Integrating Abduction and Induction in Machine Learning

Raymond J. Mooney
Department of Computer Sciences
University of Texas
Austin, TX 78712-1188, USA
mooney@cs.utexas.edu

Abstract

This paper discusses the integration of traditional abductive and inductive reasoning methods in the development of machine learning systems. In particular, the paper discusses our recent work in two areas: 1) The use of traditional abductive methods to propose revisions during theory refinement, where an existing knowledge base is modified to make it consistent with a set of empirical data; and 2) The use of inductive learning methods to automatically acquire from examples a diagnostic knowledge base used for abductive reasoning.

1 Introduction

Abduction is the process of inferring cause from effect or constructing explanations for observed events and is required for tasks such as diagnosis and plan recognition. Induction is the process of inferring general rules from specific data and is the primary task of machine learning. An important issue is how these two reasoning processes can be integrated, or how abduction can aid machine learning and how machine learning can acquire abductive theories. My research group has explored these issues in the development of several machine learning systems over the last eight years. In particular, we have developed methods for using abduction to identify faults and suggest repairs for theory refinement, and for inducing rule bases for abductive diagnosis. We treat induction and abduction as two distinct reasoning tasks, but have demonstrated that each can be of direct service to the other in developing AI systems for solving real-world problems. Below, I briefly review our work in these areas, focusing on the issue of how abduction and induction is integrated. [1]

2 Abduction and Induction

Precise definitions for abduction and induction are still somewhat controversial. In order to be concrete, I will generally assume that abduction and induction are both defined in the following general logical manner.

- **Given:** Background knowledge B and observations (data) O both represented as sets of formulae in first-order predicate calculus where O is restricted to ground formulae.
- **Find:** An hypothesis H (also a set of logical formulae) such that $B \cup H \not\vdash \bot$ and $B \cup H \vdash O$.

In abduction, H is generally restricted to set of atomic ground or existentially quantified formulae (called assumptions) and B is generally quite large relative to H. On the other hand, in induction, H generally consists of universally quantified Horn clauses (called a theory or knowledge base), and B is relatively small and may even be empty. In both cases, following Occam's Razor, it is preferred that H be kept as small and simple as possible.

Despite their limitations, these formal definitions encompass a significant fraction of the existing re-

[1] Additional details are available in our publications listed in the bibliography, most of which are available in postscript on the World Wide Web at `http://www.cs.utexas.edu/users/ml`.

search on abduction and induction, and the syntactic constraints on H capture at least some of the intuitive distinctions between the two reasoning methods. In abduction, the hypothesis is a specific set of assumptions that explain the observations of a particular case; while in induction, the hypothesis is a general theory that explains the observations across a number of cases. The body of logical work on abduction, e.g. [Pople, 1973; Poole, 1989; Levesque, 1989; Ng and Mooney, 1991; 1992; Kakas *et al.*, 1993], generally fits this definition of abduction and several diagnostic models [Reiter, 1987; Peng and Reggia, 1990] can be shown to be equivalent or a special case of it [Poole, 1989; Ng, 1992]. The work on *inductive logic programming* (ILP) [Muggleton, 1992; Lavrač and Džeroski, 1994] employs this definition of induction, and most machine learning work on induction can also be seen as fitting this paradigm [Michalski, 1983].

The intent of the current paper is not to debate the philosophical advantages and disadvantages of these definitions of induction and abduction; I believe this debate eventually becomes just a question of terminology. Given their acceptance by a fairly large body of researchers in both areas, a range of specific algorithms and systems have been developed for performing abductive and inductive reasoning as prescribed by these definitions. The claim of the current paper is that these existing methods can be fruitfully integrated to develop machine learning systems whose effectiveness has been experimentally demonstrated in several realistic applications.

3 Abduction in Theory Refinement

Theory refinement (*theory revision*, *knowledge-base refinement*) is the machine learning task of modifying an existing imperfect domain theory to make it consistent with a set of data. For logical theories, it can be more precisely defined as follows:

- **Given:** An initial theory T and a set of positive examples P and a set negative examples N where P and N are restricted to ground formulae.
- **Find:** A "minimally revised" consistent theory T' such that $\forall p \in P : T' \vdash p$ and $\forall n \in N : T' \not\vdash n$.

Generally, examples are ground Horn-clauses of the form C :- $B_1, \ldots, B_n$, where the body, B, gives a description of a case and the head, C, gives a conclusion or classification that should logically follow from this description (or should not follow in the case of a negative example). Revising a logical theory may require both adding and removing clauses as well as adding or removing literals from existing clauses. Generally, the ideal goal is to make the minimal syntactic change to the existing theory [Wogulis and Pazzani, 1993; Mooney, 1995]. Unfortunately, this task is computationally intractable; therefore, in practice, heuristic search methods must be used to approximate minimal syntactic change. Note that compared to the use of background knowledge in induction, theory refinement requires *modifying* the existing background knowledge rather than just adding clauses to it. Experimental results in a number of realistic applications have demonstrated that revising an existing imperfect knowledge base provided by an expert results in more accurate results than inducing a knowledge base from scratch [Ourston and Mooney, 1994; Towell and Shavlik, 1993].

Several theory refinement systems use abduction on individual examples to locate faults in a theory and suggest repairs [Ourston and Mooney, 1990; Ourston, 1991; Ourston and Mooney, 1994; Wogulis and Pazzani, 1993; Wogulis, 1994; Baffes and Mooney, 1993; Baffes, 1994; Baffes and Mooney, 1996; Brunk, 1996]. Each of these systems use abduction in a slightly different way, but the following discussion summarizes the basic approach. For each individual positive example that is not derivable from the current theory, abduction is applied to determine a set of assumptions that would allow it to be proved. These assumptions can then be used to make suggestions for modifying the theory. One potential repair is to learn a new rule for the assumed proposition so that it could be inferred from other known facts about the example. Another potential repair is to remove the assumed proposition from the list of antecedents of the rule in which it appears in the abductive explanation of the example. For example, consider the theory:

```
P(X) :- R(X), Q(X).
Q(X) :- S(X), T(X).
```

and the unprovable positive example:

```
P(a) :- R(a), S(a), V(a).
```

Abduction would find that the assumption `T(a)` makes this positive example provable. Therefore, two possible revisions to the theory are to remove the literal `T(X)` from the second clause in the theory, or to learn a new clause for `T(X)`, such as

```
T(X) :- V(X).
```

Another possible abductive assumption is `Q(a)`, suggesting the possible revisions of removing `Q(X)` from the first clause or learning a new clause for `Q(X)` such as

```
Q(X) :- V(X).
```

or

```
Q(X) :- S(X), V(X).
```

In order to find a small set of repairs that allow *all* of the positive examples to be proved, a greedy set covering algorithm can be used to select a small subset of the union of repair points suggested by the abductive explanations of individual positive examples, such that the resulting subset covers all of the positive examples. If simply deleting literals from a clause causes negative examples to be covered, inductive methods (e.g. ILP techniques like FOIL [Quinlan, 1990]) can be used to learn a new clause that is consistent with the negative examples. Continuing the example, assume the positive examples are:

```
P(a) :- R(a), S(a), V(a), W(a).
P(b) :- R(b), V(b), W(b).
```

and the negative examples are:

```
P(c) :- R(c), S(c).
P(d) :- R(d), W(d).
```

The abductive assumptions `Q(a)` and `Q(b)` are generated for the first and second positive examples respectively. Therefore, making a repair to the `Q` predicate would cover both cases. Note that the previously mentioned potential repairs to `T` would not cover the second example since the abductive assumption `T(b)` is not sufficient (both `T(b)` and `S(b)` must be assumed). Since a repair to the single predicate `Q` covers both positive examples, it is chosen. However, deleting the antecedent `Q(x)` from the first clause of the original theory would allow both of the negative examples to be proven.

Therefore, a new clause for `Q` is needed. Positive examples for `Q` are the required abductive assumptions `Q(a)` and `Q(b)`. Negative examples are `Q(c)` and `Q(d)` since these assumptions would allow the negative examples to be derived. Given the descriptions provided for `a, b, c` and `d` in the examples, an ILP system such as FOIL would induce the new clause:

```
Q(X) :- V(X).
```

since this is the simplest clause that covers both of the positive examples without covering either of the negatives. Note that although the alternative, equally-simple clause

```
Q(X) :- W(X)
```

covers both positive examples, it also covers the negative example `Q(d)`.

The EITHER [Ourston and Mooney, 1990; 1994; Ourston, 1991] and NEITHER [Baffes and Mooney, 1993; Baffes, 1994] theory refinement systems allow multiple assumptions in order to prove an example, preferring more specific assumptions, i.e. they employ *most-specific abduction* [Cox and Pietrzykowski, 1987]. AUDREY [Wogulis, 1991], AUDREY II [Wogulis and Pazzani, 1993], A3 [Wogulis, 1994], and CLARUS [Brunk, 1996] are a series of theory refinement systems that make a "single-fault assumption" during abduction. For each positive example, they find a single most-specific assumption that makes the example provable. Different constraints on abduction may result in different repairs being chosen, effecting the level of specificity at which the theory is refined. EITHER and NEITHER prefer making changes to the more specific aspects of the theory rather than modifying the top-level rules.

This general approach of using abduction to suggest theory repairs has proven quite successful at revising several real-world knowledge bases. The systems referenced above have significantly improved the accuracy of knowledge bases for detecting special DNA sequences called promoters that signal the start of a new gene [Ourston and Mooney, 1994; Baffes and Mooney, 1993], diagnosing diseased soybean plants [Ourston and Mooney, 1994], and determining when repayment is due on a student loan [Brunk, 1996]. The approach has also been successfully employed to construct rule-based models of student knowledge for over 50 students using an intelligent tutoring system for teaching concepts in C++ programming [Baffes, 1994; Baffes and Mooney, 1996]. In this application, theory refinement was used to modify correct knowledge of the domain to account for errors individual students made on a set of sample test questions. The resulting modifications to the correct knowledge base were then used to generate tailored instructional feedback for each student. In all of these cases, experiments with real training and test data were used to demonstrate that theory revision resulted in improved performance on novel, independent test data and generated more accurate knowledge than raw induction from the data alone. These results clearly demonstrate the utility of integrating abduction and induction for theory refinement.

We are currently developing a system for revising Bayesian networks [Pearl, 1988] using probabilistic abductive reasoning to isolate faults and suggest repairs [Ramachandran, 1995]. Bayesian networks are particularly appropriate for this approach since the standard inference procedures support both causal (predictive) and abductive (evidential) inference. Our technique focuses on revising a Bayesian network intended for causal inference by adapting it to fit a set of training examples of correct causal inference. Analogous to the logical approach outlined above, Bayesian abductive inference on each positive example is used to compute assumptions that would explain the correct inference and thereby suggest potential modifications to the existing network. The ability of this general approach to theory revision to employ probabilistic as well as logical methods of abduction is an interesting indication of its strength and generality.

4 Induction of Abductive Knowledge Bases

Another important aspect of integrating abduction and induction is the learning of abductive theories. Induction of abductive theories can be viewed as a variant of induction where the provability relation ($\vdash$) is itself interpreted abductively. In other words, given the learned theory it must be possible to *abductively* infer the correct conclusion for each of the training examples.

We have previously developed a learning system, LAB [Thompson and Mooney, 1994; Thompson, 1993], for inducing an abductive knowledge base appropriate for the diagnostic reasoning model of *parsimonious set covering* (PCT) [Peng and Reggia, 1990]. In PCT, a knowledge base consists of a set of `disorder` $\rightarrow$ `symptom` rules that demonstrate how individual disorders cause individual symptoms. Such an abductive knowledge base stands in contrast to the standard deductive `symptoms` $\rightarrow$ `disorder` rules used in standard expert systems and learned by traditional machine-learning methods. Given a set of symptoms for a particular case, the task of abductive diagnosis is to find a minimum set of disorders that explains all of the symptoms, i.e. a minimum covering set.

Given a set of training cases each consisting of a set of symptoms together with their correct diagnosis (set of disorders), LAB attempts to construct an abductive knowledge base such that the correct diagnosis for each training example is a minimum cover. The system uses a fairly straightforward hill-climbing induction algorithm. At each iteration, it adds to the developing knowledge base the individual `disorder` $\rightarrow$ `symptom` rule that maximally increases accuracy of abductive diagnosis over the complete set of training cases. The addition of rules terminate when the addition of any new rule fails to increase accuracy on the training data.

Using real data for diagnosing brain damage due to stroke originally assembled by [Tuhrim *et al.*, 1991], this technique was shown to produce an abductive knowledge base that, according to one important evaluation metric, was more accurate than an expert-built abductive rule base and the "deductive" knowledge bases learned by several standard machine-learning methods such as ID3 decision trees, FOIL Horn-clause rules, and neural networks trained using backpropagation.

LAB employs a fairly simple, restricted, propositional model of abduction and a simple, hill-climbing inductive algorithm. However, using techniques from inductive logic programming, the basic idea of using inductive learning methods to acquire abductive knowledge bases from examples could potentially be generalized to more expressive first-order representations. The existing results with LAB indicate the promise of exploring this approach. Finally, on-going research on the induction of Bayesian networks from data [Cooper and Herskovits, 1992] can be viewed as an alternative approach to learning knowledge that supports abductive inference.

5 Conclusions

In conclusion, we believe our previous and on-going work on integrating abduction and induction has effectively demonstrated two important points: 1) Abductive reasoning is useful in inductively revising existing knowledge bases to improve their accuracy; and 2) Inductive learning can be used to acquire accurate abductive theories. We have developed several machine-learning systems that integrate abduction and induction in both of these ways and experimentally demonstrated their ability to successfully aid the construction of AI systems for complex problems in medicine, molecular biology, and intelligent tutoring. However, our work has only begun to explore the potential benefits of integrating abductive and inductive reasoning. Further explorations into both of these general areas of integration will likely result in additional important discoveries and successful applications.

Acknowledgements

Many of the ideas reviewed in this paper were developed in collaboration with Dirk Ourston, Brad Richards, Paul Baffes, Cindi Thompson, and Sowmya Ramachandran. This research was partially supported by the National Science Foundation through grants IRI-9102926 and IRI-9310819, the Texas Advanced Research Projects program through grant ARP-003658-114, and the NASA Ames Research Center through grant NCC 2-629.

References

[Baffes and Mooney, 1993] P. Baffes and R.J. Mooney. Symbolic revision of theories with M-of-N rules. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 1135–1140, Chambery, France, Aug 1993.

[Baffes and Mooney, 1996] P. T. Baffes and R. J. Mooney. A novel application of theory refinement to student modeling. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence*, pages 403–408, Portland, OR, August 1996.

[Baffes, 1994] P. T. Baffes. *Automatic Student Modeling and Bug Library Construction using Theory Refinement.* PhD thesis, University of Texas, Austin, TX, August 1994.

[Brunk, 1996] C. A. Brunk. *An Investigation of Knowledge Intensive Approaches to Concept Learning and Theory Refinement.* PhD thesis, University of California, Irvine, CA, 1996.

[Cooper and Herskovits, 1992] G. G. Cooper and E. Herskovits. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning*, 9:309–347, 1992.

[Cox and Pietrzykowski, 1987] P. T. Cox and T. Pietrzykowski. General diagnosis by abductive inference. In *Proceedings of the 1987 Symposium on Logic Programming*, pages 183–189, 1987.

[Kakas *et al.*, 1993] A.C. Kakas, R. A. Kowalski, and F. Toni. Abductive logic programming. *Journal of Logic and Computation*, 2(6):719–770, 1993.

[Lavrač and Džeroski, 1994] N. Lavrač and S. Džeroski. *Inductive Logic Programming: Techniques and Applications.* Ellis Horwood, 1994.

[Levesque, 1989] H. J. Levesque. A knowledge-level account of abduction. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1061–1067, Detroit, MI, Aug 1989.

[Michalski, 1983] R. S. Michalski. A theory and methodology of inductive learning. *Artificial Intelligence*, 20:111–161, 1983.

[Mooney, 1995] R. J. Mooney. A preliminary PAC analysis of theory revision. In T. Petsche, S. Hanson, and J. Shavlik, editors, *Computational Learning Theory and Natural Learning Systems, Vol. 3*, pages 43–53. MIT Press, Cambridge, MA, 1995.

[Muggleton, 1992] S. H. Muggleton, editor. *Inductive Logic Programming.* Academic Press, New York, NY, 1992.

[Ng and Mooney, 1991] H. T. Ng and R. J. Mooney. An efficient first-order Horn-clause abduction system based on the ATMS. In *Proceedings of the Ninth National Conference on Artificial Intelligence*, pages 494–499, Anaheim, CA, July 1991.

[Ng and Mooney, 1992] H. T. Ng and R. J. Mooney. Abductive plan recognition and diagnosis: A comprehensive empirical evaluation. In *Proceedings of the Third International Conference on Principles of Knowledge Representation and Reasoning*, pages 499–508, Cambridge, MA, October 1992.

[Ng, 1992] H. T. Ng. *A General Abductive System with Applications to Plan Recognition and Diagnosis*. PhD thesis, University of Texas, Austin, TX, May 1992. Also appears as Artificial Intelligence Laboratory Technical Report AI 92-177.

[Ourston and Mooney, 1990] D. Ourston and R. Mooney. Changing the rules: A comprehensive approach to theory refinement. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pages 815–820, Detroit, MI, July 1990.

[Ourston and Mooney, 1994] D. Ourston and R. J. Mooney. Theory refinement combining analytical and empirical methods. *Artificial Intelligence*, 66:311–344, 1994.

[Ourston, 1991] D. Ourston. *Using Explanation-Based and Empirical Methods in Theory Revision*. PhD thesis, University of Texas, Austin, TX, August 1991. Also appears as Artificial Intelligence Laboratory Technical Report AI 91-164.

[Pearl, 1988] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, Inc., San Mateo,CA, 1988.

[Peng and Reggia, 1990] Yun Peng and James A. Reggia. *Abductive Inference Models for Diagnostic Problem-Solving*. Springer-Verlag, New York, 1990.

[Poole, 1989] D. Poole. Normality and faults in logic-based diagnosis. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1304–1310, Detroit, MI, 1989.

[Pople, 1973] Harry E. Pople, Jr. On the mechanization of abductive logic. In *Proceedings of the Third International Joint Conference on Artificial Intelligence*, pages 147–152, 1973.

[Quinlan, 1990] J.R. Quinlan. Learning logical definitions from relations. *Machine Learning*, 5(3):239–266, 1990.

[Ramachandran, 1995] Sowmya Ramachandran. Refinement of Bayesian networks by combining connectionist and symbolic techniques, 1995. Unpublished Ph.D. Thesis Proposal.

[Reiter, 1987] Raymond Reiter. A theory of diagnosis from first principles. *Artificial Intelligence*, 32:57–95, 1987.

[Thompson and Mooney, 1994] C. A. Thompson and R. J. Mooney. Inductive learning for abductive diagnosis. In *Proceedings of the Twelfth National Conference on Artificial Intelligence*, Seattle, WA, August 1994.

[Thompson, 1993] C. A. Thompson. Inductive learning for abductive diagnosis. Technical Report Masters Thesis, Department of Computer Sciences, University of Texas, Austin, TX, August 1993.

[Towell and Shavlik, 1993] G.G. Towell and J.W. Shavlik. Extracting refined rules from knowledge-based neural networks. *Machine Learning*, 13(1):71–102, 1993.

[Tuhrim *et al.*, 1991] Stanley Tuhrim, James Reggia, and Sharon Goodall. An experimental study of criteria for hypothesis plausibility. *Journal of Experimental and Theoretical Artificial Intelligence*, 3:129–144, 1991.

[Wogulis and Pazzani, 1993] J. Wogulis and M. Pazzani. A methodology for evaluating theory revision systems: Results with Audrey II. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 1128–1134, Chambery, France, 1993.

[Wogulis, 1991] J. Wogulis. Revising relational domain theories. In *Proceedings of the Eighth International Workshop on Machine Learning*, pages 462–466, Evanston, IL, June 1991.

[Wogulis, 1994] J. Wogulis. *An Approach to Repairing and Evaluating First-Order Theories Containing Multiple Concepts and Negation*. PhD thesis, University of California, Irvine, CA, 1994.