The Philosophy of Science and its relation to Machine Learning

Jon Williamson

In this chapter I discuss connections between machine learning and the philosophy of science. First I consider the relationship between the two disciplines. There is a clear analogy between hypothesis choice in science and model selection in machine learning. While this analogy has been invoked to argue that the two disciplines are essentially doing the same thing and should merge, I maintain that the disciplines are distinct but related and that there is a dynamic interaction operating between the two: a series of mutually beneficial interactions that changes over time. I will introduce some particularly fruitful interactions, in particular the consequences of automated scientific discovery for the debate on inductivism versus falsificationism in the philosophy of science, and the importance of philosophical work on Bayesian epistemology and causality for contemporary machine learning. I will close by suggesting the locus of a possible future interaction: evidence integration.

1 Introduction

Since its genesis in the mid 1990s, data mining has been thought of as encompassing two tasks: using data to test some pre-determined hypothesis, or using data to determine the hypothesis in the first place. The full automation of both these tasks – hypothesising and then testing – leads to what is known as *automated discovery* or *machine learning*. When such methods are applied to science, we have what is called *automated scientific discovery* or *scientific machine learning*. In this chapter, we shall consider the relationship between the philosophy of science and machine learning, keeping automated scientific discovery particularly in mind.

Section 2 offers a brief introduction to the philosophy of science. In Sect. 3 it is suggested that the philosophy of science and machine learning admit mutually fruitful interactions because of an analogy between hypothesis choice in the philosophy

J. Williamson (✉)
University of Kent, Canterbury, UK
e-mail: j.williamson@kent.ac.uk

M.M. Gaber (ed.), *Scientific Data Mining and Knowledge Discovery: Principles and Foundations*, DOI 10.1007/978-3-642-02788-8_4,

of science and model selection in machine learning. An example of the benefit of machine learning for the philosophy of science is provided by the importance of work on automated scientific discovery for the debate between inductivism and falsificationism in the philosophy of science (Sect. 4). On the other hand, the influence of philosophical work on Bayesianism and causality provides an example of the benefits of the philosophy of science for machine learning (Sect. 5). Section 6 hypothesises that evidence integration may become the locus of the further fruitful interaction between the two fields.

2 What is the Philosophy of Science?

In the quest to improve our understanding of science, three fields of enquiry stand out: history of science, sociology of science, and philosophy of science. Historians of science study the development of science, key scientists and key ideas. Sociologists of science study social constraints on scientific activity–e.g., how power struggles impact on the progress of science. Philosophers of science study the concepts of science and normative constraints on scientific activity. Questions of interest to philosophers of science include:

Demarcation: What demarcates science from non-science? One view is that empirical testability is a necessary condition for a theory to count as scientific.

Unity: To what extent is science a unified or unifiable field of enquiry? Some take physics to be fundamental and the elements of other sciences to be reducible to those of physics. Others argue that science is a hotch-potch of rather unrelated theories, or that high-level complexity is not reducible to low-level entities and their arrangement.

Realism: Are the claims of science true? To what extent are we justified in believing contemporary scientific theories? Realists hold that scientific theories aim to describe an independent reality and that science gradually gets better at describing that reality. On the other hand, instrumentalists hold that science is an instrument for making predictions and technological advances and that there is no reason to take its claims literally, or – if they are taken at face value – there are no grounds for believing them.

Explanation: What is it to give a scientific explanation of some phenomena? One view is that explaining is the act of pointing to the physical mechanism that is responsible for the phenomena. Another is that explanation is subsumption under some kind of regularity or law.

Confirmation: How does evidence confirm a scientific theory? Some hold that evidence confirms a hypothesis just when it raises the probability of the hypothesis. Others take confirmation to be a more complicated relation, or not a binary relation at all but rather to do with coherence with other beliefs.

Scientific Method: How are the goals of science achieved? What is the best way of discovering causal relationships? Can one justify induction? While many maintain that in principle one can automate science, others hold that scientific discovery is an essentially human, intuitive activity.

Concepts of the Sciences: How should one interpret the probabilities of quantum mechanics? Does natural selection operate at the level of the individual, the population or the gene? Each science has its particular conceptual questions; even the interpretation of many general concepts – such as probability and causality – remains unresolved.

3 Hypothesis Choice and Model Selection

There is a clear link between the philosophy of science on the one hand and the area of machine learning and data mining on the other. This link is based around an analogy between *hypothesis choice* in science and *model selection* in machine learning. The task of determining a scientific hypothesis on the basis of current evidence is much like the task of determining a model on the basis of given data. Moreover, the task of evaluating the resulting scientific hypothesis is much like the task of evaluating the chosen model. Finally, the task of deciding which evidence to collect next (which experiments and observations to perform) seems to be similar across science and machine learning. Apparently, then, scientific theorising and computational modeling are but two applications of a more general form of reasoning.

How is this general form of reasoning best characterised? It is sometimes called *abductive inference* or *abduction*, a notion introduced by C.S. Peirce. But this nomenclature is a mistake: the form of reasoning alluded to here is more general than abduction. Abduction is the particular logic of moving from observed phenomena to an explanation of those phenomena. Science and machine learning are interested in the broader, iterative process of moving from evidence to theory to new evidence to new theory and so on. (This broader process was sometimes called "induction" by Peirce, though "induction" is normally used instead to refer to the process of moving from the observation of a feature holding in each member of a sample to the conclusion that the feature holds of unobserved members of the population from which the sample is drawn.) Moreover, explanation is just one use of hypotheses in science and models in machine learning; hypotheses and models are also used for other forms of inference such as prediction. In fact, while explanation is often the principal target in science, machine learning tends to be more interested in prediction. When explanation is the focus, one is *theorising*; when prediction is the focus, the process is better described as *modeling*. Clearly, then, the general form of reasoning encompasses both theorising and modeling. This general form of reasoning is sometimes called *discovery*. But that is not right either: the form of reasoning under consideration here is narrower than discovery. "Discovery" applies to finding out new particular facts as well as to new generalities, but we are interested purely in generalities here. In want of a better name, we shall call this general form of reasoning *systematising*, and take it to encompass theorising and modeling.

Granting, then, that hypothesis choice in science and model selection in machine learning are two kinds of systematising, there are a variety of possible views as to the relationship between the philosophy of science and machine learning.

One might think that since the philosophy of science and machine learning are both concerned with systematising, they are essentially the same discipline, and hence some kind of merger seems sensible [1]. This position is problematic, though. As we saw in Sect. 2, the philosophy of science is not only concerned with the study of systematising, but also with a variety of other topics. Hence the philosophy of science can at best be said to intersect with machine learning. Moreover, even where they intersect the aims of the two fields are rather different: e.g., the philosophy of science is primarily interested in explanation and hence theorising, while the area of machine learning and data mining is primarily interested in prediction and hence modeling. Perhaps automated scientific discovery is one area where the aims of machine learning and the philosophy of science coincide. In which case automated scientific discovery is the locus of intersection between the philosophy of science and machine learning. But this rather narrow intersection falls far short of the claim that the two disciplines are the same.

More plausibly, then, the philosophy of science and machine learning are not essentially one, but nevertheless they do admit interesting connections [2, 189; 3, 4]. In [4], I argue that the two fields admit a *dynamic interaction*. There is a dynamic interaction between two fields if there is a connection between them which leads to a mutually beneficial exchange of ideas, the direction of transfer of ideas between the two fields changes over time, and the fields remain autonomous [5]. Here, we shall take a look at two beneficial points of interaction: the lessons of automated scientific discovery for the study of scientific method (Sect. 4) and the influence of work on Bayesian epistemology and probabilistic causality on machine learning (Sect. 5).

4 Inductivism Versus Falsificationism

Scientific method is an important topic in the philosophy of science. How do scientists make discoveries? How *should* they make discoveries?

One view, commonly called *inductivism* and advocated by Bacon [6], is that science should proceed by first making a large number of observations and then extracting laws via a procedure that is in principle open to automation. An opposing position, called *falsificationism* and held by popper [7], is that the scientist first conjectures in a way that cannot be automated and then tests this conjecture by observing and experimenting to see whether or not the predictions of the conjecture are borne out, rejecting the conjecture if not.

While examples from the history of science have tended to support falsificationism over inductivism, the successes of automated scientific discovery suggest that inductivism remains a plausible position [8]. The approach of machine learning is to collect large numbers of observations in a dataset and then to automatically extract a predictive model from this dataset. In automated scientific discovery this model is usually also meant to be explanatory; hence, the model plays the same role as scientific laws. To the extent that such procedures are successful, inductivism is successful. Gillies [8], Sect. 2.6 cites the GOLEM inductive logic programming system as an example of a machine learning procedure that successfully induced scientific laws concerning protein folding; this success was achieved with the help of humans who encoded background knowledge [8, Sect. 3.4]. Journals such as Data Mining and Knowledge Discovery and the Journal of Computer-Aided Molecular Design show that the inductive approach continues to produce advances. Moreover, the investment of drug and agrochemical companies suggests that this line of research promises to pay dividends. While the hope is that one day such companies might "close the inductive loop" – i.e., automate the whole cyclic procedure of data collection, hypothesis generation, further data collection, hypothesis reformulation ... – the present reality is that machine successes are achieved in combination with human expertise. The use by Dow AgroSciences of neural networks in the development of the insecticide spinetoram offers a recent example of successful human–machine collaboration [9]; spinetoram won the US Environmental Protection Agency 2008 Designing Greener Chemicals Award.

Perhaps human scientists proceed by applying falsificationism while machine science is inductivist. Or perhaps falsificationism and inductivism are but different approximations to a third view which better explicates scientific method. What could this third view be? It is clear that human scientists base their conjectures on a wide variety of different kinds of evidence, not just on a large number of homogeneous observations. (This – together with the fact that it can be hard for scientists to pin-point all the sources of evidence for their claims and hard for them to say exactly how their evidence informs their hypotheses – makes it hard to see how hypothesis generation can be automated. It is natural to infer, with falsificationists, that hypothesis generation can't be automated, but such an inference may be too quick.) On the other hand, most machine learning algorithms do take as input a large number of homogenous observations; this supports inductivism, but with the proviso that the successes of automated scientific discovery tend to be achieved in concert with human scientists or knowledge engineers. Human input appears to be important to fully utilise the range of evidence that is available. The third view of scientific method, then, is that a theory is formulated on the basis of extensive evidence (including background knowledge), but evidence which is often qualitative and hard to elucidate. The theory is revised as new evidence is accrued, and this new evidence tends to be accrued in a targeted way, by testing the current theory. Can this third way be automated? Contemporary machine learning methods require well-articulated quantitative evidence in the form of a dataset, but, as I suggest in Sect. 6, there is scope for relaxing this requirement and taking a fuller spectrum of evidence into account.

In sum, while not everyone is convinced by the renaissance of inductivism [see, e.g., 10], it is clear that automated scientific discovery has yielded some successes and that these have sparked new life into the debate between inductivism and falsificationism in the philosophy of science.

5 Bayesian Epistemology and Causality

Having looked at one way in which machine learning has had a beneficial impact on the philosophy of science, we now turn to the other direction: the impact of the philosophy of science on machine learning.

Epistemologists are interested in a variety of questions concerning our attitudes towards the truth of propositions. Some of these questions concern propositions that we already grant or endorse: e.g., do we know that $2 + 2 = 4$? if so, why? how should a committee aggregate the judgements of its individuals? Other questions concern propositions that are somewhat speculative: should I accept that all politicians are liars? to what extent should you believe that it will rain tomorrow?

Philosophers of science have been particularly interested in the latter question: to what extent should one believe a proposition that is open to speculation? This question is clearly relevant to scientific theorising, where we are interested in the extent to which we should believe current scientific theories. In the twentieth century philosophers of science developed and applied *Bayesian epistemology* to scientific theorising. The ideas behind Bayesian epistemology are present in the writings of some of the pioneers of probability theory – e.g., Jacob Bernoulli, Thomas Bayes – but only in recent years has it widely caught on in philosophy and the sciences.

One can characterise contemporary Bayesian epistemology around the norms that it posits:

Probability: The strengths of an agent's beliefs should be representable by probabilities. For example, the strength to which you believe it will rain tomorrow should be measurable by a number $P(r)$ between 0 and 1 inclusive, and $P(r)$ should equal $1 - P(\neg r)$, where $\neg r$ is the proposition that it will not rain tomorrow.

Calibration: These degrees of belief should be calibrated with the agent's evidence. For example, if the agent knows just that between 60% and 70% of days like today have been followed by rain, she should believe it will rain tomorrow to degree within the interval $[0.6, 0.7]$.

Equivocation: Degrees of belief should otherwise be as equivocal as possible. In the above example, the agent should equivocate as far as possible between r and $\neg r$, setting $P(r) = 0.6$, the value in the interval [0.6, 0.7] that is closest to total equivocation, $P_{=}(r) = 0.5$.

So-called *subjective Bayesianism* adopts the Probability norm and usually the Calibration norm too. This yields a relatively weak prescription, where the extent to which one should believe a proposition is largely left up to subjective choice. To limit the scope for arbitrary shifts in degrees of belief, subjectivists often invoke a further norm governing the updating of degrees of belief: the most common such norm is *Bayesian conditionalisation*, which says that the agent's new degree of belief $P'(a)$ in proposition a should be set to her old degree of belief in a conditional on the new evidence e, $P'(a) = P(a|e)$.

In contrast to subjective Bayesianism, *Objective Bayesianism* adopts all three of the earlier norms – Probability, Calibration and Equivocation. If there are finitely many basic propositions under consideration (propositions that are not composed out of simpler propositions), these three norms are usually cashed out using the *maximum entropy principle*: an agent's degrees of belief should be representable by a probability function, from all those calibrated with evidence, that has maximum entropy $H(P) = -\sum_{\omega \in \Omega} P(\omega) \log P(\omega)$. (The maximum entropy probability function is the function that is closest to the maximally equivocal probability function $P_=$ which gives the same probability $P_=(\omega) = 1/2^n$ to each conjunction $\omega \in \Omega = \{\pm a_1 \wedge \cdots \wedge \pm a_n\}$ of the basic propositions $a_1, \ldots, a_n$ or their negations, where distance from one probability function to another is understood in terms of cross entropy $d(P, Q) = \sum_{\omega \in \Omega} P(\omega) \log P(\omega)/Q(\omega)$.) Since these three norms impose rather strong constraints on degrees of belief, no further norm for updating need be invoked [11]. The justification of these norms and the relative merits of subjective and objective Bayesian epistemology are topics of some debate in philosophy [see, e.g., 12].

The development of Bayesian epistemology has had a profound impact on machine learning [13]. The field of machine learning arose out of research on expert systems. It was quickly realised that when developing an expert system it is important to model the various uncertainties that arise on account of incomplete or inconclusive data. Thus the system MYCIN, which was developed in the 1970s to diagnose bacterial infections, incorporated numerical values called "certainty factors." Certainty factors were used to measure the extent to which one ought to believe certain propositions. Hence, one might think that Bayesian epistemology should be applied here. In fact, the MYCIN procedure for handling certainty factors was non-Bayesian, and MYCIN was criticised on account of its failing to follow the norms of Bayesian epistemology. From the late 1970s, it was common to handle uncertainty in expert systems using Bayesian methods. And, when the knowledge bases of expert systems began to be learned automatically from data rather than elicited from experts, Bayesian methods were adopted in the machine learning community.

But Bayesian methods were rather computationally intractable in the late 1970s and early 1980s, and consequently systems such as Prospector, which was developed in the second half of the 1970s for mineral prospecting, had to make certain simplifying assumptions that were themselves questionable. Indeed considerations of computational complexity were probably the single biggest limiting factor for the application of Bayesian epistemology to expert systems and machine learning.

It took a rather different stream of research to unleash the potential of Bayesian epistemology in machine learning. This was research on causality and causal reasoning. In the early twentieth century, largely under the sceptical influence of Mach, Pearson and Russell, causal talk rather fell out of favour in the sciences. But it was clear that while scientists were reluctant to talk the talk, they were still very much walking the walk: associations between variables were being interpreted causally to predict the effects of interventions and to inform policy. Consequently, philosophers of science remained interested in questions about the nature of causality and how one might best reason causally.

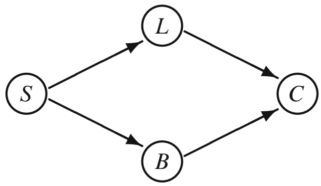

Fig. 1 Smoking (S) causes lung cancer (L) and bronchitis (B) which in turn cause chest pains (C)

Under the *probabilistic* view of causality, causal relationships are analysable in terms of probabilistic relationships – more specifically in terms of patterns of probabilistic dependence and independence [14]. Reichenbach, Good, Suppes and Pearl, pioneers of the probabilistic approach, developed the concept of a *causal net*. A causal net is a diagrammatic representation of causes and effects – such as that depicted in Fig. 1 – which has probabilistic consequences via what is now known as the *Causal Markov Condition*. This condition says that each variable in the net is probabilistically independent of its non-effects, conditional on its direct causes. If we complete a causal net by adding the probability distribution of each variable conditional on its direct causes, the net suffices to determine the joint probability distribution over all the variables in the net. Since the probabilities in the net tend to be interpreted as rational degrees of belief, and since the probabilities are often updated by Bayesian conditionalisation, a causal net is often called a *causal Bayesian net*. If we drop the causal interpretation of the arrows in the graph, we have what is known as a *Bayesian net*. The advantage of the causal interpretation is that under this interpretation the Markov Condition appears quite plausible, at least as a default constraint on degrees of belief [15].

Now, a causal Bayesian net – and more generally a Bayesian net – can permit tractable handling of Bayesian probabilities. Depending on the sparsity of the graph, it can be computationally feasible to represent and reason with Bayesian probabilities even where there are very many variables under consideration. This fact completed the Bayesian breakthrough in expert systems and machine learning. By building an expert system around a causal net, efficient representation and calculation of degrees of belief were typically achievable. From the machine learning perspective, if the space of models under consideration is the space of Bayesian nets of sufficiently sparse structure, then learning a model will permit efficient inference of appropriate degrees of belief. If the net is interpreted causally, we have what might be considered the holy grail of science: a method for the machine learning of causal relationships directly from data.

In sum, Bayesian epistemology offered a principled way of handling uncertainty in expert systems and machine learning, and Bayesian net methods overcame many of the ensuing computational hurdles. These lines of work had a huge impact: the dominance of Bayesian methods – and Bayesian net methods in particular – in the annual conferences on Uncertainty in Artificial Intelligence (UAI) from the 1980s is testament to the pervasive influence of Bayesian epistemology and work on probabilistic causality.

6 Evidence Integration

We now have some grounds for the claim that there is a dynamic interaction between machine learning and the philosophy of science: the achievements of automated scientific discovery have reinvigorated the debate between inductivists and falsificationists in the philosophy of science; on the other hand, work on Bayesian epistemology and causality has given impetus to the handling of uncertainty in machine learning. No doubt the mutually supportive relationships between philosophy of science and machine learning will continue. Here, we will briefly consider one potential point of interaction, namely the task of *evidence integration*.

The dominant paradigm in machine learning and data mining views the machine learning problem thus: given a dataset learn a (predictively accurate) model that fits (but does not overfit) the data. Clearly, this is an important problem and progress made on this problem has led to enormous practical advances. However, this problem formulation is rather over-simplistic in the increasingly evidence-rich environment of our information age. Typically, our evidence is *not* made up of a single dataset. Typically, we have a variety of datasets – of varying size and quality and with perhaps few variables in common – as well as a range of qualitative evidence concerning measured and unmeasured variables of interest – evidence of causal, logical, hierarchical and mereological relationships for instance. The earlier problem formulation just doesn't apply when our evidence is so multifarious. The Principle of Total Evidence, which holds that one should base one's beliefs and judgements on all one's available evidence, is a sound epistemological precept, and one that is breached by the dominant paradigm.

The limitations of the earlier problem formulation are increasingly becoming recognised in the machine-learning community. This recognition has led to a spate of research on what might be called *forecast aggregation*: a variety of models, each derived from a different dataset, are used to make predictions, and these separate predictions are somehow aggregated to yield an overall prediction. The aggregation operation may involve simple averaging or more sophisticated statistical meta-analysis methods.

But forecast aggregation itself has several limitations. First, it still falls foul of the Principle of Total Evidence: each model is based on a single dataset but qualitative evidence tends to be ignored. (Not always: qualitative evidence about relationships between the variables is sometimes invoked in the preprocessing of the datasets – if the knowledge engineer sees that a variable in one dataset is a subcategory of a variable in another dataset, these two variables might be unified in some way. But this data grooming is typically done by hand. As datasets involve more and more variables it is increasingly important that qualitative evidence be respected *as a part of the automation process*.) Second, it is unclear how far one should trust aggregated forecasts when they are often generated by models that not only disagree but are based on mutually inconsistent assumptions. Obviously in such cases at most one of the mutually inconsistent models is true; surely it would be better to find and use the true model (if any) than to dilute its predictions with the forecasts of false models. But this requires collecting new evidence rather than forecast aggregation. Third, the general problem of judgement aggregation – of which forecast aggregation is but a special case – is fraught with conceptual problems; indeed the literature on judgement aggregation is replete with impossibility results, not with solutions [16].

In view of these problems, a better approach might be to construct a single model which is based on the entirety of the available evidence – quantitative and qualitative – and to use that model for predictions. Combining evidence is often called *knowledge integration*. However, available evidence may not strictly qualify as knowledge because it may, for reasons that are not evident, not all be true; hence *evidence integration* is better terminology.

Bayesian epistemology provides a very good way of creating a single model on the basis of a wide variety of evidence. As discussed in Sect. 5, the model in this case is (a representation of) the probability function that captures degrees of belief that are appropriate for an agent with the evidence in question. The evidence is integrated via the Calibration norm: each item of evidence imposes constraints that this probability function must satisfy. (Some kind of consistency maintenance procedure must of course be invoked if the evidence itself is inconsistent.) A dataset imposes the following kind of constraint: the agent's probability function, when restricted to the variables of that dataset, should match (fit but not overfit) the distribution of the dataset, as far as other evidence permits. Qualitative evidence imposes another kind of equality constraint, as follows. A relation R is an *influence relation* if learning of a new variable that does not stand in relation R to (i.e., does not influence) the current variables does not provide grounds for changing one's degrees of belief concerning the current variables. Arguably causal, logical, hierarchical and mereological relationships are influence relations. Hence, evidence of such relationships imposes equality constraints of the form: the agent's probability function, when restricted to variables that are closed under influence, should match the probability function that the agent would have adopted were she only to have had evidence concerning that subset of variables, as far as other evidence permits. Hence, both quantitative and qualitative evidence impose certain equality constraints on degrees of belief. See [15] and [17] for the details and motivation behind this kind of approach.

In sum, evidence integration has greater potential than forecast aggregation to circumvent the limited applicability of the current machine learning paradigm. Bayesian epistemology is strikingly well-suited to the problem of evidence integration. Hence, there is scope for another fruitful interaction between philosophy of science and machine learning.

Example: Cancer Prognosis

As an illustration of the kind of approach to evidence integration that Bayesian epistemology offers, we shall consider an application of objective Bayesian epistemology to integrating evidence for breast cancer prognosis. This application is described in detail in [18].

When a patient has breast cancer and has had surgery to remove the cancer it is incumbent on the relevant medical practitioners to make an appropriate onward treatment decision. Broadly speaking, more effective treatments are more aggressive in the sense that they have harsher side effects. Such treatments are only warranted to the extent that the cancer is likely to recur without them. The more strongly the medical practitioner believes the cancer will recur, the more aggressive the treatment that will be instigated. It is important, then, that the agent's degree of belief in recurrence is appropriate given the available evidence.

Evidence here – as in many realistic applications – is multifarious. There are clinical datasets detailing the clinical symptoms of past patients, genomic datasets listing the presence of various molecular markers in past patients, scientific papers supporting particular causal claims or associations, medical experts' causal knowledge, and information in medical informatics systems, including medical ontologies, and previous decision support systems such as argumentation systems.

In [18] we had the following sources of evidence available. First, we had a clinical dataset, namely the SEER study, which involves three million patients in the US from 1975–2003, including 4,731 breast cancer patients. We also had two molecular datasets, one with 502 cases and another with 119 cases; the latter dataset also measured some clinical variables. Finally, we had a published study which established a causal relationship between two variables of interest.

These evidence sources impose constraints on an agent's degrees of belief, as outlined earlier. The agent's degrees of belief should match the dataset distributions on their respective domains. Moreover, degrees of belief should respect the equality constraints imposed by knowledge of causal influence. While the Probability norm holds that the strengths of the agent's beliefs should be representable by a probability function, the Calibration norm holds that this probability function should satisfy the constraints imposed by evidence.

These two norms narrow down the choice of belief function to a set of probability functions. But objective Bayesian epistemology imposes a further norm, Equivocation. Accordingly, the agent's degrees of belief should be representable by a probability function from within this set that is maximally equivocal. On a finite domain, this turns out to be the (unique) probability function in this set that has maximum entropy. So, objective Bayesian epistemology recommends that treatment decisions be based on this maximum entropy probability function.

As discussed in Sect. 5, from a computational point of view it is natural to represent a probability function by a Bayesian net. A Bayesian net that represents a probability function that is deemed appropriate by objective Bayesian epistemology is called an *objective Bayesian net* [19]. This Bayesian net can be used for inference – in our case to calculate degree to which one ought to believe that the patient's cancer will recur. Figure 2 depicts the graph of the objective Bayesian net in our cancer application. At the top is the recurrence node, beneath which are clinical variables. These are connected to five molecular variables at the bottom of the graph. Hence, one can use both molecular markers and clinical symptoms to predict

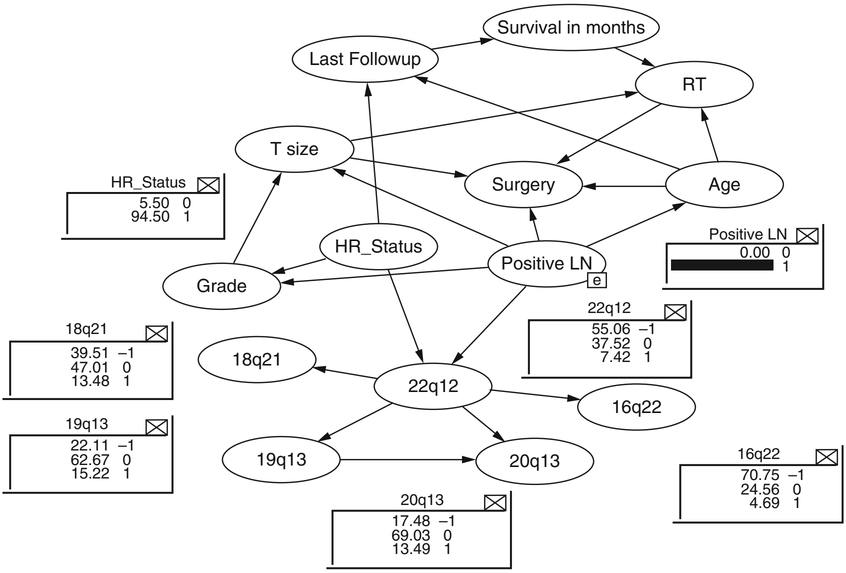

Fig. 2 Graph of an objective Bayesian net for breast cancer prognosis

the patient's survival, even though no dataset contains information about all these variables together. The objective Bayesian net model succeeds in integrating rather disparate evidence sources.

7 Conclusion

Machine learning in general and automated scientific discovery in particular have a close relationship with the philosophy of science. On the one hand, advances in automated scientific discovery have lent plausibility to inductivist philosophy of science. On the other hand, advances in probabilistic epistemology and work on causality have improved the ability of machine learning methods to handle uncertainty.

I have suggested that inductivism and falsificationism can be reconciled by viewing these positions as approximations to a third view of scientific method, one which considers the full range of evidence for a scientific hypothesis. This third way would be an intriguing avenue of research for philosophy of science. I have also suggested that the current single-dataset paradigm in machine learning is becoming increasingly inapplicable, and that research in machine learning would benefit from serious consideration of the problem of formulating a model on the basis of a broad range of evidence. Bayesian epistemology may be a fruitful avenue of research for tackling evidence integration in machine learning as well as evidence integration in the philosophy of science.

Acknowledgements This research was supported by the Leverhulme Trust. I am very grateful to Drake Eggleston and Peter Goodfellow for helpful leads.

References

1. K.B. Korb, Machine Learning as Philosophy of Science, ed. by K. Korb, H. Bensusan. in *Proceedings of the ECML-PKDD-01 Workshop on Machine Learning as Experimental Philosophy of Science*, 2001
2. P. Thagard, *Computational Philosophy of Science* (MIT, Cambridge, MA, 1988)
3. H. Bensusan, in *Is Machine Learning Experimental Philosophy of Science?* ed. by A. Aliseda, D. Pearce. ECAI2000 Workshop Notes on Scientific Reasoning in Artificial Intelligence and the Philosophy of Science (ECAI, Berlin, 2000), pp. 9–14
4. J. Williamson, Mind Mach. **14**(4), 539–549 (2004)
5. D. Gillies, Y. Zheng, Theoria **16**(3), 437–459 (2001)
6. F. Bacon, *The New Organon*, ed. by L. Jardine, M. Silverthorne (Cambridge University Press, Cambridge, 2000)
7. K.R. Popper, *Conjectures and Refutations* (Routledge and Kegan Paul, 1963)
8. D. Gillies, *Artificial Intelligence and Scientific Method* (Oxford University Press, Oxford, 1996)
9. T.C. Sparks, G.D. Crouse, J.E. Dripps, P. Anzeveno, J. Martynow, C.V. DeAmicis, J. Gifford, J. Comp. Aided Mol. Des. **22**, 393–401 (2008)
10. J.F. Allen, BioEssays **23**, 104–107 (2001)
11. J. Williamson, Synthese (2009), doi 10.1007/s11229-009-9515-y
12. J. Williamson, in *Motivating Objective Bayesianism: From Empirical Constraints to Objective Probabilities*, ed. by W.L. Harper, G.R. Wheeler. Probability and Inference: Essays in Honour of Henry E. Kyburg Jr. (College Publications, London, 2007), pp. 151–179
13. D. Gillies, in *Handling Uncertainty in Artificial Intelligence, and the Bayesian Controversy*, ed. by F. Stadler. Induction and Deduction in the Sciences (Kluwer, Dordrecht, 2003)
14. J. Williamson, *Probabilistic Theories of Causality*, ed. by H. Beebee, C. Hitchcock, P. Menzies. The Oxford Handbook of Causation (Oxford University Press, Oxford, 2008d)
15. J. Williamson, *Bayesian Nets and Causality: Philosophical and Computational Foundations* (Oxford University Press, Oxford, 2005a)
16. J. Williamson, J. Logic Comput. **19**, 461–473 (2009)
17. J. Williamson, in A. Carsetti (ed.), Causality, meaningful complexity and knowledge construction, Springer 2010
18. S. Nagl, M. Williams, J. Williamson, in *Objective Bayesian Nets for Systems Modeling and Prognosis in Breast Cancer*, ed. by D. Holmes, L. Jain. Innovations in Bayesian Networks: Theory and Applications (Springer, Berlin, 2008)
19. J. Williamson, in *Objective Bayesian Nets*, ed. by S. Artemov, H. Barringer, A.S. d'Avila Garcez, L.C. Lamb, J. Woods. We Will Show Them! Essays in Honour of Dov Gabbay, vol. 2 (College Publications, London, 2005b), pp. 713–730