Tutorial: Causal Inference in Biomedical Data Analytics

Basics and Recent Advances

May D. Wang

Depts. of BME, ECE, CSE

Georgia Institute of Technology and Emory University

Atlanta, GA, US

maywang@gatech.edu

Hang Wu

Department of Biomedical Engineering

Georgia Institute of Technology

Atlanta, GA, US

hangwu@gatech.edu

ABSTRACT

Causal questions are being answered every day in the biomedical domain, and have significant impact on biomedical experimentation design, data analysis, and healthcare decision making. In this tutorial, we focus on identifying the (quantitative) causal effect of interventions. We will introduce a) introduction of causal inference, and popular frameworks for formulating causal inference; b) basics of causal effect identification algorithms; c) state-of-the-art methods by incorporating deep learning and machine learning; d) applications in biomedical data analytics, as well as challenges and opportunities moving forward.

CCS CONCEPTS

- Computing methodologies~Causal reasoning and diagnostics

KEYWORDS

Causal Inference, Counterfactual Reasoning

ACM Reference format:

May D. Wang and Hang Wu. 2019. ACM-BCB' 19 Tutorial: Causal Inference in Biomedical Data Analytics - Basics and Recent Advances In *Proceedings of ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB'19), September 7–10, 2019, Niagara Falls, NY, USA*, 1 pages. <https://doi.org/10.1145/3307339.3343177>

1. Description

Decision-making in biomedical and healthcare domain often involves asking causal and counterfactual questions, such as: *does medication A lead to reduced blood pressure? Had we discharged the patient later, would she still be readmitted? Does the release of chemical compound B from factories cause the increase of*

cancer cases in the region? Such questions are being answered in practice every day and are impacting people's well-being, the course of biomedical experimentation design, and public health policy making. It's thus important for biomedical research to help answer these causal questions, by designing data-driven causal inference algorithms leveraging the growing amount of biomedical data and accumulated domain causal knowledge.

Causal inference algorithms studies 1) whether there exists causal relationship between variables of interest (e.g. genotypes and phenotypes) and 2) the quantitative causal effect of the interventions (e.g. taking medications and performing procedures). In this tutorial, we will focus on the latter part: identifying the (quantitative) causal effect of interventions. This topic is one of the key problems in health informatics such as epidemiology, and also have connections and applications in bioinformatics research.

2. Outline

The key focus of the tutorial is algorithms for identifying causal effect when RCTs are not available. i.e. when we cannot manipulate any variables and have to use observational data (e.g. electronic health records). We start with conventional causal inference algorithms and present the algorithms in two groups based on whether confounding variables are assumed to exist or not. We will then introduce recent advances in incorporating machine learning and deep learning algorithms for causal inference. In the last section of the tutorial, we show representative applications of causal inference in biomedical data analytics and discuss best practices in applying causal inference, as well as opportunities and challenging moving forward.

Acknowledgement

This tutorial was partly supported by research grants from CHOA, CDC (HHSD2002015F62550B), NSF (1651360), NSF (1844455), AHA, Carol Ann and David Flanagan Faculty Fellow Award to Prof. May D. Wang.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

ACM-BCB'19, September 7–10, 2019, Niagara Falls, NY, USA

© 2019 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-6666-3/19/09.

<https://doi.org/10.1145/3307339.3343177>