Machine Learning and the Philosophy of Science: a Dynamic Interaction

Jon Williamson*

Draft of June 1, 2001

Abstract

I posit here a dynamic interaction between machine learning and the philosophy of science, and illustrate this claim with the use of a case study involving the foundations of Bayesian networks.

Contents

*Department of Philosophy, King's College, London. jon.williamson@kcl.ac.uk.

§1
Introduction

How is machine learning related to the philosophy of science? Answers to this question have typically focussed on the thought that machine learning is experimental philosophy of science — that we can use machine learning algorithms to test out ideas in the philosophy of science. I would like to argue here for a two-way relationship: the philosophy of science can guide machine learning as much as machine learning can help the philosophy of science.

More particularly I shall argue that there is a *dynamic interaction* between the philosophy of science and machine learning.[§2] I will give an overview of evidence for this relationship in §3, and examine a case study in closer detail in §4.

§2
Dynamic Interactions

Donald Gillies and Yuxin Zheng proposed the concept of a dynamic interaction between two disciplines in [Gillies & Zheng 1998] and characterised it as follows:

> (1) A connection is established between two different fields or subjects (A and B say).
>
> (2) Both sides may then benefit from this connection, or, more specifically, from the resulting interaction.
>
> (3) The relationship between the two sides is not static but dynamic. Suppose, for example, the flow of ideas is at first principally from A to B. It will then characteristically happen that, after a while, the direction of the flow of ideas is reversed and ideas from B start to influence A. Typically the interaction between A and B has two phases separated by a turning point, although, as is to be expected, the turning point is not precisely defined.
>
> (4) Although the two sides interact, it is important that each one should at the same time preserve some degree of autonomy. As our above analysis of the dynamic nature of their relationship in (3) shows, this is a necessary condition for further development.[1]

[1] [Gillies & Zheng 1998] §1.

In the same paper Gillies and Zheng provide two examples. They posit a dynamic interaction between the philosophy of mathematics and the philosophy of science: the Vienna Circle, who were crucially influenced by the philosophy of mathematics, invigorated the philosophy of science; later Lakatos sparked new life into an ailing philosophy of mathematics by reapplying his methodology of scientific research programmes. Their second example is a dynamic interaction between the philosophy of mathematics and computer science: the Turing machine, the λ-calculus, type theory and first order logic arose from problems in the foundations of mathematics and were in turn a strong influence on many aspects of computer science; automated theorem proving has led to a renewed interest in the notion of mathematical proof.[2]

A dynamic interaction is, by definition, a good thing for both sides. Such a symbiotic relationship revitalises the two disciplines and ensures their continued survival. An understanding and analysis of the concept of dynamic interaction is also important, for two reasons. The first is retrospective: the existence of dynamic interactions helps explain the success and longevity of a field. The second is prospective: the prospect of dynamic interactions can be used to justify inter-disciplinary research, inter-community discussion and cross-fertilisation of ideas.

§3
Overview of the Interaction

We can look both retrospectively and the prospectively when considering the relationship between machine learning and the philosophy of science. In this section I shall give an overview of some of the interactions between the two subjects to date, and some of the prospects for future interactions. In §4 we shall see in more detail how interactions can help both subjects by looking at a case study.

At the methodological level, the philosophy of science has had a pervasive influence on machine learning. The 1920s saw the establishment of the Vienna Circle study group and the beginning of the rise of logical positivism as a key movement in the philosophy of science. Integral to their approach was the use of logic as a framework for scientific reasoning. The movement was extremely important half way through the twentieth century when Turing put forward his vision of learning machines,[3] and the early machine learning systems of the 1960s were quick to put the learning problem into the logi-

[2]See [Gillies 2000] and [Barendregt 1997] for further insights into this interaction.
[3][Turing 1950] §7.

cal framework. By the end of the 1960s the positivists were in decline and an aversion to the logical approach to science had set in, in favour of the more sociological positions championed by Kuhn and Lakatos. Some scepticism towards the symbolic approach to AI followed,[4] and progress on the foundations of neural networks allowed an alternative line of development in machine learning. However it was not long before a resurgence of formal methods gripped the philosophy of science, in the shape of Bayesianism. Bayesian philosophy became important in AI in the 1980s and 1990s with the development of probabilistic expert systems and probabilistic frameworks for learning. Now Bayesian techniques for the machine learning of formal models are widespread, as are Bayesian models themselves.

Machine learning has influenced the controversy in the philosophy of science between inductivism and falsificationism: Gillies argues that the success of the GOLEM machine learning program at learning a scientific hypothesis provides evidence for inductivism.[5] Proponents of falsificationism remain sceptical about machine learning,[6] but further successes in the automation of scientific hypothesis generation may dampen their ardour. Machine learning has also proven invaluable to the philosophy of science as a means of testing formal models of scientific reasoning, including inductive reasoning,[7] abductive reasoning,[8] coherence-based reasoning,[9] analogical reasoning,[10] causal reasoning,[§4] and theory revision.[11]

There is a clear analogy between hypothesis choice in science and model selection in machines: this has allowed a transferral of ideas between the two disciplines and may the locus of future interactions. As mentioned above, the Bayesian approach to hypothesis confirmation has become very important in machine learning. Simplicity is an important issue in both disciplines, especially the question of whether simple hypotheses are more likely to be true than complicated ones, and the question of how one should balance the simplicity of a model with its degree of fit with evidence. [Bensusan 2000] expresses cautious optimism that future interactions will shed light on simplicity, as well as the role of theoretical terms and the viability of scientific realism.

---

[4] [Dreyfus 1972].
[5] [Gillies 1996].
[6] [Allen 2001].
[7] [Muggleton & de Raedt 1994], [Flach 1995].
[8] [Dimopoulos & Kakas 1996], [Poole 1998], [Flach & Kakas 2000].
[9] [Thagard 1988], [Thagard 1992], [Thagard 2000].
[10] [Hofstadter & Group 1995].
[11] [Darden 1998].

§4
Case study: Bayesian Networks

Bayesian networks have been the hotbed of a particularly productive interaction between machine learning and the philosophy of science, largely due to the fact that the two disciplines share a common interest in causality and probability.

Philosophers of science have been concerned with the relationship between causality and probabilistic independence and dependence since the 1950s.[12] By the 1980s there had been much exploration of the issues, but little consensus as to the exact relationship, and there was little hope that further philosophical consideration of the issue would lead to a resolution.

Then in the 1980s Judea Pearl took Bayesian philosophy of probability and the *causal Markov condition*, which says that a cause is probabilistically independent of any of its non-effects conditional on its direct causes, as a starting point for his theory of Bayesian networks.[13] A Bayesian network contains two components: a directed acyclic graph that represents the causal relations on a domain of variables, and probability tables in which the probability distribution of each variable conditional on its direct causes is specified. The causal Markov condition is assumed as a link between the causal graph and the specified probabilities, and then a Bayesian network determines a joint probability distribution over the variables.[14] Bayesian networks were originally intended as a formalism for causal reasoning in expert systems, but are now also used as a general framework for representing and learning probability functions. In fact, Bayesian networks were very rapidly adopted by the expert system and machine learning communities,[15] and the question naturally arose as to whether this is because the causal Markov condition is a plausible assumption: the success of Bayesian networks might be taken as evidence in favour of the causal Markov condition as a link between causality and probability.

If the causal Markov condition is a valid assumption, the question also arises as to whether the condition can be used to characterise the causal relation itself: a causal structure may in some sense be the smallest structure to satisfy the condition. Indeed machine learning techniques for learning Bayesian networks have been proposed as a way of automating the task of

[12] [Reichenbach 1956], [Suppes 1970], [Salmon 1980], [Spohn 1980].

[13] See [Pearl 1988] and also [Gillies 2001] for an account of this development.

[14] See [Pearl 1988] or [Neapolitan 1990] for the details.

[15] See for example the proceedings of the conferences on Uncertainty in Artificial Intelligence at www.auai.org.

causal discovery,[16] and the success of such techniques can be used to support a probabilistic analysis of causality based on the causal Markov condition. Such considerations have sparked intense philosophical debate on the nature of causality,[17] and it is clear that machine learning has invigorated issues in the philosophy of science.

But what has the philosophy of science to offer machine learning regarding these developments? I shall now outline some thoughts in this respect — these are more thoroughly documented in [Williamson 2000].

The causal Markov condition implies the *principle of the common cause*, which says that any two probabilistically dependent events, neither of which causes the other, must be effects of one or more common causes, and the two events must be probabilistically independent conditional on these common causes. Ever since Reichenbach's pioneering writings on the subject, the principle of the common cause has been central to the philosophical discussions of the relationship between causality and probability. Indeed the philosophical literature contains many counter-examples to the principle! If these counter-examples are to be taken seriously, then the causal Markov condition cannot hold in general — and in fact it turns out that the failures of the causal Markov condition are quite widespread, and that where causal and probabilistic knowledge is incomplete (as it often is in the application of Bayesian networks) the causal Markov condition is unlikely to hold.[18] Thus a Bayesian network is unlikely to represent the objective probability distribution over its domain.

So what explains the success of Bayesian networks? In my view their success may be attributed to the fact that, even though Bayesian networks are an imperfect representation of objective reality, they are the best models available given the limited knowledge at hand. This can be made precise as follows. Bayesian networks contain two items of knowledge: a causal graph and probability specification. Now according to the objective Bayesian interpretation of probability,[19] an agent should set her belief distribution over the variables in question to the probability function that maximises entropy given the constraints imposed by her background knowledge. But the probability function that maximises entropy given the knowledge in the causal graph and probability specification of a Bayesian network is precisely the prob-

[16] [Spirtes et al. 1993].

[17] See [Pearl 1988], [Spirtes et al. 1993], [McKim & Turner 1997], [Glymour & Cooper 1999], [Hausman 1999], [Hausman & Woodward 1999], [Korb 1999] and [Pearl 2000].

[18] [Williamson 2000].

[19] See [Jaynes 1998] [Corfield & Williamson 2001].

ability function determined by the Bayesian network itself.[20] Hence, even though the causal Markov condition may fail and a Bayesian network may not correctly represent the objective probability distribution on its variables, the network is the model one ought to adopt if one's background knowledge is represented by the components of the network. A Bayesian network is the best model given limited knowledge.

If this is the case, then one should adopt a two-stage methodology when employing Bayesian networks: first form a Bayesian network involving causal and probabilistic knowledge to hand (this is the best model from an objective Bayesian point of view); secondly when there is enough domain data refine this Bayesian network using machine learning techniques to better represent objective probability.[21] This allows one to integrate causal background knowledge with machine learning techniques in a unified formalism, and shows how philosophical considerations to do with the interpretations of probability can have a bearing on machine learning methodology.

In sum, by following the story of Bayesian networks we see how the philosophy of science can interact dynamically with machine learning. The Bayesian interpretation of probability inspired the formulation of Bayesian networks. Machine learning techniques were soon developed to learn Bayesian networks and these techniques allow the possibility of testing philosophical analyses of causality. Meanwhile considerations arising from the philosophy of probability motivate a new approach to learning Bayesian networks and integrating background knowledge.[22]

§5
Conclusion

The conditions of Gillies and Zheng for a dynamic interaction appear to have been met in the case of the philosophy of science and machine learning: several connections exist between the two disciplines to the benefit of each side, the relationship varies through time, and both sides remain autonomous.

By adding this interaction to those of §2 we obtain a chain of interactions from computer science to machine learning — as in Figure 1. Of course several further connections directly link computer science and machine learning, but they are not altogether autonomous disciplines so there is no direct dynamic interaction between them.

[20] [Williamson 2000].

[21] See [Williamson 2000] for a more detailed proposal here.

[22] For further applications of Bayesian networks to the philosophy of science see [Hartmann & Bovens 2000], [Bovens & Hartmann 2000] and [Bovens & Olsson 2000].

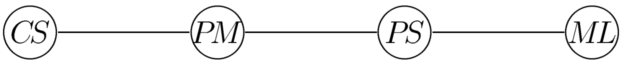

Figure 1: Dynamic interactions: computer science, the philosophy of mathematics, the philosophy of science and machine learning.

This new dynamic interaction, and the openness to external influences that accompanies it, may contribute to the success of both the philosophy of science and machine learning, and prospects for future interactions can be invoked to encourage proponents of both sides to look for further points of cross-over and pursue inter-disciplinary research.[23]

References

[Allen 2001] John F. Allen: 'Bioinformatics and discovery: induction beckons again', BioEssays 23, pages 104-107.

[Barendregt 1997] Henk Barendregt: 'The impact of the lambda calculus in logic and computer science', Bulletin of Symbolic Logic 3(2), pages 181-215.

[Bensusan 2000] Hilan Bensusan: 'Is machine learning experimental philosophy of science?', in ECAI2000 Workshop notes on Scientific Reasoning in Artificial Intelligence and the Philosophy of Science, Atocha Aliseda & David Pearce eds., pages 9-14.

[Boden 1990] Margaret A. Boden(ed.): 'The philosophy of artificial intelligence', Oxford University Press.

[Bovens & Hartmann 2000] Luc Bovens & Stephan Hartmann: 'Coherence, belief expansion and Bayesian networks', Proc. NMR 2000, Breckenridge, Colorado.

[Bovens & Olsson 2000] Luc Bovens & Erik J. Olsson: 'Coherentism, reliability and Bayesian networks', Mind 109, pages 685-719.

[Corfield & Williamson 2001] David Corfield & Jon Williamson(eds.): 'Foundations of Bayesianism', Kluwer Applied Logic Series, to appear.

[Darden 1998] Lindley Darden: 'Anomaly-driven theory redesign: computational philosophy of science experiments,' in T.W. Bynum and J.H.

[23] Thanks to the UK Arts and Humanities Research Board for funding this research.

Moor (eds.): ‘The Digital Phoenix: How Computers are Changing Philosophy’, New York: Blackwell Publishers, pages 62-78.

[Dimopoulos & Kakas 1996] Y. Dimopoulos & A. Kakas: ‘Abduction and Inductive Learning’, in L. De Raedt (ed.): ‘Advances in Inductive Logic Programming’, IOS Press, pages 144-171.

[Dreyfus 1972] Hubert L. Dreyfus: ‘What computers can’t do’, Cambridge, Massachusetts: The M.I.T. Press Press, revised edition ‘What computers still can’t do: a critique of artificial reason’ 1992.

[Flach 1995] Peter Flach: ‘An inquiry concerning the logic of induction’, ITK Dissertation Series.

[Flach & Kakas 2000] Peter A. Flach & Antonis C. Kakas: ‘On the relation between abduction and inductive learning’, in Dov M. Gabbay & Rudolf Kruse (eds.): ‘Handbook of defeasible reasoning and uncertainty management systems Vol. 4: Abductive reasoning and learning’, Kluwer Academic Publishers, pages 1–33.

[Gillies 1996] Donald Gillies: ‘Artificial intelligence and scientific method’, Oxford University Press.

[Gillies 2000] Donald Gillies: ‘Logicism and the development of computer science’, in [Kakas & Sadri 2000].

[Gillies 2001] Donald Gillies: ‘Probability in artificial intelligence’, in L. Floridi (ed.): ‘The Blackwell guide to the philosophy of computing and information’, Blackwell, to appear.

[Gillies & Zheng 1998] Donald Gillies & Yuxin Zheng: ‘Dynamic interactions with the philosophy of mathematics’, Theoria, to appear.

[Glymour & Cooper 1999] Clark Glymour & Gregory F. Cooper(eds.): ‘Computation, causation, and discovery’, Cambridge, Massachusetts: The M.I.T. Press.

[Hartmann & Bovens 2000] Stephan Hartmann & Luc Bovens: ‘The import of auxiliary theories of the instrument: a Bayesian-network approach’, in ECAI2000 Workshop notes on Scientific Reasoning in Artificial Intelligence and the Philosophy of Science, Atocha Aliseda & David Pearce eds., pages 21-27.

[Hausman 1999] Daniel M. Hausman: 'The mathematical theory of causation', review of [McKim & Turner 1997], British Journal for the Philosophy of Science 50, pages 151-162.

[Hausman & Woodward 1999] Daniel M. Hausman & James Woodward: 'Independence, invariance and the causal Markov condition', British Journal for the Philosophy of Science 50, pages 521-583.

[Hofstadter & Group 1995] Douglas Hofstadter & Fluid Analogies Research Group: 'Fluid concepts and creative analogies: computer models of the fundamental mechanisms of thought', Penguin, 1998.

[Jaynes 1998] E.T. Jaynes: 'Probability theory: the logic of science', http://bayes.wustl.edu/etj/prob.html.

[Kakas & Sadri 2000] A. Kakas & F. Sadri: 'Essays in honour of Robert Kowalski', to appear.

[Korb 1999] Kevin B. Korb: 'Probabilistic causal structure', in H. Sankey (ed.): 'Causation and laws of nature', Kluwer, pages 265-311.

[McKim & Turner 1997] Vaughn R. McKim & Stephen Turner: 'Causality in crisis? Statistical methods and the search for causal knowledge in the social sciences', University of Notre Dame Press.

[Muggleton & de Raedt 1994] Stephen Muggleton & Luc de Raedt: 'Inductive logic programming: theory and methods', in Journal of Logic Programming 12, pages 1-51.

[Neapolitan 1990] Richard E. Neapolitan: 'Probabilistic reasoning in expert systems: theory and algorithms', New York: Wiley.

[Pearl 1988] Judea Pearl: 'Probabilistic reasoning in intelligent systems: networks of plausible inference', San Mateo, California: Morgan Kaufmann.

[Pearl 2000] Judea Pearl: 'Causality: models, reasoning, and inference', Cambridge University Press.

[Poole 1998] David Poole: 'Learning, Bayesian Probability, Graphical Models, and Abduction', in Peter Flach & Antonis Kakas (eds.): 'Abduction and Induction: essays on their relation and integration', Kluwer.

[Reichenbach 1956] Hans Reichenbach: 'The direction of time', Berkeley & Los Angeles, University of California Press, reprinted 1971.

[Salmon 1980] Wesley C. Salmon: 'Probabilistic causality', in [Salmon 1998], pages 208-232.

[Salmon 1998] Wesley C. Salmon: 'Cauality and explanation', Oxford: Oxford University Press.

[Spirtes et al. 1993] Peter Spirtes, Clark Glymour & and Richard Scheines: 'Causation, Prediction, and Search', Lecture Notes in Statistics 81, Springer-Verlag.

[Spohn 1980] Wolfgang Spohn: 'Stochastic independence, causal independence, and shieldability', Journal of Philosophical Logic 9, pages 73-99.

[Suppes 1970] Patrick Suppes: 'A probabilistic theory of causality', Amsterdam: North-Holland.

[Thagard 1988] Paul Thagard: 'Computational Philosophy of Science', Cambridge, MA: Cambridge, Massachusetts: The M.I.T. Press Press/Bradford Books.

[Thagard 1992] Paul Thagard: 'Conceptual Revolutions', Princeton: Princeton University Press.

[Thagard 2000] Paul Thagard: 'Probabilistic networks and explanatory coherence', Cognitive Science Quarterly 1, pages 91-114.

[Turing 1950] Alan M. Turing: 'Computing machinery and intelligence', in [Boden 1990], pages 40-66.

[Williamson 2000] Jon Williamson: 'Foundations for Bayesian networks', in [Corfield & Williamson 2001].

Index