{\tau=0,1,2} [\hat{\rho}_{\tau} \hat{\gamma}_{\tau}^{E_{ij}} (1 - \hat{\gamma}_{\tau})^{N_{ij} - E_{ij}}]^{I_{\Omega_{ij}=\tau}}} \\ = & \prod_{i,j} \frac{\prod_{\tau=0,1,2} [\hat{\rho}_{\tau} \hat{\gamma}_{\tau}^{E_{ij}} (1 - \hat{\gamma}_{\tau})^{N_{ij} - E_{ij}}]^{I_{\Omega_{ij}=\tau}}}{\sum_{\Omega_{ij}=0,1,2} \prod_{\tau=0,1,2} [\hat{\rho}_{\tau} \hat{\gamma}_{\tau}^{E_{ij}} (1 - \hat{\gamma}_{\tau})^{N_{ij} - E_{ij}}]^{I_{\Omega_{ij}=\tau}}} \\ = & \prod_{i,j} \prod_{\tau=0,1,2} K_{ij}^{(\tau)}^{I_{\Omega_{ij}=\tau}} \end{aligned} \quad (15)$$

其中

$$K_{ij}^{(\tau)} = \frac{\hat{\rho}_{\tau} \hat{\gamma}_{\tau}^{E_{ij}} (1 - \hat{\gamma}_{\tau})^{N_{ij} - E_{ij}}}{\sum_{\tau'} \hat{\rho}_{\tau'} \hat{\gamma}_{\tau'}^{E_{ij}} (1 - \hat{\gamma}_{\tau'})^{N_{ij} - E_{ij}}} \quad (16)$$

我们只需要根据公式 (13) (14) 和 (16) 不断迭代至收敛，则根据 (15) 得到最优的 Ω 。

三、实验及讨论

算法部分我们从时间序列数据中生成了15个因果数据集，来模拟“大量含有噪声”的多个场景。这些时间序列为合成与真实的基因调控网络数据 (GRN)。分别采用了AUC和SS_Mean作为评价指标，均与ground truth相比，越接近则评价指标越大。具体的实验设置细节可参考论文。

TABLE IV
EXPERIMENTAL RESULTS OF OUR PROPOSED METHOD ON DATASETS 1-7.

Dataset	Index	Time series data	$\sigma=0.001$			$\sigma=0.003$			$\sigma=0.005$		
			CDFCM	OWA	$\Phi(\text{mean} \pm \text{std.})$	CDFCM	OWA	$\Phi(\text{mean} \pm \text{std.})$	CDFCM	OWA	$\Phi(\text{mean} \pm \text{std.})$
1	AUC	synthetic_2	0.948	0.810	0.926 \pm 0.019	0.936	0.726	0.893 \pm 0.022	0.936	0.717	0.870 \pm 0.026
		synthetic_3	0.918	0.726	0.856 \pm 0.012	0.893	0.700	0.829 \pm 0.014	0.892	0.700	0.813 \pm 0.014
	SS_Mean	synthetic_2	0.945	0.944	0.907 \pm 0.023	0.885	0.868	0.866 \pm 0.028	0.885	0.881	0.839 \pm 0.031
		synthetic_3	0.891	0.839	0.829 \pm 0.014	0.850	0.837	0.798 \pm 0.016	0.829	0.700	0.781 \pm 0.016
2	AUC	synthetic_2	0.994	0.836	0.881 \pm 0.020	0.950	0.796	0.837 \pm 0.023	0.927	0.730	0.814 \pm 0.025
		synthetic_3	0.930	0.746	0.813 \pm 0.014	0.913	0.732	0.773 \pm 0.014	0.880	0.708	0.754 \pm 0.014
	SS_Mean	synthetic_2	0.983	0.909	0.784 \pm 0.035	0.926	0.877	0.701 \pm 0.042	0.840	0.798	0.659 \pm 0.045
		synthetic_3	0.891	0.814	0.829 \pm 0.014	0.838	0.754	0.585 \pm 0.027	0.791	0.738	0.544 \pm 0.028
3	AUC	synthetic_2	0.983	1.000	0.948 \pm 0.012	0.924	0.991	0.900 \pm 0.024	0.924	0.988	0.876 \pm 0.022
		synthetic_3	0.908	0.913	0.881 \pm 0.008	0.889	0.898	0.848 \pm 0.010	0.884	0.906	0.827 \pm 0.014
	SS_Mean	synthetic_2	0.964	0.826	0.911 \pm 0.020	0.916	0.635	0.862 \pm 0.023	0.892	0.521	0.840 \pm 0.024
		synthetic_3	0.886	0.760	0.850 \pm 0.009	0.856	0.555	0.805 \pm 0.013	0.835	0.353	0.775 \pm 0.013
4	AUC	synthetic_2	0.932	0.847	0.879 \pm 0.012	0.940	0.838	0.839 \pm 0.017	0.926	0.804	0.822 \pm 0.018
		synthetic_3	0.905	0.729	0.857 \pm 0.014	0.901	0.725	0.813 \pm 0.011	0.888	0.688	0.789 \pm 0.012
	SS_Mean	synthetic_2	0.924	0.882	0.824 \pm 0.028	0.924	0.836	0.730 \pm 0.042	0.916	0.826	0.688 \pm 0.045
		synthetic_3	0.867	0.811	0.775 \pm 0.026	0.856	0.826	0.683 \pm 0.022	0.847	0.783	0.633 \pm 0.024
5	AUC	synthetic_2	0.932	0.838	0.878 \pm 0.012	0.939	0.835	0.839 \pm 0.017	0.924	0.788	0.822 \pm 0.018
		synthetic_3	0.903	0.711	0.847 \pm 0.013	0.898	0.710	0.809 \pm 0.011	0.890	0.674	0.788 \pm 0.012
	SS_Mean	synthetic_2	0.924	0.882	0.821 \pm 0.028	0.916	0.836	0.728 \pm 0.043	0.908	0.826	0.689 \pm 0.044
		synthetic_3	0.856	0.798	0.755 \pm 0.026	0.853	0.793	0.675 \pm 0.023	0.847	0.773	0.629 \pm 0.046
6	AUC	synthetic_2	0.932	0.845	0.879 \pm 0.012	0.940	0.835	0.839 \pm 0.017	0.926	0.801	0.822 \pm 0.018
		synthetic_3	0.910	0.729	0.845 \pm 0.014	0.897	0.705	0.806 \pm 0.012	0.892	0.689	0.784 \pm 0.012
	SS_Mean	synthetic_2	0.924	0.882	0.824 \pm 0.028	0.924	0.836	0.729 \pm 0.042	0.924	0.826	0.688 \pm 0.046
		synthetic_3	0.872	0.788	0.749 \pm 0.029	0.855	0.780	0.665 \pm 0.024	0.845	0.751	0.620 \pm 0.025
7	AUC	synthetic_2	0.968	0.836	0.833 \pm 0.026	0.961	0.812	0.802 \pm 0.028	0.938	0.782	0.780 \pm 0.030
		synthetic_3	0.923	0.761	0.757 \pm 0.022	0.898	0.660	0.732 \pm 0.024	0.861	0.620	0.719 \pm 0.021
	SS_Mean	synthetic_2	0.926	0.925	0.732 \pm 0.028	0.840	0.840	0.689 \pm 0.031	0.792	0.769	0.661 \pm 0.029
		synthetic_3	0.823	0.818	0.594 \pm 0.029	0.723	0.694	0.555 \pm 0.027	0.671	0.653	0.636 \pm 0.066

数据集1-7 (GRN合成数据)：我们在生成因果数据集的时间序列中增加了不同水平的高斯噪声。从上表中可以看出，我们的算法 (CDFCM) 在大部分数据集上性能最优，且受噪声影响最小。

TABLE V
EXPERIMENTAL RESULTS OF OUR PROPOSED METHOD ON DATASETS 8-9.

Dataset	Index	EFCM	$\sigma=0.001;p=10\%$			$\sigma=0.001;p=20\%$		
			OWA	$\Phi(\text{mean} \pm \text{std.})$	EFCM	OWA	$\Phi(\text{mean} \pm \text{std.})$	EFCM
8	AUC	1.000	0.861	0.958 \pm 0.031	1.000	0.819	0.916 \pm 0.050	
		1.000	0.929	0.931 \pm 0.047	1.000	0.846	0.880 \pm 0.062	
9	AUC	1.000	0.895	0.904 \pm 0.029	1.000	0.801	0.860 \pm 0.047	
		1.000	0.929	0.800 \pm 0.071	1.000	0.686	0.715 \pm 0.106	
Dataset	Index	EFCM	$\sigma=0.003;p=10\%$			$\sigma=0.003;p=20\%$		
			OWA	$\Phi(\text{mean} \pm \text{std.})$	EFCM	OWA	$\Phi(\text{mean} \pm \text{std.})$	EFCM
8	AUC	1.000	0.812	0.910 \pm 0.046	0.967	0.832	0.859 \pm 0.061	
		1.000	0.889	0.869 \pm 0.057	0.889	0.800	0.815 \pm 0.072	
9	AUC	1.000	0.833	0.858 \pm 0.039	0.933	0.763	0.806 \pm 0.055	
		1.000	0.808	0.701 \pm 0.089	0.750	0.640	0.606 \pm 0.120	
Dataset	Index	EFCM	$\sigma=0.005;p=10\%$			$\sigma=0.005;p=20\%$		
			OWA	$\Phi(\text{mean} \pm \text{std.})$	EFCM	OWA	$\Phi(\text{mean} \pm \text{std.})$	EFCM
8	AUC	1.000	0.711	0.873 \pm 0.055	0.983	0.756	0.814 \pm 0.068	
		1.000	0.846	0.827 \pm 0.065	0.889	0.800	0.768 \pm 0.077	
9	AUC	1.000	0.772	0.828 \pm 0.048	0.950	0.764	0.759 \pm 0.066	
		1.000	0.800	0.649 \pm 0.100	0.750	0.615	0.521 \pm 0.135	

数据集8-9 (GRN合成数据) : 我们在生成因果数据集的时间序列中增加了不同水平的噪声和缺失值。从上表看出CDFCM具有很强的鲁棒性。

TABLE VI
EXPERIMENTAL RESULTS OF OUR PROPOSAL ON DATASETS 10-11.

Dataset	Index	CDFCM	$\sigma=0.001$		$\sigma=0.003$		$\sigma=0.005$			
			OWA	$\Phi(\text{mean} \pm \text{std.})$	OWA	$\Phi(\text{mean} \pm \text{std.})$	OWA	$\Phi(\text{mean} \pm \text{std.})$		
10	AUC	0.987	0.848	0.897 ± 0.032	0.983	0.798	0.849 ± 0.032	0.972	0.754	0.845 ± 0.036
	SS_Mean	0.939	0.885	0.843 ± 0.054	0.907	0.861	0.752 ± 0.083	0.900	0.816	0.742 ± 0.091
11	AUC	0.972	0.764	0.850 ± 0.022	0.956	0.714	0.814 ± 0.025	0.941	0.699	0.793 ± 0.028
	SS_Mean	0.887	0.807	0.771 ± 0.062	0.839	0.767	0.704 ± 0.080	0.826	0.755	0.665 ± 0.090

数据集10-11 (GRN合成数据) : 在此数据集中, 我们利用了6种不同的FCM学习方法来得到因果数据集, 模拟真实情况中, 对于决策者选择模型困难的问题。从上表中可以看出CDFCM的性能远高于这些FCM学习方法性能的均值。

TABLE VII

CAUSAL DISCOVERY ON DATASETS 12-13. WE INVESTIGATE THE CAUSAL DISCOVERY PROCESS USING A RANGE OF KEY PARAMETER VALUES, λ_i , RANGING BETWEEN 0 AND 0.05 IN INCREMENTS OF 0.001, WHERE $i = 1, 2, \dots, 40$. THE PARAMETER λ IS PARTITIONED INTO DISTINCT RANGES: $1 \sim 10$, $1 \sim 20$, $1 \sim 30$, AND $1 \sim 40$, EACH ENCOMPASSING SPECIFIC SUBSETS OF λ VALUES. FOR INSTANCE, EMPLOYING LASSO-FCM WITH λ_i GENERATES ONE CANDIDATE FCM, AND IF THE INDEX i CORRESPONDS TO $1 \sim 10$, THIS PROCESS RESULTS IN TEN CANDIDATE FCMs.

Dataset	Index	CDFCM	$i:1 \sim 10$		$i:1 \sim 20$		$i:1 \sim 30$		$i:1 \sim 40$				
			OWA	$\Phi(\text{mean} \pm \text{std.})$	CDFCM	OWA	$\Phi(\text{mean} \pm \text{std.})$	CDFCM	OWA	$\Phi(\text{mean} \pm \text{std.})$			
12	AUC	0.851	0.574	0.839 ± 0.030	0.861	0.571	0.855 ± 0.027	0.848	0.569	0.847 ± 0.030	0.803	0.567	0.825 \pm 0.049
	SS_Mean	0.865	0.839	0.816 ± 0.061	0.865	0.852	0.848 ± 0.054	0.884	0.871	0.849 ± 0.047	0.848	0.877	0.825 ± 0.059
13	AUC	0.797	0.592	0.765 ± 0.038	0.785	0.592	0.729 ± 0.046	0.763	0.592	0.700 ± 0.056	0.756	0.590	0.675 ± 0.065
	SS_Mean	0.784	0.744	0.705 ± 0.067	0.784	0.640	0.634 ± 0.087	0.784	0.599	0.572 ± 0.116	0.784	0.560	0.515 ± 0.142

数据集12-13 (GRN真实数据) : 此数据集的生成利用了Lasso-FCM方法和真实的基因调控网络数据, 模拟真实情况中, 不同超参数对性能带来的偏差影响。从上表可以看出, 在大多数参数范围内, CDFCM都可以表现出较高的性能。

TABLE VIII

CAUSAL DISCOVERY ON DATASET 14-15. λ_1 , λ_2 , AND α ARE THREE CRITICAL PARAMETERS THAT SIGNIFICANTLY AFFECT THE PERFORMANCE OF ONLINE-FCM. THE RANGE OF λ_1 , λ_2 , AND α IS $[2^{-5}, 2^5]$, AND THE INTERVAL IS 1. WE DIVIDE INTO FIVE GROUPS ACCORDING TO THE NUMBER OF PARAMETERS δ . FOR EXAMPLE, $\delta = 2$ SHOWS THAT λ_1 , λ_2 , AND α EACH CONTAIN TWO VALUES, AND THERE ARE EIGHT COMBINATIONS OF THE THREE PARAMETERS, WHICH RESULTS IN EIGHT CANDIDATE FCMS.

Dataset	Index	$\delta = 2$			$\delta = 3$			$\delta = 4$			$\delta = 5$			$\delta = 6$			
		$\lambda_1: 2^4 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-4};$ $\alpha: 2^4 \sim 2^5;$	$\lambda_1: 2^3 \sim 2^5;$ $\lambda_2: 2^{-4} \sim 2^{-3};$ $\alpha: 2^3 \sim 2^5;$	$\lambda_1: 2^2 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-2};$ $\alpha: 2^2 \sim 2^5;$	$\lambda_1: 2^1 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-1};$ $\alpha: 2^1 \sim 2^5;$	$\lambda_1: 2^0 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-0};$ $\alpha: 2^0 \sim 2^5;$	$\lambda_1: 2^4 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-4};$ $\alpha: 2^4 \sim 2^5;$	$\lambda_1: 2^3 \sim 2^5;$ $\lambda_2: 2^{-4} \sim 2^{-3};$ $\alpha: 2^3 \sim 2^5;$	$\lambda_1: 2^2 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-2};$ $\alpha: 2^2 \sim 2^5;$	$\lambda_1: 2^1 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-1};$ $\alpha: 2^1 \sim 2^5;$	$\lambda_1: 2^0 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-0};$ $\alpha: 2^0 \sim 2^5;$	$\lambda_1: 2^4 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-4};$ $\alpha: 2^4 \sim 2^5;$	$\lambda_1: 2^3 \sim 2^5;$ $\lambda_2: 2^{-4} \sim 2^{-3};$ $\alpha: 2^3 \sim 2^5;$	$\lambda_1: 2^2 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-2};$ $\alpha: 2^2 \sim 2^5;$	$\lambda_1: 2^1 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-1};$ $\alpha: 2^1 \sim 2^5;$	$\lambda_1: 2^0 \sim 2^5;$ $\lambda_2: 2^{-5} \sim 2^{-0};$ $\alpha: 2^0 \sim 2^5;$	
14	AUC	0.757 0.819	0.734±0.015 0.734±0.015	0.767 0.877	0.709±0.086 0.709±0.086	0.777 0.883	0.691±0.033 0.691±0.033	0.752 0.877	0.677±0.034 0.677±0.034	0.736 0.870	0.668±0.034 0.668±0.034	0.186 0.305	0.147 0.204	0.349 0.349	0.739±0.036 0.739±0.036	0.816 0.825	0.729±0.037 0.729±0.037
	SS_Mean	0.651 0.506	0.589±0.038 0.589±0.038	0.607 0.303	0.499±0.091 0.499±0.091	0.588 0.222	0.416±0.132 0.416±0.132	0.559 0.204	0.349±0.147 0.349±0.147	0.549 0.186	0.305±0.147 0.305±0.147						
15	AUC	0.812 0.706	0.781±0.012 0.728±0.028	0.814 0.698	0.802 0.558	0.766±0.026 0.684±0.068	0.814 0.690	0.810 0.461	0.752±0.033 0.632±0.104	0.816 0.668	0.816 0.397	0.739±0.036 0.583±0.130	0.816 0.678	0.816 0.357	0.729±0.037 0.544±0.144	0.825 0.357	0.729±0.037 0.544±0.144
	SS_Mean	0.706 0.678	0.728±0.028 0.684±0.068	0.698 0.558	0.684±0.068 0.684±0.068	0.690 0.461	0.632±0.104 0.632±0.104	0.668 0.461	0.597 0.397	0.583±0.130 0.583±0.130	0.678 0.357	0.544±0.144 0.544±0.144	0.825 0.357	0.825 0.357	0.729±0.037 0.544±0.144	0.825 0.357	0.729±0.037 0.544±0.144

数据集14-15 (GRN真实数据) : 此数据集的生成利用了Online-FCM方法和真实的基因调控网络数据, 从上表可以看出, 在更复杂的参数组合中, CDFCM都可以表现出较高的性能。

更多详细的性能分析与鲁棒性讨论可见论文。

结论

因果推理是人工智能的一个重要领域, 其特点是各种因果推理方法的涌入。然而这些方法各有优缺点, 给决策者提供更好的因果关系同时也带来了新的挑战。我们假设, 大量的因果推理方法可以产生丰富但有噪声的因果关系数据集。我们的论文没有建立新的因果发现的方法, 而是解决了一个新颖而关键的问题: 如何从这些丰富而含有噪声的数据集中推断出强大的因果关系?

为了解决这一挑战, 我们提出了一个基于贝叶斯和模糊理论的无参数模型, 旨在从候选因果关系数据集中准确估计真实系统中的因果关系。我们的研究结果证明了我们的方法在几个关键方面的有效性:

- 1)在噪声的数据环境中保持稳定性;
- 2)协调算法的多样性, 以获得精确的因果推断;
- 3)缓解超参数引起的波动。