Learning Causal Structures from Data

JOHAN SCHUBERT, RONNIE JOHANSSON

FOI-R--4882--SE
ISSN 1650-1942 December 2019

Johan Schubert, Ronnie Johansson

Learning Causal Structures from Data

Titel	Learning Causal Structures from Data
Title	Inlärning av kausala strukturer från data
Rapportnr/Report no	FOI-R--4882--SE
Månad/Month	December
Utgivningsår/Year	2019
Antal sidor/Pages	18
ISSN	1650-1942
Kund/Customer	Försvarsdepartementet/Ministry of Defence
Forskningsområde	Ledningsteknologi
FoT-område	Inget FoT-område
Projektnr/Project no	A61021
Godkänd av/Approved by	Cecilia Dahlgren
Ansvarig avdelning	Försvars- och säkerhetssystem

Bild/Cover: Johan Schubert

Sammanfattning

I denna rapport genomför vi en avskanning av forskningen inom inlärning och upptäckt av kausala relationer inom artificiell intelligens (AI). Motivationen för att undersöka detta forskningsfält är finna hur viktig kausalitet är i militärt beslutsfattande. Området maskininlärning inom AI har sett stora framsteg under de senaste åren. Det finns emellertid en inneboende begränsning i den dominerande sortens maskininlärningsmetoder, vilka bygger på att hitta korrelationer i data. En korrelation mellan två händelser säger inget om huruvida den ena har orsakat den andra eller om en tredje händelse orsakar båda två. Genom att även ge algoritmer en uppfattning om kausalitet blir det möjligt att bättre förstå och resonera om omvärlden. För att tillåta användning av maskininlärning i militära system som normalt endast identifierar korrelationer mellan händelser och fenomen så måste man veta alla möjliga orsakssamband på förhand. Detta kräver en mycket hög förståelse av alla möjliga händelser och fenomen. Om denna kunskap inte är tillgänglig så är metoder för att lära sig kausala relationer nödvändiga.

Nyckelord: Artificiell intelligens, maskininlärning, kausal inlärning, kausal upptäckt, beslutsstöd, lägesbild, militär.

Summary

In this report we perform a horizon scanning over the research field of causal learning and discovery within artificial intelligence (AI). The motivation for investigating this field of research is to find out how important causality is in military decision-making. The area of machine learning within AI has seen great progress in recent years. However, there is an inherent limitation in the machine learning methods that focus on finding correlations in data. If correlation between two events is detected, we do not know if either one has caused the other or whether a third event is causing both. By also giving algorithms an idea of causality, it becomes possible to better understand and reason about the outside world. In order to allow the use of machine learning in military systems that normally only identify correlations between events and phenomena, one must know all possible causal relationships in advance. This requires a very high understanding of all possible events and phenomena. If this knowledge is not available, methods for learning causal relationships are necessary.

Keywords: Artificial intelligence, machine learning, causal learning, causal discovery, decision support, common operational picture, military.

Contents

1 Introduction

The machine learning area has seen major breakthroughs and reaped significant successes in the 2010s (for example in image recognition, self-driving cars, Jeopardy profits, Go games, etc.) and thereby rekindled interest in the use of *artificial intelligence* (AI) in both the civilian and military arena [1][2]. The development of deep learning is probably the biggest success of AI of the decade.

Casual learning is now a topic expanding into larger AI conferences. For example, at the annual *33rd Conference on Neural Information Processing Systems* there was a workshop entitled *"Do the right thing": machine learning and causal inference for improved decision making*[1]."

Judea Pearl, one of the pioneers in the AI field, has long studied causality (i.e., *causation*) and believes that there is an inherent limitation in the AI methods currently being explored [3]. According to Pearl, the current research focus is on understanding associations, i.e., correlations in data. By also giving algorithms an idea of causality, it allows them to better understand and reason about our world [4][5] and thus take a proper step towards general artificial intelligence that the current direction cannot achieve.

If the data under investigation (e.g., *to make predictions*) come from a time series or from a hierarchy of data sources (e.g., *parts of a military organization*) then a higher understanding is gained by looking for causal relationships instead of only correlations, by using machine learning methods that take causality into account.

For the military sector, the benefit of AI is to deliver decisive support when time is too short or when the number of choices is too large for decision makers to analyze all alternative actions. Using machine learning that discovers correlations only may not be fully satisfying as a means of military decision support. With discovered correlations between two events, we do not know if one has caused the other or if a third event causes both. It is thus necessary to discover the actual causal effects involved and to describe them in decision support systems that interact with a decision maker and communicate the true causes of events.

Consequently, the motivation to investigate this field of research is to find out just how important causality could be in military decision making by studying the research frontier in *causal discovery and learning*. This should be evaluated from the benefits that can be achieved by the adopting methodologies in the state-of-the-art of the field from a military perspective. Possible questions to consider are: What new applications will be possible? How can existing applications be adapted and improved upon? To what extent can a *common operational picture* (COP) be improved when causal relationships are considered between events in the COP?

This document reports on the internal project *Learning Causality* conducted at the Swedish Defence Research Agency (FOI)[2].

---

[1] https://neurips.cc/Conferences/2019/Schedule?showEvent=13176 (december 2019)

[2] https://www.foi.se/en/foi.html (december 2019)

2 Horizon Scanning of Causal Learning and Discovery

In this literature study, we performed horizon scanning in the area of learning causal relations within AI. This area is scanned with FOI's computer tool HSTOOL [7][8]. It is a computer tool that implements a methodology for scanning scientific literature for the purpose of detecting scientific trends. With this methodology, literature within a research field can be automatically grouped into clusters by subject content, and ranked with respect to its influence within each subject area.

With HSTOOL we have searched the literature in the *Web of Science* [3] (WOS) *Core Collection* database at Thomson Reuters.

When we analyse the material found through a search on causal learning, we find a slowly rising trend in the number of publications per year from 2001 until 2011, followed by a rapidly increasing number of publications between 2011 and 2018, rising from 20 to 68 per year, see figure 1. This growth in publications coincides with the high growth observed within the larger area of machine learning in general.

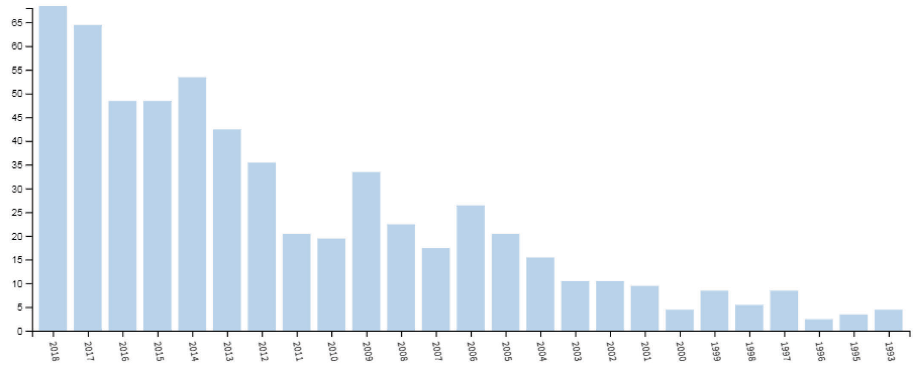

Figure 1. The number of research publications on learning causal relations per year 1993–2018.

The search question used in the scanning of WOS is:

```
TS = (causal* NEAR/2 (learn* OR discover*)) AND
        SU = (Computer science OR Mathematics)
```

where `TS` is a field label which means *Topic*. With `TS` we search with our own keywords in all subject fields. These subject fields include titles, abstracts, keywords and indexing fields such as systematics, taxonomic terms and descriptors for each scientific publication. The `*`-operator is a truncation wild card. The field label `SU` means *Research Area*. All scientific articles in WOS are classified with a specified research area. `SU` limits the search to the specified research areas. The `NEAR/2` operator used above means that the surrounding search terms should not be separated by more than two intermediate words. AND and OR are logical operators.

This search term is selected to catch both the *causal learning* and *causal discovery* terms used within the research area. We limit the search to *computer science* or *mathematics* to avoid research on learning in social science and the humanities.

[3] http://www.webofknowledge.com (december 2019)

We find that most publications are within the computer science research area with a strong focus on AI, figure 2.

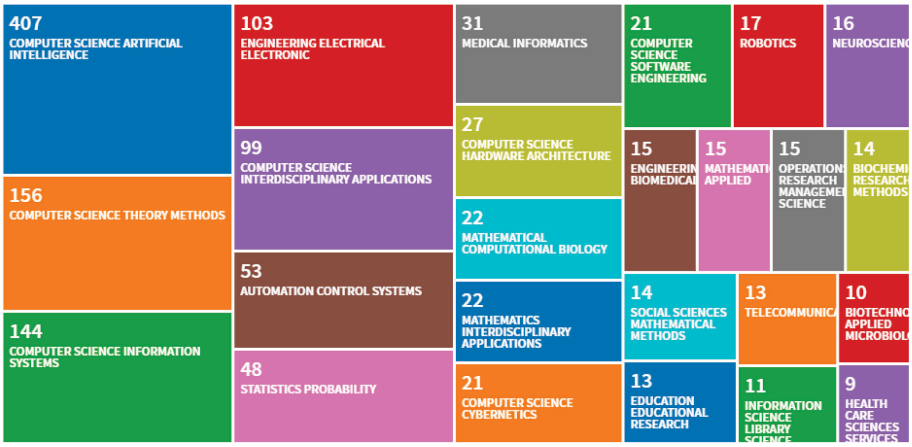

Figure 2. Research areas of publications (i.e., WOS categories) with the number of publications for each area.

The trend is even clearer when looking at the number of citations per year. We observe a strong citation trend since around 2010 rising from approximately 200 citations per year to more than 1800 citations in 2018 (figure 3).

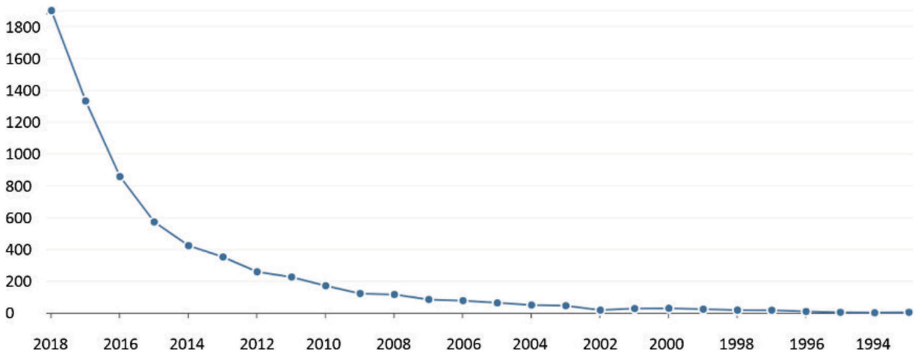

Figure 3. Number of citations per year.

Looking at which countries are performing the majority of research within the area, we find the United States followed by Europe, China and Australia. Within Europe, there are significant contribution from Germany, UK, and the Netherlands (figure 4).

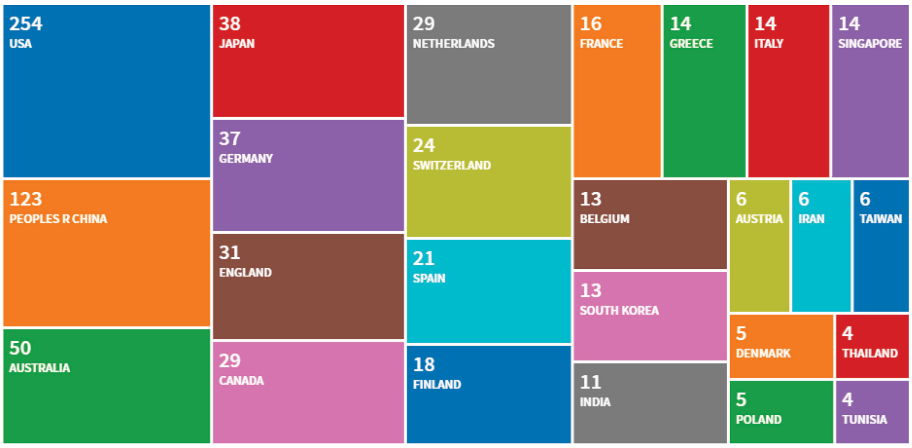

Figure 4. Countries performing research on learning causal relations (with number of publications).

In China, research within the area tends to be funded by their National Natural Science Foundation. In the US, research is funded by the NSF, NIH, DoD, etc., and in Europe, by the EU, ERC and the Academy of Finland (figure 5).

Figure 5. Funding agencies for research on learning causal relations.

Finally, we list the largest research performers with the fields of causal learning and causal discovery. We notice Max Plank and ETH in Europe, and University of Pennsylvania, Carnegie Mellon, University of California, and the University Pittsburgh in the United States among the top seven performers.

Figure 6. Research performing organizations.

3 The Field of Learning Causality

Before we start investigating the research literature on causal learning, let us first provide an overview over the field of causality research. We briefly cover the concept of causality, representations, and different types of learning.

3.1 The Concept of Causality

We will not fully penetrate or dwell on the philosophical aspects of causality, such as "Is causality even a 'real-thing' in the world or is it just a tool for the conscious mind to interpret the world?" It is however, worth highlighting how elusive and relative the concept is.

Merriam-Webster's online dictionary[4] provides the following definition of causality:

1. a causal quality or agency,
2. the relation between a cause and its effect or between regularly correlated events or phenomena.

The definition emphasises a "relation between a cause and its effect". In order to grasp the concept, also "cause" needs to be explained.

Merriam-Webster proposes a number of explanations of "cause" including[5]:

1. a reason for an action or condition : motive,
2. something that brings about an effect or a result,
3. a person or thing that is the occasion of an action or state,
4. sufficient reason.

We tend to have an intuitive understanding of causality, which perhaps best fits with the second definition of "cause" above, i.e. "something that brings about change". For instance, due the cause "Rain", "Wet lawn" became the effect. In a military context, "something" would, e.g., be an agent with the power to reason and act, such as an adversarial military force.

Definitions one and three, i.e. motive or a person, as the cause, could also be considered depending on the needs of the reasoner. For instance, compare "Peter's passion for bowling caused the bowling pin to fall" with "The ball that Peter rolled made the bowling pin fall". Depending on the needs of the reasoner, either Peter, the rolling ball, or even the existence of the game of bowling, could be interpreted as the "cause". Hence, inevitably causality incorporates a subjective sense of abstraction and interpretation.

For further reading on various interpretations of causality, we refer to Waldmann and Mayrhofer [9] who present the contending dependency, dispositional and process frameworks for causality.

[4] https://www.merriam-webster.com/dictionary/causality (december 2019)

[5] https://www.merriam-webster.com/dictionary/cause (december 2019)

3.2 Representation of causality

A *causal model* is "a mathematical abstract that quantitatively describes the causal relationships between variables" [10]. A popular causal model is the so-called *structural causal model* [11], which consists of a *causal graph* and *structural equations*. A causal graph is Bayesian network (consisting of a set of nodes representing a set of variables, V, and a set of directed edges representing dependence, E, a so-called directed acyclic graph or DAG) with the requirement that a dependence strictly represents causality.[6] The structural equations "specify the causal effects represented by the directed edges in the graph." For instance, for the causal graph G in figure 7, with variables A, B and C, the structural equation for C, f_C, yields $C = f_C(A, B, X_C)$, where X_C is a hidden random variable representing uncertainty concerning the causal relationship between A, B, and C.

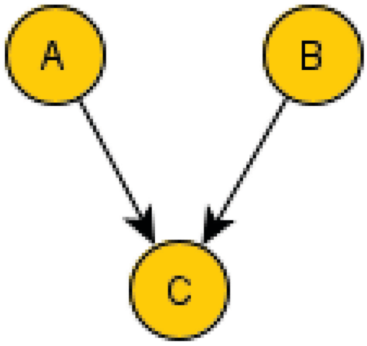

Figure 7. A causal graph is a special case of a Bayesian network

Although causal graphs are frequently used to represent causality, it should be stressed that alternative types of representations also have the expressiveness to capture causality, including if-then-rules, finite state machines, and Petri nets.

3.3 Types of Learning

When categorizing the various forms of causal learning, we follow the presentation by Guo *et al.* (2019) [10].

The domain of causal learning is subdivided into two categories:

- Causal inference,
- Causal discovery.

Simply put, *causal inference* (also learning causal effects) deals with learning about how different variables respond to changes in other variables, given an existing causal model. For instance, how does the expected value of variable C (in Figure 7) vary with changes in variable B?

According to [10], most research in causal learning has been devoted to causal inference. Our focus in this report, on the other hand, is on the second category, i.e. *causal discovery* (also learning causal relationships). Causal discovery concerns learning causal relationships based on a set of data, and possibly some assumptions.

[6] Although in designing Bayesian networks it often assumed that directed edges represent causality, it is not compulsory.

There are basically two classes of methods for causal discovery: *constraint-based* and *score-based*. Constraint-based[7] methods try to learn a structure which respects the conditional independences between variables present in the data. For instance, the two-step Peter-Clark algorithm first learns an undirected causal graph (a so-called skeleton graph) from data, by starting with a fully connected graph, and then systematically removing edges between variables which are conditionally independent in the data. In the second step, the directions of the edges are estimated. Score-based methods tend to appear the optimization problem

$G^* = \arg\max_{G \text{ DAG over } V} Q(V, G)$, where the typically enormous space of possible graphs is heuristically explored to find the best fit model, G^*(with respect to Q).

[7] Sometimes called *independence-based*.

4 Analysis of the Literature on Causality

The papers downloaded from WOS are processed in several steps within HSTOOL. They are first clustered into groups by a *Gibbs sampling Dirichlet multinomial mixture model* (GSDMM) algorithm [6] for clustering and we introduce a complementary method to determine the optimal number of clusters [7][8]. We then use scientometric measures that identify articles that have made a significant impact. This influence is measured using citation statistics and is used to rank all articles within each group, see figure 8.

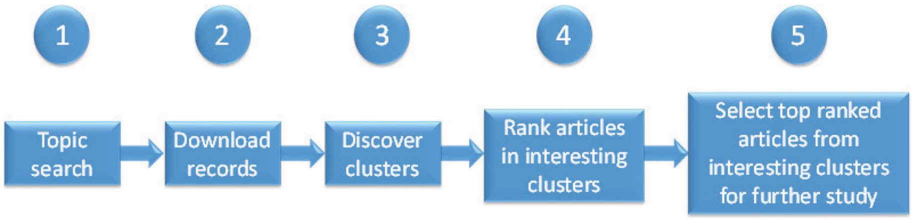

Figure 8. Proposed workflow for horizon scanning of scientific literature.

Using these methods we may be able to automatically discover previously unknown categories of research papers in the field of causal discovery and learning. Within each such category all articles are ranked by their importance.

What follows next is a review of some of the most important findings of this study. In total we found 554 publications using the search question that was shown in section 2. These were automatically clustered into 89 different categories. The articles we mention below are the most highly cited articles in the last five years. Some of them are new, being the latest articles within an important trend, and some of them are slightly older and often are early articles that define the new trend of causal learning and discovery.

Methods using causal discovery in situations with non-experimental data is a problem type with many different applications. While it is not possible prove the full correctness of any causal model derived from observational data, these methods are important in situations where it is impossible to conduct experiments. Shimizu *et al.* (2006) [12] have developed a methodology for discovery of the complete causal structure of a problem with continuous data. This method handles the situation when data is generated by a linear process and there are no non-observable confounding variables (including additional requirements). The methodology is based on *independent component analysis* (ICA)[8]. The authors have developed a complete MATLAB package to perform analysis with their method; linear non-Gaussian acyclic model. Their paper is one of the most influential in causal discovery and learning to date, being cited 685 times since its publication (as of 3 December 2019).

Shimizu *et al.* (2011) [13] has also developed a non-iterative method for estimating the causal order between multiple variables. Previous iterative methods using structural equation models and Bayesian networks to analyze causal relationships between continuous variables were never guaranteed to converge to a solution. Unlike the previous methods, this new method does not require any algorithmic parameters and is guaranteed to converge to the right solution within a small fixed number of steps if all model assumptions are met.

Later Peters *et al.* (2014) [14] investigated how to learn causal acyclic graphs from a joint probability distribution[9]. Such graphs can be used to predict even developments when the available information is deficient. The authors show that if the observation probability

[8] https://en.wikipedia.org/wiki/Independent_component_analysis (december 2019)

[9] https://en.wikipedia.org/wiki/Joint_probability_distribution (december 2019)

distribution follows a structural equation model[10] with added noise, the directed acyclic graph becomes identifiable from the distribution under simple conditions. Experiments indicate that methods based on restricted structural equation models can outperform traditional constraint-based methods.

Bayesian networks may be used in data mining. Heckerman (1997) [15] demonstrated how to use Bayesian networks for learning causal relationships from prior knowledge. Later Ellis and Wong (2008) [16] studied directed acyclic graphs and developed a method for inference of directed acyclic graphical structures. They proposed a computationally fast algorithm using a combination of several methodologies including Markov chain Monte Carlo for determining Bayesian network structures from experimental data.

In 2012 Colombo *et al*. (2012) [17] studied the problem of learning causal relations between random variables in a directed acyclic graph when allowing for any number of non-observable variables and selection variables. They developed an algorithm called *Really Fast Causal Inference Algorithm* (RFCI). They compared it to a previous algorithm called *Fast Causal Inference Algorithm* (FCI) and they show in simulations that the two algorithms are similar in output but that RFCI is much faster than FCI. The new algorithm can be used to assess the impact of non-observable variables or selection variables. It can also be used to find boundaries for causal effects that are based on observational data from an unknown underlying causal graph. Also worth noticing is that Kalisch *et al*. (2012) [18] have developed an R-package[11] for causal structure learning.

There are various challenges associated with the causal discovery algorithms. Isozaki (2012) [19] addresses the issue that the conditional independence tests in constraint-based methods fail when data sizes are small. Isozaki presents a new conditional independence test based on results from thermodynamics and shows improvements when applied to the Peter-Clarks algorithm.

Yu *et al.* (2018) [21] address another challenge. First they point out that the property of Markov blankets (MBs) of BNs is a useful tool for causal discovery. The MB of a node in a BN represents the "node neighbourhood" of the node in question. Hence, if the MBs of all nodes can be estimated, then the complete causal structure can be uncovered. The authors further point out that the discovery of MBs dictates an assumption of *causal sufficiency*, which means that there is a hidden factor (a latent variable) which makes two variables appear in a direct causal relation. This assumption is in many cases too strong and there is a need for MB discovery techniques that relax that assumption. To deal with the problem, the authors represent the problem using a structure called *maximal ancestral graph* which allows representing latent common causes. They then propose a method to efficiently discover the MB of a variable.

Taken together, these publications demonstrate that, while this research area has a strongly growing trend (figure 1), the field has already reached some initial level of maturity with several approaches for different problem areas ready for engineering applications today. We do, however, expect several more years of growth, since the total number of publications is still very small compared to machine learning research, focusing only on finding correlations between events and phenomena. As machine learning is increasingly used to reason about high level events, causal explanations going beyond correlations become necessary.

[10] https://en.wikipedia.org/wiki/Structural_equation_modeling (december 2019)

[11] https://www.r-project.org (december 2019)

5 The Military Perspective on Causality

Understanding the causal relationships between different events and phenomena is crucial in decision support systems for both civil and military applications. Stigler [19] states that understanding causality is a *critical issue* for military officers. Casually assuming causal relationships or simply ignoring them may lead to erroneous conclusions and inappropriate, possibly catastrophic, action. On the other hand, understanding causality in collected data provides an opportunity to understand the relationships between events and thus a deeper understanding and explanation of the current situation picture and in prediction of future events. If the data is available, causal discovery could assist officers to discover unknown or unexpected causal relations and refute previous causal relations held to be true. Methods of causal discovery and learning can be used in any system that seeks to understand the outside world when data comes either from time series or from different parts of hierarchically ordered structures such as military units.

To allow using machine learning in military systems that normally only identify correlations between events and phenomena, one must know all causal relations in advance. This requires a very high prior understanding of all possible events and phenomena. If this knowledge is unavailable, methods for learning causality are necessary.

Although we have not yet found examples in the research literature of learning causality in military applications and of its use in military decision support systems, there are examples of civilian use. We therefore anticipate that causality learning and discovery will be of importance in future military systems if they have components that use machine learning.

6 References

[1] Schubert, J. (2017). Artificiell intelligens för militärt beslutsstöd. FOI-R--4552--SE. Swedish Defence Research Agency, Stockholm. [Online]. Available: https://www.foi.se/rapportsammanfattning?reportNo=FOI-R--4552--SE

[2] Allen, G., Chan, T. (2017). Artificial intelligence and national security. Cambridge, MA: Harvard Kennedy School. [Online]. Available: https://www.belfercenter.org/publication/artificial-intelligence-and-national-security

[3] Pearl, J. (2009). *Causality*. New York: Cambridge University Press. [Online]. Available: https://www.cambridge.org/9781139632997

[4] Zhang, K., Schlpkopf, B., Spirtes, P., Glymor, C. (2018). Learning causality and causality-related learning: some recent progress, *National science review* **5**(1):26–29. doi:10.1093/nsr/nwx137

[5] Harnett, K. (2018, 15 May). To build truly intelligent machines, teach them cause and effect, *Quanta Mag.* [Online] Available: https://www.quantamagazine.org/to-build-truly-intelligent-machines-teach-them-cause-and-effect-20180515/

[6] Yin, J., Wang, J. (2014). A Dirichlet multinomial mixture model-based approach for short text clustering, in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge discovery and data mining* (KDD'14). New York: ACM, pp. 233–242. doi:10.1145/2623330.2623715

[7] Karasalo, M., Schubert, J. (2019). HSTOOL for Horizon Scanning of Scientific Literature. FOI-R--4760--SE, Swedish Defence Research Agency. [Online]. Available: https://www.foi.se/rapportsammanfattning?reportNo=FOI-R--4760--SE

[8] Karasalo, M., Schubert, J. (2019). Developing horizon scanning methods for the discovery of scientific trends, in *Proceedings of the 15th International Conference on Document Analysis and Recognition* (ICDAR 2019), Sydney, Australia, 20–25 September 2019. Piscataway, NJ: IEEE, pp. 1055–1062. doi:10.1109/ICDAR.2019.00172

[9] Waldmann, M. R., Mayrhofer, R. (2016). Hybrid causal representations, in *Psychology of Learning and Motivation*. Amsterdam: Elsevier (Vol. 65, Ch. 3). doi:10.1016/bs.plm.2016.04.001.

[10] Guo, R., Cheng, L., Li, J., Hahn, P.R., Liu, H. (2019). A survey of learning causality with data: problems and methods, under review.

[11] Pearl, J. (2009). Causal inference in statistics: An overview, *Statistics Surveys* 3:96–146. doi:10.1214/09-SS057

[12] Shimizu, S., Hoyer, P. O., Hyvarinen, A., Kerminen, A. (2006). A linear non-Gaussian acyclic model for causal discovery, *Journal of Machine Learning Research* **7**:2003–2030. [Online] Available: http://www.jmlr.org/papers/v7/shimizu06a.html

[13] Shimizu, S., Inazumi, T., Sogawa, Y., Hyvärinen, A., Kawahara, Y., Washio, T., Hoyer, P. O., Bollen, K. (2011). DirectLiNGAM: A direct method for learning a linear non-Gaussian structural equation model, *Journal of Machine Learning Research* **12**:1225–1248. [Online] Available: http://www.jmlr.org/papers/v12/shimizu11a.html

[14] Peters, J., Mooij, J. M., Janzing, D., Schölkopf, B. (2014). Causal discovery with continuous additive noise models, *Journal of Machine Learning Research* **15**:2009–2053. [Online] Available: http://www.jmlr.org/papers/v15/peters14a.html

[15] Heckerman, D. (1997). Bayesian networks for data mining. *Data mining and knowledge discovery* **1**(1):79–119. doi:10.1023/A:1009730122752

[16] Ellis, B., Wong, W. H. (2008). Learning causal Bayesian network structures from experimental data, *Journal of the American Statistical Association* **103**(482):778–789. doi:10.1198/016214508000000193

[17] Colombo, D., Maathuis, M. H., Kalisch, M., Richardson, T. S. (2012). Learning high-dimensional directed acyclic graphs with latent and selection variables, *The Annals of Statistics* **40**(1):294–321. [Online] Available: https://www.jstor.org/stable/41713636

[18] Kalisch, M., Mächler, M., Colombo, D., Maathuis, M. H., Bühlmann, P. (2012). Causal inference using graphical models with the R package pcalg, *Journal of Statistical Software* **47**(11):1–26. doi:10.18637/jss.v047.i11

[19] Isozaki, T. (2012). Learning causal Bayesian networks using minimum free energy principle, *New Generation Computing* **30**(1):17–52. doi:10.1007/s00354-012-0103-1

[20] Yu, K., Liu, L., Li, J., Chen, H. (2018). Mining Markov blankets without causal sufficiency, *IEEE Transactions on Neural Networks and Learning Systems* **29**(12):6333–6347, doi:10.1109/TNNLS.2018.2828982

[21] Stigler, A. L. (2015). Asessing causality in a complex security environment, *Joint Force Quarterly* **76**(1):35–39. [Online] Available: https://ndupress.ndu.edu/JFQ/Joint-Force-Quarterly-76/Article/577586/assessing-causality-in-a-complex-security-environment/

FOI
Defence Research Agency
SE-164 90 Stockholm

Phone: +46 8 555 030 00
Fax: +46 8 555 031 00

www.foi.se