Strategy Summit

How do we link theory and practice in strategy?

A proposal and a set of papers to share with the panelists led by Martin Reeves

Moving toward evidence-based interventions: Contributions by the Community of Causal Data Science (CDS)[1]

Paul Hünermund and Gwen Lee[2]

June 3, 2024

ABSTRACT

We introduce to Strategy Summit the community of Causal Data Science (CDS) where research meets practice. Our aim is to bridge the gap between theory and practice by advancing real-world applications of causal data science that connect causal machine learning and business decision making. Moreover, we provide an overview of the tools and methods that help research scientists engage in discussions with industry professionals and policy-makers about the role and impact of causality in machine learning.

LINKING THEORY AND PRACTICE IN STRATEGY BY CONNECTING CAUSAL MACHINE LEARNING AND BUSINESS DECISION MAKING

Establishing causality beyond association, influence, or correlation is key to informed managerial decisions and interventions (Lee & Bettis, 2023). By contrast, standard machine learning approaches remain purely correlational and prediction-based, confining them to analytical insights that can only partly address a wide variety of managerial decision problems (Hünermund et al., 2022). To inform evidence-based interventions, causal knowledge is critical for strategic and organizational decision-making. As argued by Felin and Zenger (2017:258) in proposing the theory-based view of strategy, novel or "great" strategies come from theories. Building on the theory-based view, we submit that to bridge the gap between theory and practice, the field of strategy can move faster toward evidence-based interventions by connecting causal machine learning and business decision-making. Data-augmented decision-making provides analytical insights and requires theory that is fundamental in causal inference.

---

[1] CDS website: https://www.causalscience.org/

[2] phu.si@cbs.dk https://orcid.org/0000-0001-9163-038X (PH) and gwendolyn.lee@warrington.ufl.edu https://orcid.org/0000-0003-4276-7982 (GL)

CAUSAL DATA SCIENCE

Causal data science (CDS) distinguishes between three levels of inferential tasks: (1) associational, (2) interventional, (3) counterfactual (the "Ladder of Causation" or Pearl Causal Hierarchy; Bareinboim et al., 2022). Associational tasks predict a certain outcome and are thus useful for deciding when to apply a certain policy that is known to be effective (e.g., when to send promotions to customers who are at risk of churning?). In comparison, interventional and counterfactual tasks ask questions that distinguish "correlation" from "causation". Ascending the ladder of causation, interventional tasks ask questions to examine what types of policies are most effective (e.g., which CEO personality profiles will likely affect firm performance most favorably?). At the top of the ladder, counterfactual tasks use forward-looking or retrospective reasoning at the individual level and are thus able to answer "why?" questions (e.g., why did our operational efficiency improve? Was it due to the merger last year?). The relevance of CDS for strategy follows both from practical (what works? [interventional]) as well as epistemological (why does it work? [counterfactual]) considerations. These "causation" questions always require theoretical background knowledge derived from expert domain knowledge and can never be answered in a fully data-driven way. As such, CDS is highly compatible with the aforementioned theory-based view of strategy and entrepreneurship.

CDS, WHERE RESEARCH MEETS PRACTICE

Interest in CDS among industry professionals is growing, showing an ongoing shift among practitioners toward applying causal data science methods for business decision-making (Hünermund et al., 2022). The annual CDS Meeting, which started in 2020 and has attracted more than 3,500 participants including academics and practitioners, bridges industry and academia in causal data science (https://www.causalscience.org). CDS is increasingly featured in many special sessions at key academic conferences such as NeurIPS, KDD, and CLeaR. The Gartner Hype Cycle 2023 lists Causal AI under its "Innovation Trigger" category: https://www.gartner.com/en/articles/what-s-new-in-the-2023-gartner-hype-cycle-for-emerging-technologies

TOOLS & METHODS

The analytical tools and computational methods of CDS cover a complete and fully automatable causal inference framework for tackling a wide variety of data science tasks in applied research (see a recent survey by Hünermund & Bareinboim, 2023 on Causal Data Fusion). We highlight a few of these tools and methods. Directed Acyclic Graphs (DCGs) are used for model and variable selection (Hünermund et al., 2022). Causal forests are used to estimate heterogeneous treatment effects, moving beyond average treatment effects and, therefore, highlighting distinctiveness (Wager & Athey, 2017). Transportability analysis is used to guide research designs and economize on experimental costs in practice (Lee, 2024). Causal reinforcement learning is used to determine where to invest scarce resources for intervention (Lee & Bareinboim, 2020). Causal discovery is used to analyze statistical properties of purely observational data for machine learning-enhanced exploratory research that is supported by algorithms (Glymour et al., 2019). The DoWhy Library in Python offers a comprehensive workflow and causal AI pipeline for practitioners to follow and get started (https://www.datacamp.com/tutorial/intro-to-causal-ai-using-the-dowhy-library-in-python; also see https://crl.causalai.net for further references on causal AI).

REFERENCES

Bareinboim, Elias, Juan D. Correa, Duligur Ibeling, and Thomas Icard. (2022) On Pearl's hierarchy and the foundations of causal inference. In ***Probabilistic and Causal Inference: The Works Of Judea Pearl***, chapter 1 pp. 507–556. https://causalai.net/r60.pdf

Felin, Teppo, and Todd R Zenger. (2017). The theory-based view: economic actors as theorists. ***Strategy Science***, 2(4), 258–271. https://doi.org/10.1287/stsc.2017.0048

Glymour, Clark, Kun Zhang, and Peter Spirtes. (2019) Review of causal discovery methods based on graphical models. ***Frontiers in Genetics, Sec. Statistical Genetics and Methodology***, 10(524), 1–15. https://doi.org/10.3389/fgene.2019.00524

Hünermund, Paul, Beyers Louw and Mikko Rönkkö (2022) The choice of control variables: How causal graphs can inform the decision. In ***Academy of Management Proceedings*** (No. 1, p. 15534). Briarcliff Manor, NY 10510: Academy of Management. https://doi.org/10.5465/AMBPP.2022.294

Hünermund, Paul, and Elias Bareinboim (2023) Causal inference and data fusion in econometrics ***The Econometrics Journal*** https://doi.org/10.1093/ectj/utad008

Hünermund, Paul, Jermain Kaminski, and Carla Schmitt. (2022) Causal machine learning and business decision making. http://dx.doi.org/10.2139/ssrn.3867326

Lee, Gwendolyn K. (2024) How transportability analysis can be useful for cumulative theory testing in management research. ***Journal of Management Scientific Reports***, 2(2), 179–197. https://doi.org/10.1177/27550311241249137

Lee, Gwendolyn K., and Rich Bettis. (2023) Structural causal modeling of managerial interventions: What if managers had not intervened by doing this? ***Strategy Science***, 8(1), 24–43. https://doi.org/10.1287/stsc.2022.0169

Lee, Sanghack, and Elias Bareinboim. (2020) Characterizing optimal mixed policies: Where to intervene and what to observe. ***Advances In Neural Information Processing Systems***, 33, 8565–8576. https://proceedings.neurips.cc/paper_files/paper/2020/file/61a10e6abb1149ad9d08f303267f9bc4-Paper.pdf

Wager, Stefan, and Susan Athey. (2017) Estimation and inference of heterogeneous treatment effects using random forests. arXiv:1510.04342v4