Leveraging Deep Generative Model For Causal Effect Estimation in Healthcare

Yushin Kim
Computer Science & Engineering
Major in Bio-Artificial Intelligence
Hanyang University
Ansan, South Korea
hpwgg045@hanyang.ac.kr

Sejong Lee
Computer Science & Engineering
Major in Bio-Artificial Intelligence
Hanyang University
Ansan, South Korea
kingsaejong@hanyang.ac.kr

Sunghyun Cho
Computer Science & Engineering
Hanyang University
Ansan, South Korea
chopro@hanyang.ac.kr

***Abstract*—Deep generative models have risen to prominence in diverse domains, including healthcare. In particular, their application in causal effect estimation has the potential to drive significant advancements in personalized medicine. In this study, we conducted an empirical analysis to investigate the impact of selection bias on continuous treatment effect estimation using deep generative models. Our results demonstrate that the presence of selection bias can lead to estimation performance disparities of up to 8 to 9 times. Such significant deviations pose severe risks in medical applications, where accurate treatment effect estimation is crucial for patient outcomes. Therefore, this paper underscores the criticality of addressing these challenges and proposes directions for future research to ensure the robust application of AI in healthcare.**

***Index Terms*—Artificial Intelligence, Causal Effect Estimation, Deep Generative Learning**

I. Introduction

In healthcare, precise estimation of treatment effects is vital. While treatments often have fixed doses or categories, many scenarios involve continuous treatment where dosages or interventions vary in a continuous spectrum [1]. This continuous nature poses a challenge: it increases the complexity of estimating effects due to the vast range of potential dosages and their respective outcomes [2]. Furthermore, the need for accurate estimation becomes even more crucial when considering the wide array of patient responses based on individual health profiles and conditions.

Deep learning algorithms, especially generative models like Generative Adversarial Networks (GANs), offer a promising avenue for addressing the challenges of continuous treatment effect estimation [3]. The core principle of GANs is a duel between a generator, which creates synthetic data samples, and a discriminator, which tries to distinguish between genuine and synthetic samples. As the training progresses, the generator becomes adept at producing increasingly convincing fake outcomes. In the context of treatment effect estimation, these fake outcomes represent potential counterfactual results, becoming more aligned with what might have been observed in real-world scenarios. This iterative refinement makes GANs exceptionally suited for modeling the intricacies of continuous treatments, offering a dynamic method to bridge the gap between observed and unobserved potential outcomes.

In this paper, we explore the potential applications of deep generative model based causal effect estimation in the medical field. We conduct an empirical study to utilize the latest causal effect estimation techniques and analyze the results. In addition, from a medical artificial intelligence perspective, we present the future direction for causal effect estimation technology to assist clinicians.

II. Deep Learning-based Causal Effect Estimation

Efforts to estimate causal effects using deep learning are diverse, depending on the treatment's form and assumptions. At its fundamentals, most research employs a binary treatment to assess whether an intervention is applied [4]. There are also studies that consider continuous treatment to estimate outcome curves based on intervention intensity [5]. When multiple interventions are possible, research might assume multiple treatments to identify the optimal treatment [6]. There's also a focus on models robust to biases from hidden confounding variables [7]. Amidst various research directions, this paper emphasizes the potential of continuous treatment effect estimation techniques, aiming to derive precise drug-response curves for patients. To this end, we investigate SCIGAN, a state-of-the-art model for continuous treatment effect estimation that assumes unconfoundedness [8].

SCIGAN introduces a novel approach to estimate the dose-response curve using a deep generative model. The generator of SCIGAN takes factual data as input, producing counterfactual outcomes for arbitrary treatments and dosages. Generated counterfactual outcomes, along with the real factual outcomes, are paired as (dosage, outcome) and provided to the Discriminator, which aims to classify the factual outcomes. This Discriminator contains two hierarchically structured models trained to classify treatment types and dosages respectively. Once trained, the generator can produce counterfactual dose-response functions for given factual data, comprising features, treatment, dosage, and outcome. Leveraging these generated counterfactual datapoints, a separate inference network can then be trained, enabling the inference of potential outcomes for any set of features. SCIGAN also provides a theoretical analysis, underscoring the utility of the GAN framework and the hierarchical discriminator.

III. Empirical Study

In this empirical study, we utilize SCIGAN to evaluate its ability to estimate simple treatment effects. While SCIGAN is originally designed to support multiple treatments, we narrow our focus to continuous treatment, measuring its performance exclusively in a single treatment variable context. Adopting a similar experimental setup to SCIGAN, we extract 16 features from the MIMIC-III dataset, which includes metrics like SpO2 and tidal volume. The treatment variable was assigned values between 0 and 1, based on a beta distribution with the mode representing the optimal dosage ($\alpha = 3$). We simulate two diverse dose-response curves: a sine wave and a thresholded linear function.

Figure 1 presents a comparison of the ground-truth dose-response functions and SCIGAN's predictions. Upon close examination, particularly around the optimal dosage—the crest for the sine wave and the value of 1 for the thresholded linear function—it is evident that SCIGAN's predictions are remarkably close to the actual values. The predicted dosage error stands at less than 0.015 relative to the true optimal dosage across both simulation scenarios. This results in an outcome error of 1.15 for the sine wave and a mere 0.20 for the thresholded linear function. However, it is crucial to highlight that as we move away from the optimal dosage, SCIGAN's performance diminishes. The gap between the ground-truth response and the model's prediction widens, with the outcome error dramatically escalating to 10.42 for the sine wave and 1.77 for the thresholded linear function.

We hypothesize that the performance degradation is linked to the selection bias from the treatment assignment using the beta distribution. The sparsity of observational data intensifies as we shift away from the optimal dosage. The very nature of generative models, where the generator learns indirectly based on feedback from the discriminator, becomes particularly vulnerable under such data scarcity. In medical contexts, this translates to tangible risks: treatment guidelines often inadvertently produce selection biases, causing certain patients to consistently receive non-optimal dosages. Moreover, such misjudgments can lead to overtreatment or undertreatment, presenting serious risks for patient outcomes.

To address the performance degradation due to selection bias, a novel mechanism is necessary to comprehensively leverage data from underrepresented dosage intervals. One plausible approach involves integrating domain knowledge. For example, E2B method proposed by Bahadori create arbitrary response function sets using domain insights and adjust the sample weights of the data [2]. However, these methods might still introduce biases if domain knowledge is insufficient. Despite the critical importance of this issue in the medical field, research targeting the challenges of selection bias in continuous treatment effect estimation remains sparse. Before AI-driven treatment effect estimation models can be reliably implemented in clinical settings, there's an evident need for multifaceted research to address these challenges.

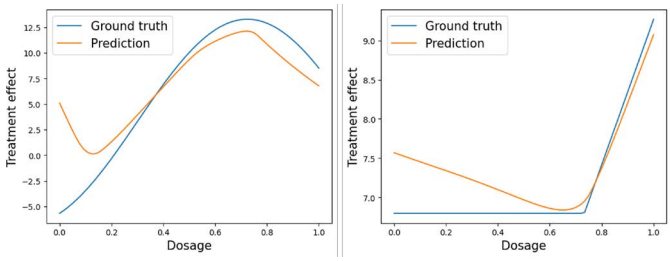

Fig. 1. Confusion matrix of labels between Experts 4 and 3 (left), and Experts 5 and 3 (right)

IV. Conclusion

We have investigated a state-of-the-art deep generative model for estimating continuous treatment effects, particularly when confronted with selection bias. As the medical community increasingly looks towards AI for decision-making support, it's imperative to understand and address these challenges. Our study not only underscores this need but also paves the way for future research aiming to improve the robustness and reliability of AI-driven causal effect estimation in healthcare.

Acknowledgment

This work was supported by Institute for Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT) (No. RS-2023-00261068, Development of Lightweight Multimodal Anti-Phishing Models and Split-Learning Techniques for Privacy-Preserving Anti-Phishing)

References

[1] Bica, Ioana, et al. "From real-world patient data to individualized treatment effects using machine learning: current and future methods to address underlying challenges." Clinical Pharmacology & Therapeutics 109.1 (2021): 87-100.

[2] Bahadori, Taha, Eric Tchetgen Tchetgen, and David Heckerman. "End-to-end balancing for causal continuous treatment-effect estimation." International Conference on Machine Learning. PMLR, 2022.

[3] Yao, Liuyi, et al. "A survey on causal inference." ACM Transactions on Knowledge Discovery from Data (TKDD) 15.5 (2021): 1-46.

[4] Shi, Claudia, David Blei, and Victor Veitch. "Adapting neural networks for the estimation of treatment effects." Advances in neural information processing systems 32 (2019).

[5] Schwab, Patrick, et al. "Learning counterfactual representations for estimating individual dose-response curves." Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 34. No. 04. 2020.

[6] Ma, Haixu and Zeng, Donglin and Liu, Yufeng. "Learning Individualized Treatment Rules with Many Treatments: A Supervised Clustering Approach Using Adaptive Fusion." Advances in Neural Information Processing Systems 35 (2022).

[7] Jesson, Andrew, et al. "Scalable sensitivity and uncertainty analyses for causal-effect estimates of continuous-valued interventions." Advances in Neural Information Processing Systems 35 (2022).

[8] Bica, Ioana, James Jordon, and Mihaela van der Schaar. "Estimating the effects of continuous-valued interventions using generative adversarial networks." Advances in Neural Information Processing Systems 33 (2020).