Computers and Concrete, *Vol. 31, No. 4 (2023) 277-292*
https://doi.org/10.12989/cac.2023.31.4.277

Causality, causal discovery, causal inference and counterfactuals in Civil Engineering: Causal machine learning and case studies for knowledge discovery

M.Z. Naser*[1,2] and Arash Teymori Gharah Tapeh[1]

[1]School of Civil & Environmental Engineering and Earth Sciences (SCEEES), Clemson University, USA
[2]Artificial Intelligence Research Institute for Science and Engineering (AIRISE), Clemson University, USA

(*Received November 9, 2022, Revised December 19, 2022, Accepted December 19, 2022*)

Abstract. Much of our experiments are designed to uncover the cause(s) and effect(s) behind a phenomenon (i.e., data generating mechanism) we happen to be interested in. Uncovering such relationships allows us to identify the true workings of a phenomenon and, most importantly, to realize and articulate a model to explore the phenomenon on hand and/or allow us to predict it accurately. Fundamentally, such models are likely to be derived via a causal approach (as opposed to an observational or empirical mean). In this approach, causal discovery is required to create a causal model, which can then be applied to infer the influence of interventions, and answer any hypothetical questions (i.e., in the form of What ifs? Etc.) that commonly used prediction- and statistical-based models may not be able to address. From this lens, this paper builds a case for causal discovery and causal inference and contrasts that against common machine learning approaches - all from a civil and structural engineering perspective. More specifically, this paper outlines the key principles of causality and the most commonly used algorithms and packages for causal discovery and causal inference. Finally, this paper also presents a series of examples and case studies of how causal concepts can be adopted for our domain.

Keywords: causal discovery; causal inference; civil engineering; machine learning

1. Introduction

Seeking causal knowledge is a foundational pursuit with branching philosophical, epistemological, and ontological ties [1]. Causal knowledge accurately describes how a phenomenon, *Y*, comes to be by answering key questions such as, what causes *Y*? How, when, and why does *Y* occur? Etc. (Bunge 1979). Arriving at precious answers to the above questions can be ambitious, and thus, causal knowledge can be difficult to pinpoint or measure (Schölkopf 2019).

In the civil and environmental engineering domain, we often design experiments in which a set of parameters is identified or assumed to be responsible for a given phenomenon. For example, say we are trying to understand the concept of flexural capacity in beams. First, we would fabricate a Beam A with certain geometrical and material features (i.e., W16×36, Grade A992). Then, we add boundary conditions (i.e., simply supported) to load this beam in a manner that enables us to capture its flexural response. We load the beam[2] and report that this beam fails once the level of the applied moment reaches 361.6 kN.m.

At this point, a link is then obtained by associating the reported moment at failure to the geometrical and material features as well as the loading configuration of Beam A. This link draws from the following observation: Applying a bending moment of 361.6 kN.m to Beam A has caused it to fail. A series of questions may arise: 1) What has caused the failure of Beam A? And 2) Was failure triggered due to the geometric configurations of W16×36? Or, perhaps due to the material properties of Grade A992? What about the effect of boundary conditions? Or Loading configuration?

The above are examples of causal questions. Another set of questions that may also arise includes: Other things constant [*Ceteris Paribus*] 3) Would Beam A have failed if it was not for the presence of the bending moment? 4) Would this beam fail at 361.6 kN.m if it had been a W18×40? Or had it been made from Grade A36? These are counterfactual questions that also belong to the causal family. Answering the above two sets of questions requires a causal investigation.

Fortunately, our domain knowledge can aid, if not substitute, the need for a thorough causal investigation. For example, we know that in the case of W-shaped steel beams under bending, the geometric features are lumped into the plastic modulus, *Z*, and that the material features are represented by the yield strength of structural steel, f_y. Both are also tied via Eq. (1), which represents a multiplication form to estimate the moment capacity (i.e., resistance) of a given W-shaped steel beam.

$$Resistance = Z \times f_y \tag{1}$$

*Corresponding author, Assistant Professor
E-mail: mznaser@clemson.edu
Website: www.mznaser.edu

[1] A cohesive look at causality from the lens of philosophy, epistemology, and ontology, can be found in (Michotte 2017, Salmon 2003).
[2] Say with one point load at mid-span.

Copyright © 2023 Techno-Press, Ltd.
http://www.techno-press.org/?journal=cac&subpage=8
ISSN: 1598-8198 (Print), 1598-818X (Online)

https://doi.org/10.5121/ijdkp.2015.5201.

Martí-Vargas, J.R., Ferri, F.J. and Yepes, V. (2013), "Prediction of the transfer length of prestressing strands with neural networks", *Comput. Concrete*, **12**(2), 187-209. https://doi.org/10.12989/cac.2013.12.2.187.

Michotte, A. (2017), *The Perception of Causality*, Routledge, New York, NY, USA.

Mitchell, T. (1997), *Machine Learning*, McGraw Hill, New York, NY, USA.

Muggleton, S. (1991), "Inductive logic programming", *New Gen. Comput.*, **8**, 295-318. https://doi.org/10.1007/BF03037089.

Naser, M.Z. and Alavi, A.H. (2021), "Error metrics and performance fitness indicators for artificial intelligence and machine learning in engineering and sciences", *Arch. Struct. Constr.*, **2021**, 1-19. https://doi.org/10.1007/s44150-021-00015-8.

Naser, M.Z. and Kodur, V.K. (2022), "Explainable machine learning using real, synthetic and augmented fire tests to predict fire resistance and spalling of RC columns", *Eng. Struct.*, **253**, 113824. https://doi.org/10.1016/j.engstruct.2021.113824.

Nogueira, A.R., Gama, J. and Ferreira, C.A. (2021), "Causal discovery in machine learning: Theories and applications", *J. Dyn. Games*, **8**(3), 203. https://doi.org/10.3934/jdg.2021008.

Nogueira, A.R., Pugnana, A., Ruggieri, S., Pedreschi, D. and Gama, J. (2022), "Methods and tools for causal discovery and causal inference", *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.*, **12**(2), e1449. https://doi.org/10.1002/widm.1449.

pcalg (2022), Methods for Graphical Models and Causal Inference [R package pcalg version 2.7-6], Comprehensive R Archive Network (CRAN).

Pearl, J. (2009a), "Causal inference in statistics: An overview", *Stat. Surv.*, **3**, 96-146. https://doi.org/10.1214/09-SS057.

Pearl, J. (2009b), *Causality*, Cambridge University Press, Cambridge, UK.

Pearl, J. (2013), "Causal diagrams and the identification of causal effects", *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge, UK.

Pearl, J. and Mackenzie, D. (2018a), *The Book of Why: The New Science of Cause and Effect-Basic Books*, Basic Books, New York, NY, USA.

Pearl, J. and Mackenzie, D. (2018b), *The Book of Why: The New Science of Cause and Effect, Notices of the American Mathematical Society*, Basic Books, New York, NY, USA.

Ramsey, J., Glymour, M., Sanchez-Romero, R. and Glymour, C. (2017), "A million variables and more: The fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images", *Int. J. Data Sci. Anal.*, **3**, 121-129. https://doi.org/10.1007/s41060-016-0032-z.

Rubin, D.B. (2005), "Causal inference using potential outcomes", *J. Am. Stat. Assoc.*, **100**(469), 322-331. https://doi.org/10.1198/016214504000001880.

Salmon, W.C. (2003), *Causality and Explanation*, Oxford University Press, Oxford, UK.

Sanjayan, G. and Stocks, L.J. (1993), "Spalling of high-strength silica fume concrete in fire", *ACI Mater. J.*, **90**(2), 170-173. https://doi.org/10.14359/4015.

Scheines, R. (1996), *An Introduction to Causal Inference*, Carnegie Mellon University, Pittsburgh, PA, USA.

Schölkopf, B. (2019), "Causality for machine learning", *arXiv preprint*, **1911**, 10500.

Sharma, A. and Kiciman, E. (2019), "DoWhy: A Python package for causal inference", https://github. com/microsoft/dowhy.

Spirtes, P., Glymour, C. and Scheines, R. (2000), "Causation, prediction, and search (Springer lecture notes in statistics)", *Lecture Notes in Statistics*, MIT Press, Cambridge, MA, USA.

Spirtes, P. and Zhang, K. (2016), "Causal discovery and inference: concepts and recent methodological advances", *Appl. Informat.*, **3**(1), 1-28. https://doi.org/10.1186/s40535-016-0018-x.

Surveys|NCSES|NSF (2022), https://www.nsf.gov/statistics/surveys.cfm.

Thelwall, M. (2018), "Dimensions: A competitor to Scopus and the Web of Science?", *J. Informetr.*, **12**(2), 430-435. https://doi.org/10.1016/j.joi.2018.03.006.

TIGRAMITE (2022), GitHub - Jakobrunge/Tigramite: Tigramite is a Python Package for Causal Inference with a Focus on Time Series Data, https://github.com/jakobrunge/tigramite.

Tong, T. and Yu, T.E. (2018), "Transportation and economic growth in China: A heterogeneous panel cointegration and causality analysis", *J. Transp. Geogr.*, **73**, 120-130. https://doi.org/10.1016/j.jtrangeo.2018.10.016.

Triantafillou, S. and Tsamardinos, I. (2016), "Score based vs constraint based causal learning in the presence of confounders", *CEUR Workshop Proceedings*.

Uber Technologies (2020), About Causal ML - Causalml Documentation, Uber Technologies Inc., San Francisco, USA. https://causalml.readthedocs.io/en/latest/about.html.

Vowels, M.J., Camgoz, N.C. and Bowden, R. (2021), "D'ya like DAGs? A survey on structure learning and causal discovery", *ACM Comput. Surv.*, **55**(4), 1-36. https://doi.org/10.1145/3527154.

Wagner, C.H. (1982), "Simpson's paradox in real life", *American Statistician*, **36**(1), 46-48.

Wardhana, K. and Hadipriono, F.C. (2003a), "Analysis of recent bridge failures in the United States", *J. Perfor. Constr. Facil.*, **17**(3), 144-150. https://doi.org/10.1061/(ASCE)0887-3828(2003)17:3(144).

Wasserman, L. (2021), "Causal inference", Statistics & Data Science, Carnegie Mellon University, Pittsburgh, PA, USA.

Yaswanth, K.K., Revathy, J. and Gajalakshmi, P. (2021), "Artificial intelligence for the compressive strength prediction of novel ductile geopolymer composites", *Comput. Concrete*, **28**(1), 55-68. https://doi.org/10.12989/cac.2021.28.1.055.

Yu, K., Li, J. and Liu, L. (2016), "A review on algorithms for constraint-based causal discovery", *arXiv preprint*, **1611**, 03977.