A Review of Multiple Causal Inference

Mingzhang Yin

May 15, 2020

1 The Multiple-cause Confounding and Deconfounder

In this section, we review the multiple causal inference paper by Wang and Blei [10]. We aim to elaborate the setting of multiple-cause inference, the intuition and realization of proposed methodology and the assumptions behind.

(a) Multiple causes setting

The paper studies how to infer the causal effects of many causes/treatments in the presence of unobserved confounders. Unlike traditional causal inference which usually considers a single possible cause, the multiple causal inference assumes there are m possible causes, represented as a vector $\boldsymbol{a} = (a_1, \cdots, a_m)$. The potential outcome for unit i under causes $\boldsymbol{a}_i$ is $Y_i(\boldsymbol{a}) \in \mathbb{R}$ and the observed outcome is $y_i(\boldsymbol{a}_i)$. There are measured pre-treatment covariates x_i of unit i. The observed data is $\mathcal{D} = \{\boldsymbol{a}_i, y_i(\boldsymbol{a}_i), x_i\}_{i=1}^n$. The goal is to estimate the average treatment effect of a particular array of causes $\mu = \mathbb{E}[Y_i(\boldsymbol{a})]$.

The authors consider a practical setting that there are unmeasured confoundings, which influence both cause and outcome. Consequently,

$$\mathbb{E}[Y_i(\boldsymbol{a})] \neq \mathbb{E}[Y_i(\boldsymbol{a})|A_i = \boldsymbol{a}] \tag{1}$$

where the RHS is what we can estimate from the observational data but the LHS is what we want to estimate. However, if we have observed confounders Z_i, hence the weak unconfoundedness $A_i \perp\!\!\!\perp Y_i(\boldsymbol{a})|Z_i$, then the average effect can be estimated as

$$\mathbb{E}[Y_i(\boldsymbol{a})] = \mathbb{E}[\mathbb{E}[Y_i(\boldsymbol{a})|A_i = \boldsymbol{a}, Z_i]] = \int \mathbb{E}[Y_i|A_i = \boldsymbol{a}, Z_i]p(Z_i)dZ_i \tag{2}$$

(b) Deconfounder and its intuition

The paper aims to infer the unobserved confounder as latent variable by studying the association relationship between multiple cause. The authors claim that "the confounders are effectively observed, even if not explicitly so, and embedded in the multiplicity of the causes. [11]" A factor model is used as the tool to extract this confounder information from

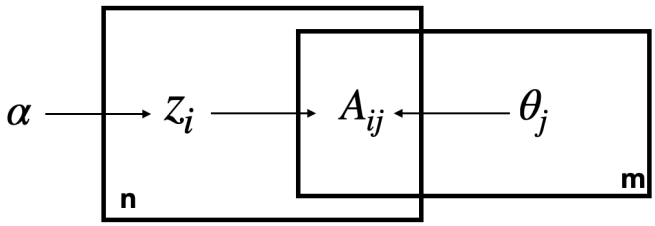

Figure 1: Illustration of the factor model

the causes. In the context of multiple causes, a factor model assumes the observed cause j of unit i is generated from

$$A_{ij} \sim p(\cdot|z_i, \theta_j); \quad Z_i \sim p(\cdot|\alpha) \tag{3}$$

where Z_i is per-data point level latent variable (a.k.a. factor loading) and θ_j is per-factor level parameter (a.k.a. factor score). The graphical model of a typical factor model is shown in Figure 1. Simple examples for factor model include: probabilistic PCA, $A_{ij} \sim \mathcal{N}(z_i^T \theta_j, \sigma^2)$; Poison factor model $A_{ij} \sim Pois(z_i^T \theta_j)$.

The proposed deconfounder is motivated by strong intuition. Assume the fitted factor model captures the marginal distribution of causes $p(\boldsymbol{a})$, then all causes are conditionally independent given the local latent factors

$$p(\boldsymbol{a}_i|z_i) = \prod_{j=1}^{m} p(a_{ij}|z_i). \tag{4}$$

The paper then claims that the inferred latent factor loading vector $\hat{Z}_i = \mathbb{E}_M[Z_i|A_i = \boldsymbol{a}_i]$ can be used as a substitute confounder (To get formal weak unconfoundedness $A_1, \cdots, A_m \perp\!\!\!\perp Y(\boldsymbol{a})|X, \hat{Z}$ rests on further assumptions, as discussed in the next section). The intuition for this claim is that "the multiple-cause confounders induce dependence among the causes; if a good factor model provides a variable that renders the causes conditionally independent then that variable captures the confounders [10]". A proof by contradiction in an intuitive sense is shown in Figure 2: if there exists an unobserved multi-cause confounder, then the conditional independence cannot hold. Here it needs to assume that there are no unobserved single-cause confounders, which are variables that affect one cause and the potential outcome.

The deconfounder algorithm is a two-stage process:

(i) Use a factor model to estimate substitute confounder from assigned causes as $\hat{z}_i = \mathbb{E}_M[Z_i|A_i = \boldsymbol{a}_i]$;

(ii) Check that the factor model captures their population distribution;

(iii) (Use an outcome model to) estimate the conditional expectation $\mathbb{E}[Y_i|A_i = \boldsymbol{a}, Z_i = \hat{z}_i, X_i = x_i]$ and estimate causal effects by stratification.

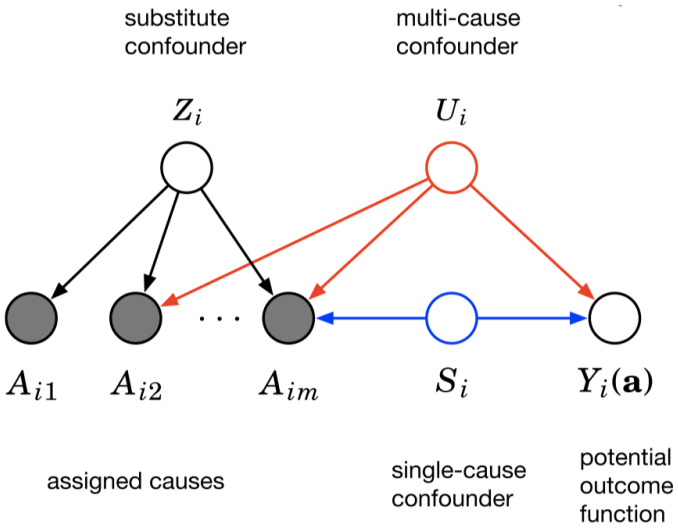

Figure 2: Illustration of the multiple causes

(c) Deconfounder assumptions

Thus, the deconfounder relies on three main assumptions: i) sutva [9]; ii) no unobserved single-cause confounders; iii) overlap [6]. The last two assumptions worth further scrutiny.

The "No unobserved single-cause confounders" assumption is weaker than the traditional weak unconfoundedness assumption for causal identification, and is considered as a major contribution of this paper. This assumption is refined and elaborated in Wang and Blei [12], which consists of two sub-parts.

Definition 1 (Multi-cause separator)**.** *A multi-cause separator Z is a smallest σ-algebra that renders all the causes conditionally independent,*

$$p(A_1, \cdots, A_m|Z) = \prod_{j=1}^{m} p(A_j|Z)$$

and none of the conditionals $p(A_j|Z)$ is a point mass.

Assumption 1 (No unobserved single-cause confounders)**.** *There exists a random variable Z that satisfies the following two requirements:*

- *It is a multi-cause separator;*
- *Weak unconfoundedness: $A_1, \cdots, A_m \perp\!\!\!\perp Y(\boldsymbol{a})|Z, X, \ \forall \boldsymbol{a}$*

The smallest σ-algebra ensures Z does not contain single cause confounder and the weak unconfoundedness ensures Z contains all the multi-cause confounders.

The "overlap" condition requires that given the confounder, the conditional probability of any vector of assigned causes is positive, i.e.

$$p(A_{ij} \in \mathcal{A}|Z_i = z_i) > 0, \quad \forall \mathcal{A} \text{ with positive measure} \tag{5}$$

If the overlap condition is violated, there are no observations at some combinations of causes and (substitude) confounders.

Capitalizing on Eq. (2),

$$\mathbb{E}[Y_i(\boldsymbol{a})] = \int \mathbb{E}[Y_i|A_i = \boldsymbol{a}, Z_i]p(Z_i)dZ_i \approx \frac{1}{n}\sum_{i=i}^{n} \mathbb{E}_Y[Y|A_i = \boldsymbol{a}, Z_i = z_i] \tag{6}$$

In deconfounder algorithm, to estimate the expected outcome, a parametric outcome model

$$\mathbb{E}_Y[Y|A_i = \boldsymbol{a}, Z_i = z] = f_{\boldsymbol{\beta}}(\boldsymbol{a}, z)$$

is trained on $\{y_i(\boldsymbol{a}_i), \boldsymbol{a}_i, \hat{z}_i\}_{i=1}^n$ but in Eq. (6), we need regression on $\{\boldsymbol{a}, \hat{z}_i\}_{i=1}^n$. When overlap condition is violated, we make predictions based on extrapolation, hence only a certain set of causal effects are identifiable [5].

Elaborated in Wang and Blei [12], to achieve the weak unconfoundedness, it requires an additional assumption

Assumption 2 (Deterministic substitute confounder)**.** *All multi-cause separators Z are pinpointed by a single deterministic function of the causes,*

$$p(Z|A_1, \cdots, A_m) = \delta_{f(A_1,\cdots,A_m)}$$

Under Assumptions 1, 2, a measured multi-cause separator Z leads to weak unconfoundedness

$$A_1, \cdots, A_m \perp\!\!\!\perp Y(\boldsymbol{a})|X, Z$$

2 Problems and Responses

The paper receives active feedback; some representative ones are: D'Amour [2], D'Amour [3], Imai and Jiang [5], Athey et al. [1], Ogburn et al. [8], and the authors' response in Wang and Blei [11, 12]. We summarize the discussions from three aspects: identification, no single cause confounder and overlap.

(a) Identification

The identification problem is mainly discussed in D'Amour [2], D'Amour [3] and in Imai and Jiang [5]. First, let's define what is the causal identification. In general, the mapping from observational data to causal relationship is a one-to-many mapping. That's to say, in the presence of unobserved confounding, the observed data distribution is compatible with many potentially contradictory causal explanations. When this is the case, we say that the causal quantity of interest, or estimand, is *not identified*. when the causal estimand can be written entirely in terms of observable probability distributions, we say the estimand is *identified*.

D'Amour [2] gives an example that the variable Z such that

$$p(A_1, \cdots, A_m|Z) = \prod_{j=1}^{m} p(A_j|Z)$$

is not unique. Then different multi-cause separators result in different causal estimations. So the Assumption 1 alone does not guarantee identification.

D'Amour [3], Imai and Jiang [5] further show, with an extended Assumptions 1, 2, the overlap assumption is violated (Wang and Blei [12] emphasizes that the weak unconfoundedness is not). When the factor model $p(\hat{Z}_i|A_i)$ degenerates, the estimand

$$\mathbb{E}[\mathbb{E}[Y_i(\boldsymbol{a})|A_i = \boldsymbol{a}, \hat{Z}_i]] = \int \mathbb{E}[Y_i|A_i = \boldsymbol{a}, \hat{Z}_i]p(\hat{Z}_i|A_i = \boldsymbol{a}_i)d\hat{Z}_i \tag{7}$$

is in general different from Eq. (2), pointed out by Imai and Jiang [5]. As a result, even if a factor model is uniquely identified and the weak unconfoundedness is achieved, without overlap condition, the nonparametric identification is nevertheless impossible.

D'Amour [3] summarizes this dilemma as "the causes A cannot be used simultaneously as measurements of the unobserved confounder Z, and as treatments whose effects are being estimated." D'Amour [2] provides the following theorem

Theorem 1. *Under Assumption 1 and assume $p(Z, A) = p(A|Z)p(Z)$ is unique, $p(Y(\boldsymbol{a}))$ is not identified, except in trivial case $p(Y(\boldsymbol{a})) = p(Y|\boldsymbol{a})$.*

Solutions:
A consensus is to add parametric assumptions to the outcome model $p(Y_i(\boldsymbol{a})|Z_i)$; though these assumptions may limit the practical applicability of the deconfounder method and rely on strong a priori knowledge.

Wang and Blei [10] lives at one extreme: they assume uniqueness in confounder and forgo all-causes overlap. To achieve causal identification in Theorem 6-8 in [10], they impose additional restrictions on the factor model and outcome model, or reduce the range of cause values to a subset of causes or causes with the same substitute confounder.

Imai and Jiang [5] adds parametric assumptions to $p(Y_i(\boldsymbol{a})|Z_i)$, which requires a linear and separable form.

The current known identification conditions and results are summarized in Wang and Blei [11].

Conclusion:
In sum, Assumption 1 alone neither induces weak unconfoundedness nor causal effect identification [2]. Assumptions 1, 2 induces weak unconfoundedness but not identification, due to the loss of overlap condition [2, 5]. Assumptions 1, 2 + parametric assumption / reducing causes range lead to the identification [10]. (The authors suggest that " not using the deconfounder with causally dependent causes" [11]; Is it an additional assumption?)

(b) No single cause confounder assumption

Ogburn et al. [8] claims the multiple causal inference paper [10] is incorrect. It summarizes the assumption 1, 2 to be
"If $\exists$ U, s.t. $A_j \perp\!\!\!\perp Y(\boldsymbol{a})|U$, $A_j \perp\!\!\!\perp A_{\setminus j}|U$ and we find Z s.t. $A_j \perp\!\!\!\perp A_{\setminus j}|Z$, then $A_1, \cdots, A_m \perp\!\!\!\perp W|Z$"
and gives some counterexamples.

However, this summarization is not a complete statement of the original assumptions 1, 2 that are needed for weak unconfoundedness. There are plenty of examples to show the claim

in this summary is not true, such as the ones in Figure 2 of Imai and Jiang [5]. Without the σ-algebra statement, the partial summary cannot distinguish a single multicause confounder and the multiple single-cause confounders it decomposes to, so it cannot exclude single cause confounders. Without Assumption 2, the weak unconfoundedness is not true. Hence the examples in Ogburn et al. [8] are counterexamples to the partial summary, but not to the original "no single cause confounder" assumption. (the statement in Counterexample 2 itself seems not correct, i.e. $(A_1, A_2) \perp\!\!\!\perp W|Z$ in the given example).

(c) Overlap assumption

As summarized in Section (a), overlap assumption is violated when the factor model degenerates. Lack of complete overlap is often considered a problem more serious than imbalance [4]. When overlap assumption is violated, additional assumptions, such as parametric assumptions in Section (a), are necessary for identification.

3 Future Directions

One important lesson I learned from these discussions is that there is no free lunch. One assumption that is weakened often means other assumptions need to be strengthened. But creating multiple versions of assumptions gives choices in practice to match the real condition to the most suitable assumption.

From theoretical perspective, we can study other assumptions that guarantee the identification for multiple causal inference. For example, when overlap is stringent, can we identify causal effects on different target population, as the one defined in Li et al. [7]?

From methodological perspective, when the substitute confounder is not pinpointed, can we identify the feasible region of causal effects? Is it beneficial to use more flexible factor models, such as the one in Yin and Zhou [13]? Can we force overlapping by adding regularization in the factor model? Can we solve the problems such as the one in Athey et al. [1] where multiple causes naturally arise?

References

[1] Susan Athey, Guido W Imbens, and Michael Pollmann. Comment on:"the blessings of multiple causes" by yixin wang and david m. blei. *Journal of the American Statistical Association*, 114(528):1602–1604, 2019.

[2] Alexander D'Amour. On multi-cause causal inference with unobserved confounding: Counterexamples, impossibility, and alternatives. *arXiv preprint arXiv:1902.10286*, 2019.

[3] Alexander D'Amour. Comment: Reflections on the deconfounder. *Journal of the American Statistical Association*, 114(528):1597–1601, 2019.

[4] Andrew Gelman and Jennifer Hill. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press, 2006.

[5] Kosuke Imai and Zhichao Jiang. Comment: The challenges of multiple causes. *Journal of the American Statistical Association*, 114(528):1605–1610, 2019.

[6] Kosuke Imai and David A Van Dyk. Causal inference with general treatment regimes: Generalizing the propensity score. *Journal of the American Statistical Association*, 99 (467):854–866, 2004.

[7] Fan Li, Kari Lock Morgan, and Alan M Zaslavsky. Balancing covariates via propensity score weighting. *Journal of the American Statistical Association*, 113(521):390–400, 2018.

[8] Elizabeth L Ogburn, Ilya Shpitser, and Eric J Tchetgen Tchetgen. Counterexamples to" the blessings of multiple causes" by wang and blei. *arXiv preprint arXiv:2001.06555*, 2020.

[9] Donald B Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical association*, 75(371):591–593, 1980.

[10] Yixin Wang and David M Blei. The blessings of multiple causes. *Journal of the American Statistical Association*, pages 1–71, 2019.

[11] Yixin Wang and David M Blei. The blessings of multiple causes: Rejoinder. *Journal of the American Statistical Association*, 114(528):1616–1619, 2019.

[12] Yixin Wang and David M Blei. Towards clarifying the theory of the deconfounder. *arXiv preprint arXiv:2003.04948*, 2020.

[13] Mingzhang Yin and Mingyuan Zhou. Semi-implicit variational inference. In *International Conference on Machine Learning*, pages 5646–5655, 2018.