Causal Inference Using Bayesian Deep Learning

by

Adi Lin

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

Australian Artificial Intelligence Institute (AAII), School of Computer Science
Faculty of Engineering and Information Technology (FEIT)
University of Technology Sydney

August, 2020

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This research is supported by the Australian Government Research Training Program.

Production Note:
Signature removed
prior to publication.

Signature of Candidate

To my family.

ACKNOWLEDGEMENTS

As I write these acknowledgements, it seems as though it was only yesterday when I arrived in Sydney. I would like to take this opportunity to express my deepest gratitude to the people who have inspired and helped me over the last few years.

First and foremost, I would like to express my sincere gratitude to my supervisors Prof. Jie Lu, Prof. Guangquan Zhang and Dr Junyu Xuan for their continuous encouragement and guidance throughout my PhD candidature. I really appreciate the opportunity to undertake research in the Australian Artificial Intelligence Institute (AAII, formerly the Centre for Artificial Intelligence (CAI)), in the Faculty of Engineering and Information Technology (FEIT) at the University of Technology Sydney.

My principal supervisor, Prof. Lu, always allowed me sufficient freedom to explore fundamental knowledge and the unknown, enabling me to investigate the world of causality. I will always be grateful for her patience. Prof. Lu provided many research suggestions from a wide perspective, and she also taught me how to present a professional paper and gave me constructive feedback and suggestions. I am also grateful for having had Prof. Zhang as my co-supervisor. Prof. Zhang shared his experience in research and life and taught me how to write critical reviews for technical articles. Dr Junyu Xuan engaged in profitable discussions with me and gave me detailed assistance in relation to research directions. I learned a great deal about research methods from Dr Xuan, and I have greatly benefited from his scientific writing skills.

I am also indebted to the other members of my committee, Dr Haiyan Lu and Dr Ling Chen, who kindly contributed their time to help me become a better researcher. I really appreciated the discussions on causal inference with Dr Fujin Zhu. I would like to express my thanks for the support from the administrative management team, Camila Cremonese, Robyn Barden, Lily Qian and Janet Stack. I would like to thank all the other students and staff from the AAII and FEIT who have helped and influenced me. I would like to particularly mention the talented researchers of the Decision Systems & e-Service Intelligence Lab, Daokun Zhang, Mingming Gong, Yuangang Pan, Pingbo Pan, and others who are not listed here.

I kindly thank Ms Jemima Moore and Ms Robyn Barden for polishing the language used in my publications and thesis. I have learnt much about academic writing from them.

I am grateful to the Australian Artificial Intelligence Institute (AAII) at the University of Technology Sydney. This study was supported by the Australian Research Council's Discovery project.

Finally and most importantly, I would like to express my heartfelt appreciation and gratitude to my parents and my sisters. During this long journey, they always believed in and supported me. I dedicate this thesis to them. I especially thank my sisters, who took care of my parents during my study and research.

Adi Lin

Sydney, Australia, 2020

ABSTRACT

Causal inference from observational data has wide application in precision medicine, economics, social sciences, computational advertising, and so on. Causal inference from observational data aims to estimate causal effects when controlled experimentation is not feasible. Causal inference is the process of identifying how a change in a cause leads to a change in the outcome. In today's data–driven world, causal inference has become a key part of the evaluation process for many purposes, such as examining the effects of medicine or the impact of an economic policy on society.

Confounding bias occurring in observational data may result in causal inference leading a wrong result. Confounding bias is the fundamental bias of causal inference from observational data. Under some specific assumptions, it is possible to estimate the causal effect from observational data with confounding bias. Although the existing literature contains some excellent models, there is room to improve their representation power and their ability to capture complex causal relationships. Furthermore, there is a research gap between deep Bayesian models and causal inference from observational data under confounding bias. In order to narrow the gap, this thesis provides algorithms to estimate the causal effects from observational data in some cases when a set of confounders exists. This result can provide effective decision support for policymakers in various areas.

This thesis recovers causal inference from observational data with observed confounding bias, unobserved confounding bias and time-dependent confounding bias. First, this thesis considers two kinds of causal inference problems when observed confounding bias exists. This thesis proposes a model with separate Gaussian processes to estimate the **C**onditional **A**verage **C**ausal Effect on the **T**reated (CACT). Each separate Gaussian process is proposed to estimate the average causal effect for the treated group and the control group. In order to estimate various kinds of causal effects, such as average, conditional average, and average treated, this thesis focuses on Bayesian generative models. A prior called Causal DP is proposed, and a generative model called CDP based on the prior is developed to estimate causal effects. The prior captures the complex relationships between covariates, treatments, and outcomes in observational data. The model is a Bayesian nonparametric generative model and is not based on the assumption of any parametric distribution. The proposed generative model performs well with missing covariates and does not suffer from overfitting. Second, this thesis proposes methods to resolve the challenges when unobserved confounding bias exists. The instrumental variable methods resolve this problem by introducing a variable that is correlated with the treatment and affects the outcome only through the treatment. This thesis presents a one-stage approach to jointly estimate the treatment distribution and the outcome generating function through a designed deep neural network structure. The one-stage method is different to existing instrumental variable methods requiring two stages to separately estimate the conditional treatment distribution and the outcome generating function. This study is the first to merge the two stages to leverage the outcome to the treatment distribution estimation. Finally, this thesis estimates the causal effect for **D**ynamic **T**reatment **R**egimes (DTRs) where time-dependent confounding bias exists. Censoring and time-dependent confounding under DTRs bring a challenge in the observational data has a declining sample size but an increasing feature dimension over time. This thesis combines outcome regression models with treatment models for high-dimensional features using uncensored subjects that are potentially small-sample. And this thesis fits deep Bayesian models for outcome regression models to unveil complex relationships between confounders,

treatments, and outcomes.

This thesis evaluates all the methods proposed in this thesis using synthetic, semi-synthetic or real-world data. Comparative experiments against several state-of-the-art methods show that the proposed methods generally perform better than or are comparative with their competitors. Given the key importance of causal inference in both theory and real-world applications, we argue that the models and algorithms proposed in this thesis contribute to both scientific research and practical applications.

Dissertation directed by

Dist. Prof. Jie Lu, Associated Prof. Guangquan Zhang, and Dr. Junyu Xuan

Australian Artificial Intelligence Institute, School of Computer Science, FEIT

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES