AI SUSTEIN

An NSF EPSCoR Track-2 RII Program

Showcase #3

Causal Inference in Longitudinal Studies Using Causal Bayesian Network with Latent Variables

by Phat Huynh, Leah Irish, Arveity Setty, Om Yadav, Trung Q. Le

Background

Longitudinal studies have been broadly used in clinical research to investigate the associations between exposures or treatments and the outcome of the diseases, such as disease onset, subsequent morbidity, and mortality. However, few studies emphasize the causal relationships between observed variables and latent, time-varying confounders. The causal Bayesian network (CBN) shows promise in handling multiple causes and effects. There are two critical issues hinder CBNs from moving towards "absolute" causality including: (1) non-instance-specific causal structure learning and (2) the identification difficulty of latent confounding variables (LVs). To address these limitations, the paper presents an extension of the Bayesian Network for Latent Variable (BN-LV) framework that quantify the causal effects of the latent variables in CBNs by imposing various constraints for the identification of latent structures and the structure learning algorithms. Specifically, we adopted the BN-LVs method to locate LVs and estimate their values in the model structure and applied traditional causal inference techniques to validate the causal structure. Specifically, we extended the BN-LVs framework to apply to longitudinal studies by extracting the temporal ordering of the exposures, outcomes, and confounders. We then sorted all time-varying variables by their temporal order and imposed a constraint in structure learning algorithms allowing only the preceding variables to cause subsequent ones. The constraints significantly reduce the complexity of the structure learning task in BN-LVs methods. To learn instance-specific causal mechanisms and validate the learned CBN, we applied unit-level causal inference methods after the CBN learned the structure. We validated the model using the case study: Temporal Associations Between Daytime Napping and Sleep Outcomes.

Methodology

The method executes in 3 main stages: model construction, structural model discovery, and unit-level causal inference. In the first stage, the latent variable identification (LVI) algorithm identifies the LVs inductively, given the measurement items in an exploratory mode. The LVI tests the independence axiom. This axiom asserts that measurement items associated with the same LV are supposed to be caused by that LV, thus they should be conditionally independent of each other given the LV. In the second stage, we discover the most parsimonious causal structure involving LVs and search for the best structure among all candidate graphs using a constraint-based PC algorithm. In the last stage, the unit-level causal inference was performed to validate the learned structure.

Results and Discussion

We systematically evaluate our proposed method in a case study investigating the temporal associations between daytime napping and sleep outcomes. The case study further extends our previous works on detecting and predicting of cardiorespiratory acute disorder's biomarkers. In the first step of the LVI algorithm, we need to specify all sets of measurement variables that satisfy the independence axiom. In the second step, we pruned the generated measurement item sets into disjoint sets. We performed 3 different Bayesian network structure learning algorithms, namely Hill Climbing, PC algorithm, and Max-Min Hill-Climbing, for constructing the CBN without LVs from the observed variables in our data. For comparison, the CBNs without and with LVs learned from those algorithms were illustrated in Figure 1 and Figure 2 respectively.

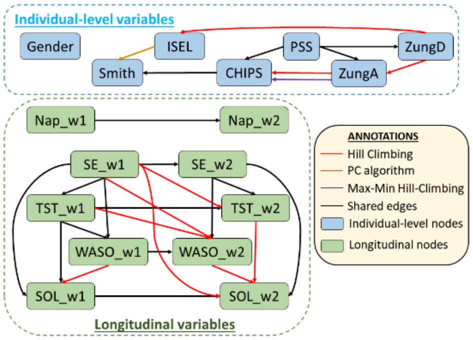

Figure 1. Learned CBN from the measurement items using the score-based method — Hill Climbing, the constraint-based method — PC algorithm, and the hybrid method — Max-Min Hill-Climbing.

We continued to perform the stable PC algorithm to learn the CBN with latent structure after having the outputs from the LVI algorithm. The learned structure with LVs and the optimal weights $\boldsymbol{\alpha}^*$ were illustrated in Figure 2.

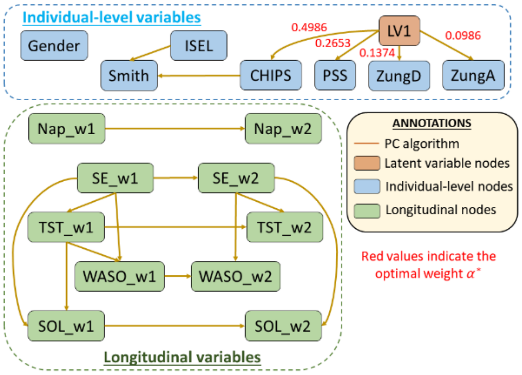

Figure 2*. Learned CBN from the measurement items with one latent variable* $LV1$ *using the stable PC algorithm.*

To justify “near” causality of the learned CBN, we applied the causal inference methods introduced in the methods. The detailed results were illustrated in Table 1.

Table 1*. Estimated Unit-level Average Causal Effects of Napping Time on 4 Sleep Outcomes for 4 Shortlisted Subjects*

Subjects	TST		SE	
	p-value	ACE	p-value	ACE
Subject 1	0.26	16.66 (+)	0.46	-3.69 (-)
Subject 2	0.38	16.75 (+)	0.13	-3.55 (-)
Subject 3	0.04	74.20 (+)	0.83	0.52 (+)
Subject 4	0.71	20.50 (+)	0.67	-3.09 (-)
	WASO		SOL	
	p-value	ACE	p-value	ACE
Subject 1	0.41	9 (+)	0.12	14.33 (+)
Subject 2	0.45	3.50 (+)	0.96	0.25 (+)
Subject 3	0.15	14.60 (+)	0.38	5.80 (+)
Subject 4	0.22	19.75 (+)	0.29	24.75 (+)

For more information, please refer to: https://ieeexplore.ieee.org/xpl/conhome/1000626/all-proceedings.

AI SUSTEIN

Showcase #3

An NSF EPSCoR Track-2 RII Program

The showcase paper will be published in the **2022 Annual Reliability and Maintainability Symposium (RAMS) IEEE conference proceedings**.