On the Effects of Irrelevant Variables in Treatment Effect Estimation with Deep Disentanglement

Ahmad Saeed Khan
Supervisors: Johannes Andreas Stork, Erik Schaffernicht
Örebro University

Abstract

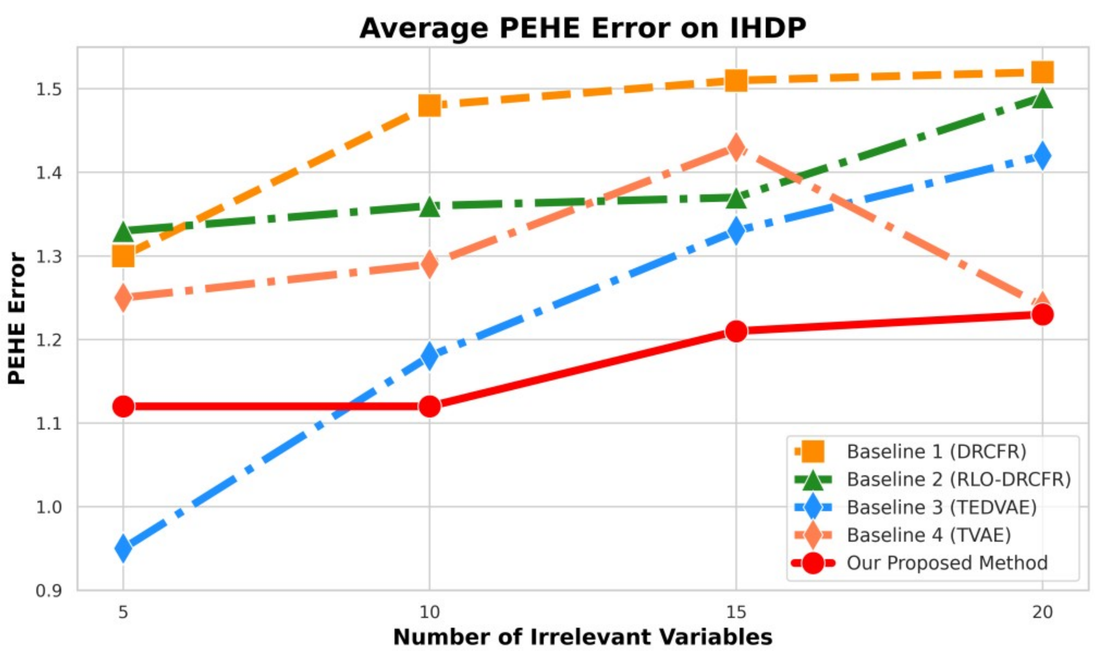
We address the challenge of irrelevant variables in treatment effect estimation from observational data. Our deep embedding method explicitly disentangles irrelevant variables alongside instrumental, confounding, and adjustment factors. Using an autoencoder and orthogonalization, we prevent irrelevant information leakage into latent spaces. Experiments on synthetic and real-world datasets show improved identification of irrelevant variables and more accurate treatment effect predictions, with robustness to increasing irrelevant dimensions compared to prior methods.

Motivation

Estimating treatment effects from observational data is critical yet challenging due to **selection bias** and **irrelevant variables**.

- **Current Limitation**: Existing disentanglement methods fail to explicitly address irrelevant variables, leading to:

  1. Poor **Precision in Estimation of Heterogeneous Effects (PEHE)**.
  2. Significant **information leakage** among latent factors.

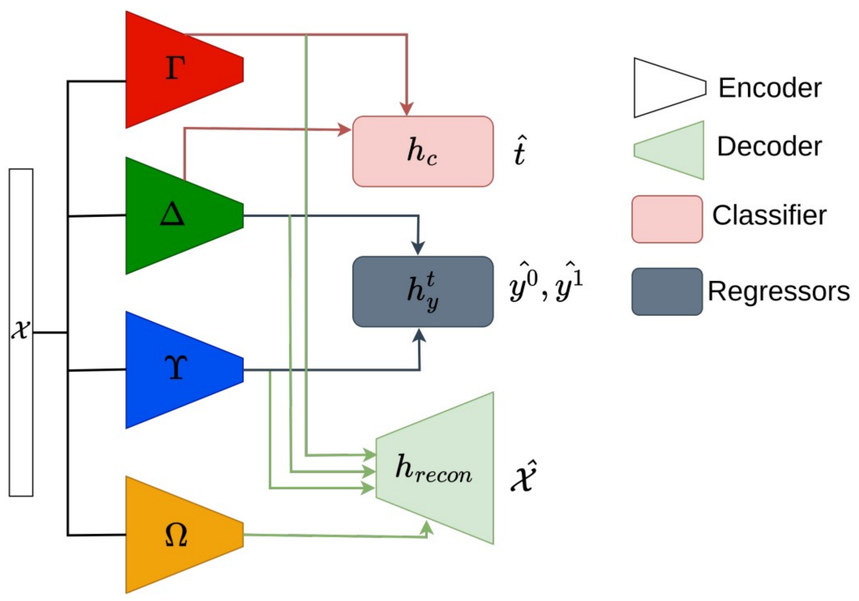


Contributions

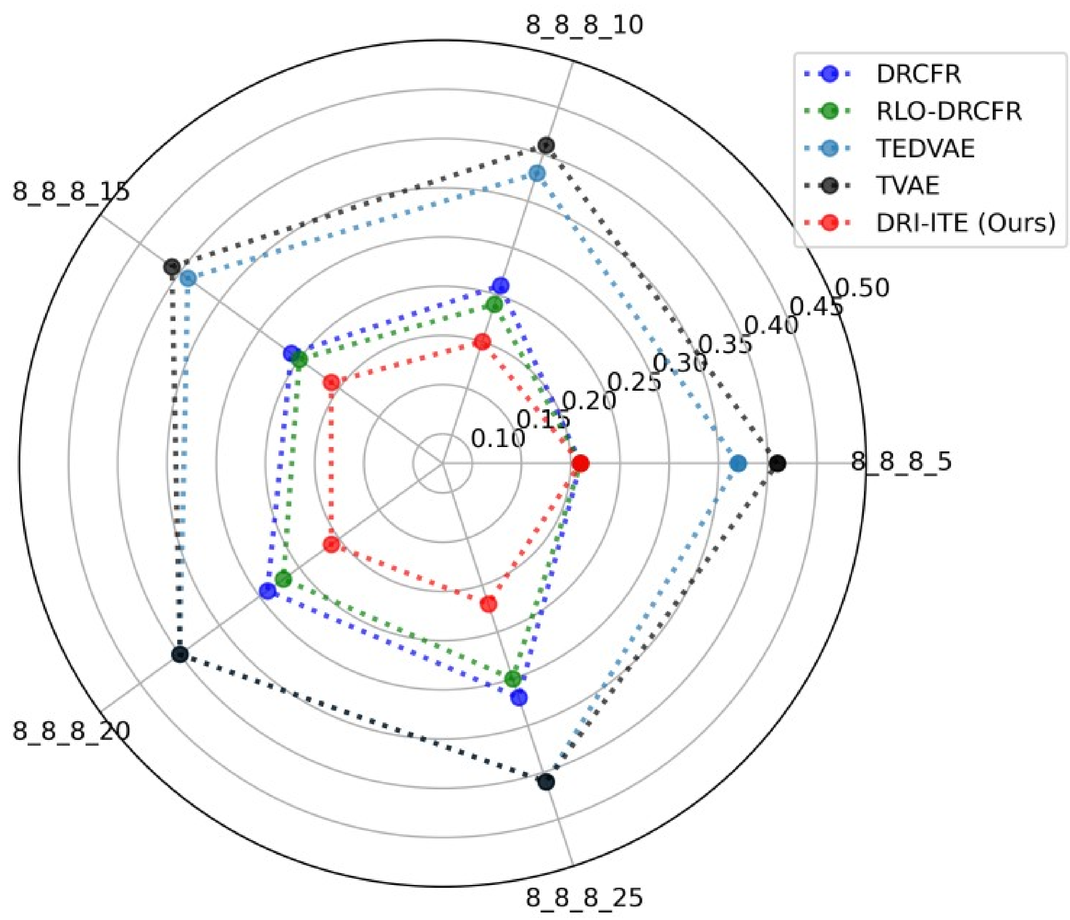
•**Explicitly identify and separate irrelevant variables (Ω)** from instrumental (Γ), confounding (Δ), and adjustment (Y) factors.

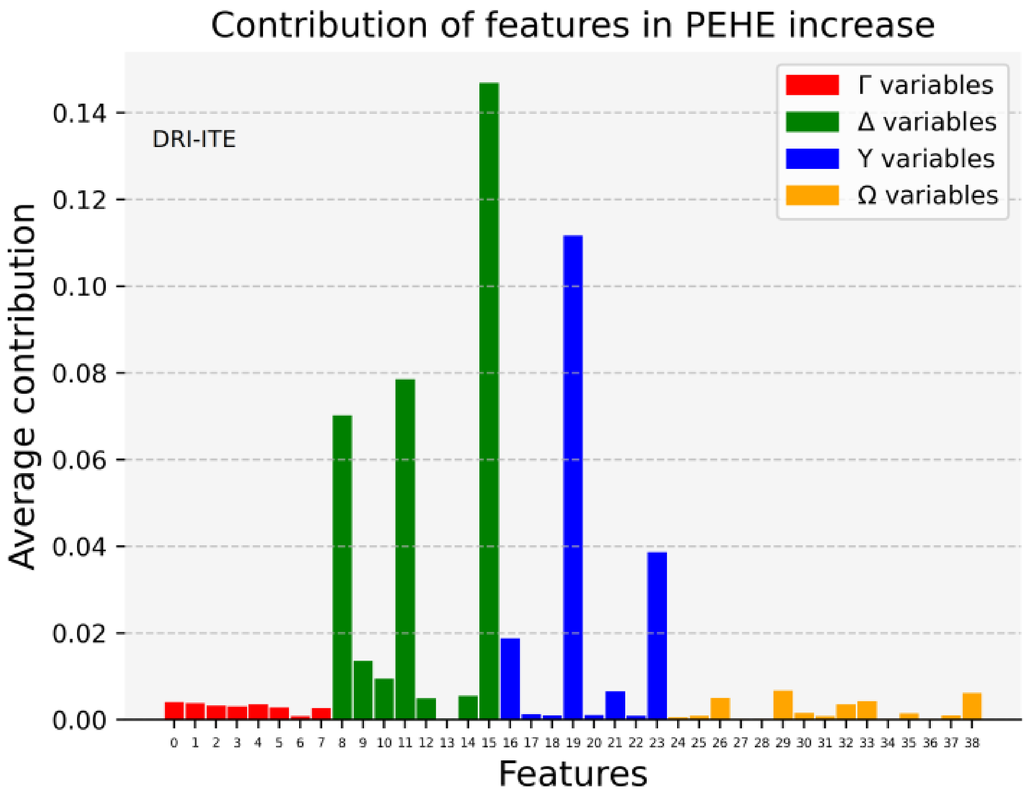
•Use a **reconstruction loss** and **orthogonality constraints** for robust disentanglement.

.



Results





Funded by

This work has been supported by the Industrial Graduate School Collaborative AI & Robotics funded by the Swedish Knowledge Foundation Dnr:20190128