Causal Learning: Improving Generalization, Representation Learning, and Efficiency using Tools from Causality

Logan Graham
DPhil Candidate
Machine Learning Research Group
University of Oxford

Bayesian Statistics in the Big Data Era, CIRM, November 2018

"Happy is the human who has been able to learn the causes of things."

- Virgil

"Happy is the human who has been able to ~~learn~~ use the causes of things."

\- Virgil

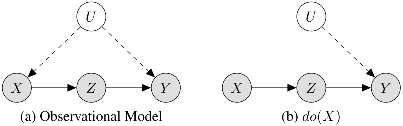
Core Lessons

1. Causal tools help beyond causal inference (structure learning, TE Estimation, IV/PSM/RDD/DiD/etc.)
2. Exploiting causal information can improve learning transfer, learning efficiency, and representation learning
3. There are opportunities to integrate into more general, traditional machine learning methods and tasks

Humans are causal

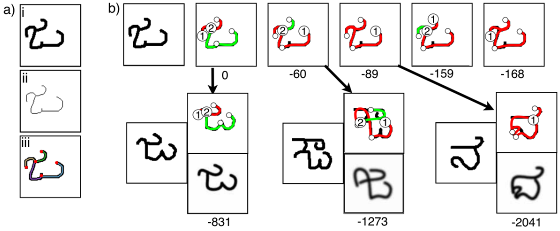
- Humans use interventions to learn unconfounded generative models to improve performance (Gopnik and Sobel, 2000, McCormack et al., 2016)
- Causal-relational structure is an inductive bias for how the world works (Lake et al., 2016, Battaglia et al., 2018)
- $\rightarrow$ Let's build this intuition into ML

Causality is Bayesian



- Modern frameworks stem from Bayesian beginnings (Pearl, 2000)
- Naturally expressed in probabilistic statements with causal DAGs
- Uncertainty over $\mathcal{M}$ ($X \to Y$ or $Y \to X$?)
- Uncertainty over $\mathcal{M}^{do(I)}$ ("counterfactual" or "interventional distribution")

Learning Transfer: Causal Relations



- Decode the generating process as (a) primitives (b) causal process (Lake et al., 2013) (also see (Kansky et al., 2017))

$$P\left(\psi, \theta^{(m)} | I^{(m)}\right) \approx \sum_{i=1}^{K} w_i \delta\left(\theta^{(m)} - \theta^{(m)[i]}\right) \delta\left(\psi - \psi^{[i]}\right)$$

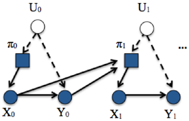
One-shot classification

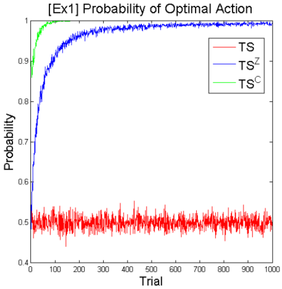
Learner	Error rate
Humans	4.5%
HBPL	4.8%
Affine	18.2 (31.8%)
HD	34.8 (68.3%)
DBM	38 (72%)
SS	62.5%
NN	78.3%

Character generation

Efficiency: Confounding-aware Actions

Bandits w/ confounded reward / inclination process

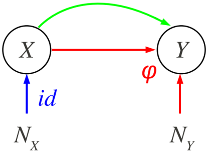




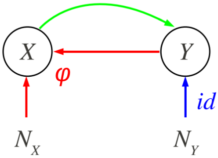
Causal Thompson Sampling: account for counterfactual reward under confounding (Bareinboim et al., 2015) (also see (Lattimore et al., 2016))

$$a \leftarrow g(P(y|X = x), E(Y_{X=x'}|X = x))$$

Exploit the "Independence of Causal Mechanisms" assumption: $P(X)$ contains no information about $P(Y|X)$ (Schoelkopf et al., 2012)

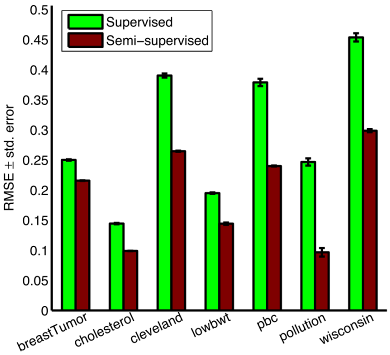


- *Smoking* $\rightarrow$ *Cancer*
- $P(Y)$ contains no information about $P(Y|X)$

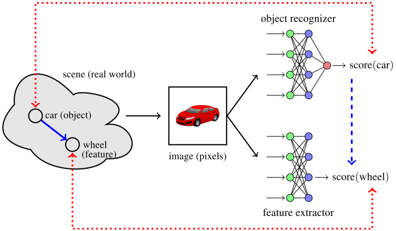


- *Cancer* $\rightarrow$ *Smoking*
- $P(Y)$ contains information about $P(X|Y)$
- Having more samples from $P(Y)$ helps

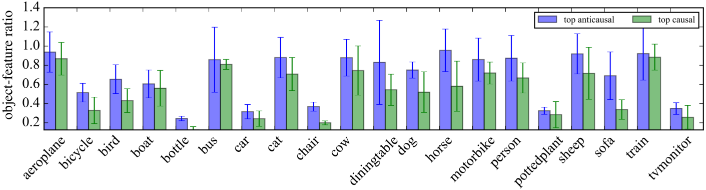
Efficiency: Semi-Supervised Learning Improvement



1. Exploit the "anticausal" relationship to better classify objects (Lopez-Paz et al., 2016), see also (Chalupka et al., 2014)
2. Learn causal directionality classifier given data $S_i = \left\{\left(x_{ij}, y_{ij}\right)\right\}_{j=1}^{m_i} \sim P^{m_i}\left(X_i, Y_i\right)$ with directionality label $\ell_i \in \{\leftarrow, \rightarrow\}$
3. Classify NN features as "causal" or "anticausal"
4. Check $\exists\ dep(f_{ac}, l_f)$



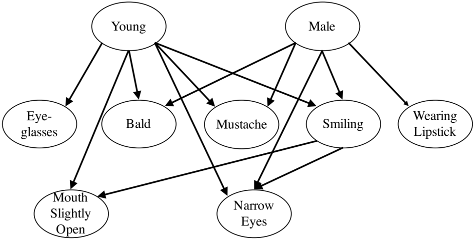
Representation: Learning Causal Features



Generating from interventional, not conditional distributions (Kocaoglu et al., 2017)

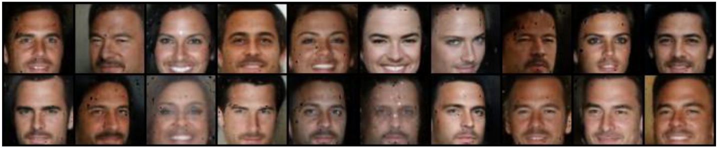
$$P(Image|do(Moustache = 1)) \neq P(Image|Moustache = 0)$$

1. Pass in a causal model $\mathcal{M}$:



2. Train a "Causal Controller" that generates labels from $\mathcal{M}^{\delta}$ or $\mathcal{M}$
3. Train GAN that takes these labels and data as input

Representation: Interventional Sampling



(a) Intervening vs Conditioning on Mustache, Top: Intervene Mustache=1, Bottom: Condition Mustache=1

(a) Intervening vs Conditioning on Bald, Top: Intervene Bald=1, Bottom: Condition Bald=1

1. Causal tools help beyond causal inference (structure learning, IV's, TE Estimation, PSM, etc.)
   - They play a role in the learning and doing process
2. Exploiting causal information can improve learning transfer, efficiency, and representation learning
3. There are opportunities to integrate into more general, traditional machine learning methods and tasks
   - Bayesian modelling, active learning & Bayesian Optimization & RL, representation learning, generalization
   - Interventions, counterfactuals, causal models

1. Upcoming review paper on this work
2. arXausality → my Github repo of all recent causal ML papers
3. Causal Learning Workshop @ NeurIPS 2018
4. Pyro,DoWhy, CDT → Packages for expressing and testing causal models
5. Causality Workbench → data & challenges for causal inference tasks

References I

E. Bareinboim, A. Forney, and J. Pearl. Bandits with Unobserved Confounders: A Causal Approach. *Nips*, (November):1–9, 2015. ISSN 10495258. URL https://papers.nips.cc/paper/5692-bandits-with-unobserved-confounders-a-causal-approach.pdf.

P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner, C. Gulcehre, F. Song, A. Ballard, J. Gilmer, G. Dahl, A. Vaswani, K. Allen, C. Nash, V. Langston, C. Dyer, N. Heess, D. Wierstra, P. Kohli, M. Botvinick, O. Vinyals, Y. Li, and R. Pascanu. Relational inductive biases, deep learning, and graph networks. jun 2018. URL http://arxiv.org/abs/1806.01261.

K. Chalupka, P. Perona, and F. Eberhardt. Visual Causal Feature Learning. dec 2014. URL http://arxiv.org/abs/1412.2309.

A. Gopnik and D. M. Sobel. Detecting blickets: How young children use information about novel causal powers in categorization and induction. *Child Development*, 71(5):1205–1222, 2000. ISSN 00093920. doi: 10.1111/1467-8624.00224. URL http://www.ncbi.nlm.nih.gov/pubmed/11108092.

K. Kansky, T. Silver, D. A. Mély, M. Eldawy, M. Lázaro-Gredilla, X. Lou, N. Dorfman, S. Sidor, S. Phoenix, and D. George. Schema Networks: Zero-shot Transfer with a Generative Causal Model of Intuitive Physics. jun 2017. URL http://arxiv.org/abs/1706.04317.

M. Kocaoglu, C. Snyder, A. G. Dimakis, and S. Vishwanath. CausalGAN: Learning Causal Implicit Generative Models with Adversarial Training. 2017. URL https://arxiv.org/pdf/1709.02023v1.pdfhttp://arxiv.org/abs/1709.02023.

B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum. One-shot learning by inverting a compositional causal process. *Advances in Neural Information Processing Systems 27 (NIPS 2013)*, pages 1–6, 2013. ISSN 10495258. URL https://papers.nips.cc/paper/5128-one-shot-learning-by-inverting-a-compositional-causal-process.pdf.

B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman. Building Machines That Learn and Think Like People, 2016. ISSN 14691825. URL http://cims.nyu.edu/{~}brenden/http://www.mit.edu/{~}tomeru/http://web.mit.edu/cocosci/josh.htmlhttp://gershmanlab.webfactional.com/index.html.

F. Lattimore, T. Lattimore, and M. D. Reid. Causal Bandits: Learning Good Interventions via Causal Inference. jun 2016. URL http://arxiv.org/abs/1606.03203.

D. Lopez-Paz, R. Nishihara, S. Chintala, B. Schölkopf, and L. Bottou. Discovering Causal Signals in Images. pages 6979–6987, 2016. doi: 10.1109/CVPR.2017.14. URL https://arxiv.org/pdf/1605.08179.pdfhttp://arxiv.org/abs/1605.08179.

S. Magliacane, T. van Ommen, T. Claassen, S. Bongers, P. Versteeg, and J. M. Mooij. Causal Transfer Learning. 2017. URL https://arxiv.org/pdf/1707.06422.pdfhttp://arxiv.org/abs/1707.06422.

T. McCormack, N. Bramley, C. Frosch, F. Patrick, and D. Lagnado. Children's use of interventions to learn causal structure. *Journal of Experimental Child Psychology*, 141:1–22, jan 2016. ISSN 00220965. doi: 10.1016/j.jecp.2015.06.017. URL https://www.sciencedirect.com/science/article/pii/S0022096515001642{#}f0005.

J. Pearl. *Causality: Models, reasoning, and inference*. 2000. ISBN 9780511803161. doi: 10.1017/CBO9780511803161.

M. Rojas-Carulla, B. Schölkopf, R. Turner, and J. Peters. Causal Transfer in Machine Learning. jul 2015. URL http://arxiv.org/abs/1507.05333.

B. Schoelkopf, D. Janzing, J. Peters, E. Sgouritsa, K. Zhang, and J. Mooij. On Causal and Anticausal Learning. jun 2012. URL http://arxiv.org/abs/1206.6471.

V. Thomas, J. Pondard, E. Bengio, M. Sarfati, P. Beaudoin, M.-J. Meurs, J. Pineau, D. Precup, and Y. Bengio. Independently Controllable Factors. aug . ISBN 9783642232169. doi: 10.1007/978-3-642-23217-6_3. URL http://arxiv.org/abs/1708.01289.