The University of Queensland

PHYS3040 - 2020 - Semester One

Benchmarking the CausalNex Package

Literature Review - Sam Bethune - 43536319

Introduction

There goes a saying in statistically oriented circles that 'correlation does not imply causation'; often this statement is accompanied by a parable of the following variety.

> Countries with higher birth rates are known to have a higher per-capita population of storks. Therefore, we see that storks must be the bringers of babies. *(Schölkopf et al., 2013).*

While this is innocuous enough, it is not difficult to imagine a more complex or malign scenario. Consider the problem of determining whether two correlated genes might indicate that one causes the other. Do criminals with a history of violent offences tend to subsequently engage in substance abuse or are things the other way around? Might there be a mediator or confounder present? In describing The Four Great Errors to which humanity was prone, Nietzsche made sure to include mistaking cause for effect and vice versa (2003).

Delving into such matters brings to our attention that the language of statistics, while able to tell us all we need know concerning correlation, speaks nothing of causation. For that we have mathematics and physics; as an example we can specify different varieties of force and so ask all-important "counterfactuals". These are questions of the sort 'what if there had been a star where before there was vacuum?' The importance of being able to ask and answer such questions we need not outline, but what if for whatever reason we are unable to obtain a working mathematical model? Surely before Newton came to his inverse square law he had to ask whether two masses gave rise to a mutual force.

Pearl (2018) takes this opportunity to introduce the role of causality, as an intermediary between statistics and mathematics. While causality subsumes statistics, mathematics and physics subsume causality (and accordingly statistics also). Most often we describe causality in the language of graph theory. Here we are interested in a computational approach to causality, in which we further restrict our descriptions to directed acyclic graphs (DAGs).

Computational work in causality typically revolves around the idea of querying counterfactuals without actually performing the intervention in question. We want to know what would happen if there were a star at a given position without actually putting one there to find out. For this we require the construction (again without actually performing any interventions if possible) of a causal graph, for which there exist many algorithms of widely varying design.

In the majority of early designs, the number of conditional independence tests (the equivalent of finding the error for a given node of a neural network during back-propagation) to compute a graph increases superexponentially with the number of nodes (Li, 2015). Now however, many models are less computationally expensive to scale and in general computing power has (according to Moore at least) increased exponentially itself. As such the distribution of open source causal algorithms able to be run on a personal computer has become widespread.

Given the recent global outbreak of COVID-19, we are interested in applying such models to a causal analysis of the related data. For reasons to be outlined in the body of this report, we are drawn to the CausalNex library for Python and its associated NOTEARS causal discovery engine in particular. In the NOTEARS release whitepaper (Aragam et al., 2018) benchmarking is performed only on synthetic data, leaving doubt concerning its performance on the COVID-19 dataframe of the future.

As such we will use a present COVID-19 dataset to evaluate its performance against more established algorithms, in particular PC and FCI in anticipation of what is to come. This literature review is intended to provide insight into the motivation and direction behind our research as well as an evaluation of the sources studied thus far. We additionally hope that it may provide a brief panorama of causal inference, in particular its relationship to Machine Learning (ML), to any who are interested.

Li, J., Liu, L., Le, T. D. (2015). *Practical Approaches to Causal Relationship Exploration*. London, United Kingdom: Springer.
Pearl, J. (2018). *The New Science of Cause and Effect (Conference Presentation)*. PyData Los Angeles 2018: University of Southern California, Los Angeles, USA. Retrieved from: `https://www.youtube.com/watch?v=ZaPV1OSEpHw`
Nietzsche, F. (2013). *Twilight of the Idols and the Antichrist (trans. Hollingdale, R. J.)*. New York, USA: Penguin Books.
Schölkopf, B., Janzing, D. (2013). *Causality (Conference Presentation).* Machine Learning Summer School 2013: Max Planck Institute for Intelligent Systems, Tübingen, Germany. Retrieved from: `https://www.youtube.com/watch?v=KsbftkwZTq4` (Part One) `https://www.youtube.com/watch?v=Y5M6qwbidu0` (Part Two)

Aragam, B., Zheng, X., Ravikumar, P., Xing, E. P. (2018).

DAGs With NOTEARS: Continuous Optimization for Structure Learning. Retrieved from: `https://arxiv.org/abs/1803.01422`

Central to our work, this whitepaper describes the NOTEARS algorithm as touched on in *Introduction*. Aragam et al. begin by surveying the landscape of causal algorithms available, focussing on constraint and score based methods. Respectively these are those that have a conditional independence test or score function for a proposed DAG as their foundation; NOTEARS is itself a score based method. The authors outline the ways in which they hope to improve on current approaches. They primarily hope for a model that is computationally frugal, able to be parallelised (iterated globally) and conceptually simple.

As opposed to the wealth of existing score based methods, NOTEARS replaces selection from a discrete pool of potential DAGs with convergence to a local optima via the gradient descent family of numerical methods. While the benefits of this approach are many-fold, Aragam et al. emphasize the inherent speed that comes with dependence on the well studied problem of gradient descent given the equality of local minima as well as the simplicity and flexibility of their implementation.

It is our conclusion that the authors are well placed in their ardent enthusiasm regarding their findings. The NOTEARS algorithm, described as state of the art in the CausalNex documentation discussed below, shows great promise in the testing published as part of this paper. While we have already mentioned that NOTEARS is yet to be benchmarked against a real causal dataset, we here emphasize the artificiality of the testing conducted by Aragam et al. Test graphs were randomly generated with designated numbers of edges for each node in a given test, not an effective representation of the engine's performance on an actual (likely sparse) DAG. In all other respects however we found the paper to be of an excellent standard, particularly as regards clarity.

CausalNex. (2020).

A First CausalNex Tutorial. Retrieved from: `https://causalnex.readthedocs.io/en/latest/03_tutorial/03_tutorial.html`

In order to benchmark the CausalNex library we will first require guidance as to its optimal use. Meeting this need, *A First CausalNex Tutorial* provides an overview of the package's workings via a Jupyter Notebook exploring a dataset describing exam performance. The tutorial covers the major aspects of the library's functionality including model initialization, DAG fitting, feature discretization, Bayesian network fitting via the NOTEARS algorithm and support for querying counterfactuals via Pearl's do-calculus. The quality of the tutorial aside, we find the package to be of a very high standard with support for many state of the art related technologies.

Returning to the quality of the tutorial itself we describe it as comprehensive, clear and practical if somewhat brief. Especially given its title, we suspect that further notebooks were/are planned but are yet to arrive. Additionally we regret that support regarding parallelisation and GPU implementation was not included as part of the documentation but wish to reaffirm our overall high opinion of *A First CausalNex Tutorial.* We note that the professional standard of CausalNex and its related infrastructure is unsurprising given the package's strong affiliation (despite being open source) with McKinsey Co.'s QuantumBlack data science consultancy, an established industry leader.

Howard, J., Thomas, R., Gruger, S. (2019).

....

Practical Deep Learning for Coders: Version Three (Lecture Series) Retrieved from:
`https://course.fast.ai/`

....

Howard et al's MOOC teaches deep learning via the use of their FastAI Python ecosystem. FastAI, as opposed to competitors such as Keras, is journalistic in nature and intends to provide users with state of the art technology via the extensive use of black boxes. Besides the importance to us of gaining an orientation in the broader landscape of data science before pursuing causality in particular, the MOOC was of great relevance given the NOTEARS algorithm's alignment with ML methods. Covering not only the use of FastAI itself but also deep learning fundamentals and industry best practice, great care is taken to position students for the future as expected given FastAI's forward thinking ethos.

The quality of the lectures is outstanding and on its own a compelling reason to investigate FastAI. Although lengthy and heavily loaded with content, they are always clear not to mention highly relevant to modern deep learning practitioners. Particularly relevant to our research was the coverage of learning rate manipulation, regularization schemes, metric adjustment and commercial GPU infrastructure.

Kalainathan, D., Goudet, O., Guyon, I., Lopez-Paz, D., Sebag, M. (2019).

....

Structural Agnostic Modelling: Adversarial Learning of Causal Graphs. Retrieved from:
`https://arxiv.org/abs/1803.04929`

....

This paper stands somewhat aside from the others listed. Initially considered along with its parent library Causal Discovery Toolkit as an alternative to NOTEARS with CausalNex, we consider Structural Agnostic Modelling (SAM) worthwhile discussing nonetheless. Rather than a constraint or score based methodology, Kalainathan et al. have here taken inspiration from the rise of Generative Adversarial Network (GAN) deep learning technology in what they call a generative approach. A most intriguing platform, their ideas combine Kolmogorov complexity and the Minimum Description Length (MDL) approximation with a single layer neural network. As in the case of NOTEARS, this allows for incorporation of the well studied problem of local minimisation via gradient descent.

Our decision to pass over SAM in favour of NOTEARS was largely due to SAM's dependence on a wealth of training data. While both models iterate via the gradient descent family, NOTEARS allows for (and is in fact reliant on) input from a 'human expert'. SAM on the other hand is initialized as a randomly weighted vector in accordance with neural network protocol. Not knowing what the COVID-19 dataset of the future will resemble, we consider NOTEARS to be a safer investment of our time. However, we include SAM in this paper for the benefit of readers who may think of a solution or otherwise be interested. For a start we propose that data augmentation methods and network pretraining (for example on influenza data) could both be viable in seeking to employ the algorithm under a small training set.

Li, J., Liu, L., Le, T. D. (2015).

Practical Approaches to Causal Relationship Exploration. London, United Kingdom: Springer.

Li et al's *Practical Approaches to Causal Relationship Exploration* is but one in Springer's *Brief's in Electrical and Computer Engineering* series. Intended to take readers with a technical background from being completely unaccustomed with causal inference to capable of implementing basic computational routines in ∼80 pages, the text focuses on constraint and score based methods. Particular attention is paid to the PC, CR-PA and CR-CS algorithms. All methodologies are introduced theoretically and composed in pseudo-code before being worked through using current open source Python libraries available for download. Discussions of the strengths and weaknesses of each algorithm along with practical commentary regarding their uses are always provided.

The necessity of the book to our research we hardly need emphasize; the descriptions of each algorithm in simple terms saved many hours consulting sources excessively detailed for our purposes. Though rarely compelling, the text fills a vital gap in the literature by allowing for the enormous number of existing professionals with transferrable skills to quickly gain a foothold in causal inference. For those looking to quickly and conveniently understand the results of causal research, we strongly recommend Li et al's brief as a first port of call before further reading.

Ng, A. (2012).

Machine Learning (Lecture Notes). Coursera, Machine Learning (Paid Online Course Materials).

A widely recognised and well established first course on ML, Ng's MOOC was instrumental to us in gaining a thorough understanding of the concepts underlying ML methods. While Howard et al's course focussed on state of the art techniques using the latest tools, Ng taught from the ground up with linear algebra based content and assessments in MATLAB/OCTAVE. Especially emphasised were cost functions, metrics, regularization, model diagnostics, pipelines and programming practices. Also covered were useful concepts such as kernel methods, which are now considered largely legacy technology in deep learning but of increasingly major relevance to causal inference.

As the reader may have begun to tell from the above reviews, the integration of causal inference with ML methods is rapidly taking hold. Given many ML inspired causality engines employ ML elements in a novel way, we found a fundamental working knowledge of them to be indispensable in our investigation. Knowledge gained from Ng's course was essential background in reading Aragam et al and Kalainathan et al's whitepapers as well as listening to Schölkopf et al's lectures. While we can wholeheartedly recommend Ng's *Machine Learning* on the basis of our experience, we understand there exist many introductory ML resources widely considered to be equally as effective. Thus we encourage the reader to find that which seems to best suit their individual background and intentions.

Pearl, J. (2018).

The New Science of Cause and Effect (Conference Presentation). PyData Los Angeles 2018: University of Southern California, Los Angeles, USA. Retrieved from: `https://www.youtube.com/watch?v=ZaPV1OSEpHw`

Following various attempts with other resources, this address of Pearl's was our first successful foray into causal inference. Aimed at a popular audience and focussing more on the philosophical motivations and implications of causality than developing any sort of technical capacity, topics such as DAGs, conditional independence and interventions are covered nonetheless. We consider this presentation to be in the same vein as Pearl's popular *The Book of Why*, which we were ironically unable to obtain due to COVID-19 itself.

Pearl is the preeminant giant of causal inference, having developed many of the field's fundamental tools and been prized with the 2011 Turing Award (equivalent to a Nobel Prize in computer science) for his work. His explanation of the the do-calculus (his own invention) as well as conveyance of the relevance of causality to machine learning are both unsurprisingly invaluable. We also note his natural ability as a public speaker, contrary to many more technically oriented causal research figures. However, Pearl's address here is clearly borrowed from a completely different event and obviously one at which he was expected to speak for much longer. Toward the end of his allotted hour he begins aimlessly flicking through his myriad slides, the content of which appears to be thoughtfully presented and highly relevant. As such we recommend this address to causal inference initiates but suggest that those who are interested consider tracking down the event for which Pearl had originally prepared his material.

Pearl, J., Glymour, M., Jewell, N. P. (2016).

Causal Inference in Statistics: a Primer. West Sussex, United Kingdom: John Wiley & Sons.

As mentioned above, Pearl is a prolific author in causal inference. In a sense this book written alongside Glymour and Jewell is intended to provide a stepping stone to his more advanced publications such as the 2009 *Causality: Models, Reasoning and Inference*, or as in our case to further causal reading in general. Assuming a background of only high school level statistics, the authors progress to a theoretically well developed description of causal inference focussing on constraint based methods. The book features complete mathematical proofs of foundational causal results and describes in depth the behaviour of colliders, properties of d-separated nodes and much more as well as vitally important discussion of the causal Markov, independent Markov and faithfulness assumptions.

The book is clear throughout and was essential to our early understanding of constraint based algorithms and fundamental ability to navigate DAGs in general. Though the computationally oriented reader often feels they are being presented with more logical and mathematical content than necessary, we consider becoming familiar with such ideas useful given it allows for broader interaction with the field of causality. In particular, we feel that one must understand the assumptions necessary in constraint based modelling before perceiving the potential advantages of abandoning them such as in Kalainathan et al and Schölkopf et al. Besides Pearl, Glymour and Jewell are each major figures in causality; this book testifies their commitment to the broader adoption of causal methodology across STEM fields. We recommend it to all intending for a long term engagement with the field.

Peters, J. (2015).

Causality (Lecture Notes). ETH Zürich, Zürich, Switzerland.

Peters' notes form the basis of a script used for a number of introductory causal seminars pitched to an interdisciplinary academic audience. In particular, being himself a well known statistician, Peters assumes a high degree of familiarity with statistics and its associated mathematical formalism. The notes cover all which is necessary to begin understanding the tools and procedures of causal inference and include much by way of mathematical proofs and deep investigations of the assumptions typically present in both constraint and score based methods. We also note the inclusion of proof oriented exercises of varying difficulty at the end of each chapter.

This source is a standout for readers with a strong background in mathematics and statistics; in under 70 terse pages Peters brings the audience from being unaware of causal inference to presenting a considerable understanding of its underlying techniques and dependencies. Especially strong is its discussion of model assumptions and their technical ramifications. In this respect we found it unmatched among the resources here listed and a rewarding investment for those taken with rigour. Much more useful to our research however was the excellent and authoritative bibliography attached, providing something of a 'who's who' of causal inference. Additionally it provided guidance on integrating observational and interventional datasets, potentially a valuable technique in the modelling of the future COVID-19 dataframe.

Schölkopf, B., Janzing, D. (2013).

Causality (Conference Presentation). Machine Learning Summer School 2013: Max Planck Institute for Intelligent Systems, Tübingen, Germany. Retrieved from:
`https://www.youtube.com/watch?v=KsbftkwZTq4` (Part One)
`https://www.youtube.com/watch?v=Y5M6qwbidu0` (Part Two)

Taken in context as a presentation at the *Max Planck Institute for Intelligent Systems*' 2013 *Machine Learning Summer School*, Schölkopf and Janzing seem to have intended these lectures for experts from a variety of data science disciplines. As such they deliver an introduction to causality before delving into kernel methods and Kolmogorov complexity optimization via the MDL, touching on advanced topics from a range of fields. Schölkopf in particular presents a modest overview of his remarkable technical accomplishments including a conditional solution to the double node problem and formulation of state of the art kernel mappings.

Considered global leaders in computational causal inference, Schölkopf and Janzing here seem more intent on piquing the interest of experts from other disciplines than establishing any firm foundations in their field among the audience. While we consider it unlikely any previously uninitiated attendees left feeling ready to begin elementary work in causality, we suspect many would have been interested in the direction of the latest research given the enormous span of the lectures. Accordingly we recommend this presentation for speculative viewing (rather than serious study) by those who have already covered the foundations of causal inference as in Pearl et al. (2016) and Li et al. (2015). For them it may be helpful in gaining an orientation in the current landscape of causal research, as it assisted us in approaching the work of Aragam et al. and Kalainathan et al.

Schölkopf, B., Janzing, D., Peters, J. (2017).

Elements of Causal Inference: Foundations and Learning Algorithms. Cambridge, Massachusetts: MIT Press

In stark contrast to the above source Schölkopf et al. present this work as a standard reference text in the field of causal inference, going so far as to make it freely available online to promote its widepsread adoption. Their coverage of causality is intended to be comprehensive at a basic level. All statements are rigorously proven and much attention is paid to filling out the historical landscape of causal inference in order to leave no doubt as to their authority regarding the subject matter. A wealth of problems are included along with a sizable bibliography; Schölkopf et al's intention to establish this textbook as a future classic is everywhere clear.

We consider the authors to have been successful in creating a reliable and practical resource able to carry entrants to the field through to an advanced level as well as serve as a reference to more seasoned practitioners. It is in the latter capacity that this work distinguishes itself however. We cannot imagine requiring a proof or result of an assumption in our work which Schölkopf et al. have not covered here. As such we consider it a valuable addition to our resource list and one which we suspect will in time become commonplace among the causal inference community. This is especially so given Schölkopf's standing as a leading and highly influential research figure.