Boosting the Correspondence Between Abduction and Learning

Helmut Prendinger and Mitsuru Ishizuka
Department of Information and Communication Engineering
School of Engineering, University of Tokyo, Japan
e-mail: {helmut,ishizuka}@miv.t.u-tokyo.ac.jp

The aim of this contribution is to strengthen the ties between a probabilistic (or cost-based) variant of logic-based abduction and the Bayesian approach to learning. Taking on an abductive perspective, we show that a variety of scientific reasoning tasks are covered by the integration of abduction and learning.

Many kinds of cognitive and scientific activity fall under one of the following categories: *evidential* reasoning, i.e., inferring causes from effects, or *causal* reasoning, i.e., predicting the effects given some cause. Independently of the type of reasoning, the domain theory T is often modeled causally, with rules "$e \leftarrow c$" denoting cause-effect relationships. As an example of the interaction between evidential and causal reasoning, consider a medical doctor who determines possible diseases based on observations, and then makes predictions of other symptoms. This scheme of reasoning can be easily represented in probabilistic logic-based abduction and Bayesian networks. Since the work of Popper, it is also a well-known pattern in scientific reasoning.

Abduction, originally invented by Ch. S. Peirce, uses a causal model in the opposite direction: given a domain model and an effect, abduction suggests some cause as a hypothesis that would explain the effect. Poole[1] discusses *probabilistic* Horn abduction, with prior probabilities $P(h_i)$ associated with hypotheses h_i, and relates it to Pearl's discrete Bayesian networks[2]. In short, a Bayesian network is a directed acyclic graph with nodes representing random variables and arcs representing directly influencing ('causing') relations. The key insight of Poole's work is that Bayes' rule $P(h|e) = P(e|h)P(h)/P(e)$ can be seen as a rule for abduction if we make some assumptions about the rule base T: it contains no directed cycles, the rules for every non-hypothesis are covering, and the bodies of rules in T are mutually exclusive. Now we may compute the prior probabilities for propositions p. Let H_p denote the set all minimal explanations of p (i.e., no proper subset of H_p is an explanation for p). Then $P(p) = P(\bigvee_{e_i \in H_p} e_i) = \sum_{e_i \in H_p} P(e_i)$. In addition, it is assumed that

[1] D. Poole. Learning, Bayesian probability, graphical models, and abduction. In P. Flach and A. Kakas, editors, *Abduction and Induction: essays on their relation and integration.* Kluwer, 1998. Forthcoming.

[2] E. Charniak. Bayesian networks without tears. *AI Magazine*, 12(4):50–63, 1991.

hypotheses are (unconditionally) independent. Hence, if $H = \{h_1, ..., h_n\}$ is a minimal explanation, then $P(h_1 \wedge ... \wedge h_n) = \prod_{i=1}^{n} P(h_i)$. If hypotheses are not independent, a new hypothesis is invented. This is justified by Reichenbach's *principle of the common cause*, i.e., if two events A and B co-occur more frequently than they occur independently, there exists a common cause for A and B. Under those restrictions, Bayesian networks can be translated to probabilistic Horn abduction and vice versa. Observe that the assumptions restrict *how* a domain is modeled, not *what* can be modeled. However, abduction is advantageous since it allows for a non-propositional representation. Arbitrary conditional probabilities are computed by using Bayes' rule as an abduction rule. First the evidence e is explained (yielding $P(e)$), and from those explanations, h is explained, which gives $P(h \wedge e)$ $(= P(e|h)P(e))$. In effect, we get the probability of h given e, $P(h|e)$, the posterior probability of h. Note however, that inference in Bayesian networks does not rely on Bayes' rule. Pearl, for instance, suggests a message-passing scheme instead, and Schurz[3] supplies probabilistic inference with an 'assumption generation' mechanism.

The task of *learning* can be seen as an evidential reasoning task, where given the data (evidence), we try to find models that 'cause' the data, and then use the models to make predictions. Bayesian learning uses Bayes' rule for the learning task, and has been applied to Bayesian networks, decision tree learning, and unsupervised learning, among others. It can be shown that Bayesian learning respects Occam's razor, "choose the shortest explanation for observed data" (via the minimum description length principle). In unsupervised learning, for example, such as *classification*, we are interested in what is the most probable classification of new evidence given the training data, rather than the most probable hypothesis (given the training data). As an example that gained much attention in the artificial intelligence community recently, we might want to learn how to assign news articles from the world-wide web to different categories. Interestingly, classification often employs a method similar to Reichenbach's principle: if certain nodes exhibit dependencies, a 'common cause', called *hidden variable*, is introduced to explain the dependency.

Our current implementation of Horn abduction[4] is *cost-based* rather than probabilistic. However, costs (associated with hypotheses) can be seen as –log probabilities and thus allow for meaningful semantics. Ohsawa and Ishizuka efficiently solve abductive problems as integer linear programming problems.

In summary, the claim of this contribution is that Bayesian learning can be performed by abduction as well, if hypotheses have associated probabilities (or costs), and the rules in the domain theory meet certain reasonable requirements. A challenging area for the application of abduction is the learning task of knowledge discovery. We hope to take up this topic soon.

---

[3] G. Schurz. Probabilistic default logic based on irrelevance and relevance assumptions. In *Proceedings First International Joint Conference on Qualitative and Quantitative Practical Reasoning (ECSQARU-FAPR'97)*, pages 536–553, 1997.

[4] Y. Ohsawa and M. Ishizuka. Networked bubble propagation: a polynomial-time hypothetical reasoning method for computing near-optimal solutions. *Artificial Intelligence*, 91:131–154, 1997.