***2 year Post-doc position**: Understanding and Explaining Complex Systems with Causal Relationships*

Background

Most ML studies are concerned with learning predictive models that allow to predict the value of some variables, based on the values of others. However, this amounts to learn ‘only’ correlations: another approach to understanding and explaining a complex system is to discover the causal relationships between variables. While randomized controlled experiments will remain the gold standard to determine causal relationships, such experiments are hampered due to practical, economical, or ethical reasons. In such cases, determining causal relationships from observational data is of primary importance ([1,2,3]).

Main activities

A first approach to pairwise causal relation learning has been proposed as a ML task: given enough examples of image pairs for which the relation is known (3 classes: independance, causality, or presence of confounders), the goal is to predict the class for the relation between 2 variables considering the available data for both variables as 2D images [4]. However, too few examples (pairs of variables with known causality relation) are available except in some very specific domains, like biological regulation networks [6] where experiments are possible. As could have been expected, what can be learned from these examples works well in other biological domains, but tends to behave poorly on completely different domains: the so-called "Mother distribution" [8], representing the distribution of distributions of pairs of variables, is too specific, making it impossible to learn a general causation model. On the other hand, domain adaptation has now reached a mature state, transferring knowledge from one domain to another (sufficiently related) one, e.g. using adversarial learning approaches [5]. The idea here is to apply domain adaptation to causation learning, in order to adapt the causation model learned from biological networks to e.g. human resources domains.

Another direction is related to the most recent advances in causal learning in the TAU team, namely SAM (Structure Agnostic Modeling, [7]). SAM also uses the idea of adversarial learning. Considering D variables, SAM builds D neural nets, each aimed to learn the i-th variable from all other variables (and a noise variable). The set of these NNs is used to reconstruct full 'fake' samples. They are jointly learned with a discriminator, aimed to recognize true examples from fake ones, using gradient backpropagation and reverse gradient backpropagation. The overall architecture is enriched with so-called slack variables, supporting the regularization of the approach and enforcing the sparsity of the NNs, under the assumption that the true underlying generative model (causal model) achieves a best trade-off between data-fitting and structural simplicity. SAM represents a great advance in the field, as it discovers the full causality graph at once, without the need to combinatorial optimization (as when starting from pairwise causation strengths and conditional independence). A current limitation of SAM lies in the lack of information sharing among the NNs. In the case where latent confounders control several variables, these would need to be identified several times. Hence a next direction is to consider a brand new neural architecture, inspired from auto-encoders with two differences: i) a fat latent layer (to ensure that all confounders are found); ii) a specific regularization scheme preventing the discovery of trivial relations while nevertheless imposing sparsity.

In both research directions above, a crucial issue is that of existing domain knowledge: in many domains, some partial causality information is available, that it would be a waste to try to rediscover. Incorporating such (symbolic) information is key to incremental learning of causality, and hence to scalability.

Skills

We are searching for motivated candidates with a PhD degree in Computer Science and with competences in machine learning (preferably with focus some knowledge about on neural networks). Knowledge about causal inference would, of course, be appreciated.

The candidate should be proficient in written and spoken English (at least B2 level according to the CEFR system).

Assignment

To apply for the position, the candidate must send an email to the list of contacts below. The email should include:

- A CV
- A statement letter explaining the candidate's motivations to apply for the position
- At least two recommendation letters

Context

This position is open at Inria, in France and is part of the HyAIAI Inria challenge. More specifically, this position is part of a collaboration between the TAU and Lacodam Inria teams. The post-doc will be located in Saclay (in the TAU team). She will actively participate in the activities of HyAIAI, in particular report on her work in the HYAIAI meetings.

Contact people

Marc Schoenauer (marc.schoenauer@inria.fr)
Michèle Sebbag (sebag@lri.fr)
Alexandre Termier (alexandre.termier@inria.fr)
Elisa Fromont (elisa.fromont@inria.fr)

Bibliography

[1] Judea Pearl: *The seven tools of causal inference, with reflections on machine learning*. Commun. ACM 62(3): 54-60 (2019).
[2] J. Peters, D. Janzing, and B. Schölkopf. *Elements of Causal Inference Foundations and Learning Algorithms*. MIT Press, 2017.
[3] Judea Pearl and Dana Mackenzie. *The Book of Why. The New Science of Cause and Effect*. Basic Books, 2018.
[4] Isabelle Guyon et. al. : **Cause Effect Pairs in Machine Learning**. Chalearn Causality Challenges 2019.
[5] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, Victor S. Lempitsky: *Domain-Adversarial Training of Neural Networks,* J. Mach. Learn. Res. 2015
[6] Karen Sachs, Omar Perez, Dana Pe'er, Douglas A. Lauffenburger, Garry P. Nolan: *Causal Protein-Signaling Networks Derived from Multiparameter Single-Cell Data*. Science, vol 308, Issue 5721, pp 523-529 (2005)
[7] Diviyan Kalainathan, Olivier Goudet, Isabelle Guyon, David Lopez-Paz, Michèle Sebag: *SAM: Structural Agnostic Model, Causal Discovery and Penalized Adversarial Learning*. CoRR abs/1803.04929 (2018)
[8] Mateo Rojas-Carulla, Marco Baroni, David Lopez-Paz: *Causal Discovery Using Proxy Variables*. CoRR abs/1702.07306 (2017)