SPRINGER BRIEFS IN
ELECTRICAL AND COMPUTER ENGINEERING

Jiuyong Li
Lin Liu
Thuc Duy Le

Practical Approaches to Causal Relationship Exploration

Springer

SpringerBriefs in Electrical and Computer Engineering

For further volumes:
http://www.springer.com/series/10059

Jiuyong Li • Lin Liu • Thuc Duy Le

Practical Approaches to Causal Relationship Exploration

Springer

Jiuyong Li
School of Information Technology
and Mathematical Sciences
University of South Australia
Adelaide, South Australia
Australia

Thuc Duy Le
School of Information Technology
and Mathematical Sciences
University of South Australia
Adelaide, South Australia
Australia

Lin Liu
School of Information Technology
and Mathematical Sciences
University of South Australia
Adelaide, South Australia
Australia

ISSN 2191-8112 ISSN 2191-8120 (electronic)
SpringerBriefs in Electrical and Computer Engineering
ISBN 978-3-319-14432-0 ISBN 978-3-319-14433-7 (eBook)
DOI 10.1007/978-3-319-14433-7

Library of Congress Control Number: 2015930835

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Recommended by Xuemin (Sherman) Shen.

Preface

Causal discovery aims to discover the cause-effect relationships between variables. The relationships provide explanations as to how events have happened and predictions as to which events will happen in the future. Causality has been studied and utilised in almost all disciplines, e.g. medicine, epidemiology, biology, economics, physics, social science, as a basis for explanation, prediction and decision making.

Randomised controlled trials are the gold standard for discovering causal relationships. However, in many cases it is impossible to conduct randomised controlled trials due to cost, feasibility and/or ethical concerns.

With the rapid explosion of data collected in various areas, it is desirable to discover causal relationships in observational data. Causal discovery in data does not only reduce the costs for many scientific explorations and assist decision making, but importantly, it also helps detect crucial signals in data which might not be identified by domain experts to prevent serious consequences. Furthermore, data provides great opportunities for automated causal discovery and exploration by exploiting existing and developing new computational methods.

Significant achievements in causal modelling and inference have been made in various disciplines. Observational studies have long been used in medical research for identifying causal factors of diseases. Causal models, such as structural equation model and potential outcome model, have been used in a range of areas. In computer science, causal discovery based on graphical models has made significant theoretical achievements in the last 30 years.

However, there is still a lack of practical methods for causal discovery in large data sets. Most existing approaches are either hypothesis driven or unable to deal with large data sets. The causal discovery in this book refers to automated exploration of causal factors in large data sets without domain experts' hypotheses (domain experts may not know what to expect). More efficient methods are needed to deal with different types of data and applications for this purpose.

We aim to introduce four practical causal discovery methods for practitioners to mine their increasing data collection for causal information. We explain the mechanisms of the methods and provide demonstrations for the use of the methods. Relevant software tools can be found at their authors' home pages. Note that causal conclusions are not guaranteed by using a method since a data set may not satisfy the assumptions of the method, which itself includes heuristics. However, the methods in this book have a major advantage over other data mining or machine learning approaches for relationship exploration. These methods detect the relationship between two variables by considering other variables, and this consideration makes the discovered relationships less likely to be spurious or volatile.

We also aim to share our understandings in causal discovery with our peer researchers. Causal discovery is the goal for scientific exploration, and causal discovery in data is what computational researchers are able to contribute greatly to our society. Although it is arguable whether causality is definable or discoverable in data, the research endeavor in various disciplines has made significant progress in the theory of causal modelling and inference, which has shown promising applications in classification and prediction. It is time for data mining and machine learning researchers to reap the theoretical results and design efficient and practical algorithms for large scale causal relationship exploration. In this book, we characterise the causal discovery as a search problem—searching for persistent associations. This characterisation hopefully paves a short path for data mining and machine learning researchers to design more efficient algorithms for causal discovery with big data.

This book is also useful for students who want to improve their knowledge in this emerging area of research and applications. Causal discovery is a truly multiple disciplinary topic. We do not assume the knowledge of readers in a specific area. We try to explain the concepts and techniques clearly and use as many examples as possible. We hope that all readers, regardless of their knowledge backgrounds, can get some benefits by reading this book.

We thank Springer editors, Melissa Fearon and Jennifer Malat, for appreciating our idea and supporting us in writing the book. We also thank the Australian Research Council for its support for our work in this important area. This book would not be possible without the support of our collaborators, and here we just name a few: Jixue Liu, Zhou Jin, Bing-yu Sun and Rujing Wang. Algorithm implementation and experiments have been helped by Xin Zhu, Saisai Ma and Shu Hu. Certainly, the unfailing support of our families is crucial for us, and our gratitude is beyond words.

Adelaide, Australia,
October 2014

Jiuyong Li
Lin Liu
Thuc Duy Le

Contents