Abduction and Induction: an AI perspective

Yannis Dimopoulos and Antonis Kakas[1]

1 Introduction

In this short note[2] we compare the two separate forms of reasoning, induction and abduction, aiming at their effective integration within Artificial Intelligence (e.g. [1], [2], [6], [8] [5]).

The two inferences, abduction and induction, as they are commonly used in Artificial Intelligence, share the same basic formal specification: In a logic-based approach, given a background knowledge T (often called domain theory) that describes our problem domain, and an observation (or a set of observations) O, find a hypothesis H such that

$$T \cup H \models O \qquad (1)$$

However the two inferences differ significantly in the way they satisfy this specification as they typically use and extend this in different ways. Peirce [7] distinguishes the two forms of reasoning as follows:

(a) Abduction derives the case from the rule and the result.
(b) Induction derives the rule from the case and the result.

According to this view of Peirce, "result" denotes the observation O as new information, and in the case of abduction the given knowledge, namely the theory T, is a general rule whereas in the case of induction the given knowledge consists of specific "case" information.

2 A duality between Abduction and Induction

This simple characterization suggests in fact that it may be possible to see abduction and induction as used in AI, *dual* to each other. Consider the common specification (1) that these are required to satisfy and suppose first that we are given an empirical inductive learning framework. The background theory T_L, contains facts for various objects, and the set O of observations consists of training examples on a concept, c, that we want to learn. The hypothesis H_L constructed by our inductive learning is a set of rules for c whose conditions refer to the predicates (relations) in T_L. We have the learning equation

$$T_L \cup H_L \models O.$$

Consider now exchanging the role of T_L and H_L and regard H_L as a given background theory in an abductive framework and the facts in T_L as the possible abductive hypothesis. So we let $T_A = H_L$. Then we can construct abductive explanations, H_A, of O where H_A is drawn from T_L. Hence the abductive equation $T_A \cup H_A \models O$ can be written as

$$H_L \cup T_L' \models O$$

where T_L' is a subset of T_L. Comparing this equation with the original equation of learning above we see that one is dual to the other.

Conversely, consider now that we start from an abductive framework with background theory T_A, observations O on a single concept c defined in T_A and an explanation H_A of O. Can we use H_A to learn back the definition of c in T_A ? Let us illustrate the problem with an example.

Let T_A be the theory,
$tall(X) \leftarrow adult(X), European(X)$
$European(X) \leftarrow French(X)$
$European(X) \leftarrow Italian(X)$
$adult(Allan)$
$adult(Paolo)$
where the predicates "French" and "Italian" are abducible predicates and the other predicates are assumed to be completely defined by the theory.

Let also the set O of observations be $\{tall(Allan), tall(Paolo)\}$ and consider an abductive explanation H_A of O, say $H_A = \{French(Allan), Italian(Paolo)\}$. If we then take as background theory T_L the theory $T_A' \cup \{French(Allan), Italian(Paolo)\}$ where T_A' is the theory T_A without the rule for the predicate "tall" we can learn back this rule for "tall" from the same observations O as positive examples for this concept. Hence again we see we can effectively exchange the abductive hypothesis to become the background theory for learning and the abductive background theory to play the role of a hypothesis generated from learning. Starting from the abductive equation

$$T_A \cup H_A \models O$$

we have the learning equation

$$H_A' \cup T_A^c \models O$$

where the background theory H_A' is the closure of $H_A \cup (T_A - T_A^c)$ and T_A^c is the subset of T_A that defines the predicate "c" to be learned back.

This analysis suggests a methodology of how to integrate abduction in learning. Abduction can be used first to explain the training data of our learning problem to generate suitable or relevant background data on which to base the inductive generalization. In general, we may have many different abductive explanations for the same set of observations O. In the example above $H_A = \{French(Allan), French(Paolo)\}$ also forms an explanation. In such a case an appropriate background theory can again be obtained from all these possible explanations using the skeptical notion of abductive entailment (i.e. truth under all abductive explanations) with respect to the theory $T_A - T_A^c$. In [3] this dual correspondence from abductive problems to learning problems,

[1] Department of Computer Science, University of Cyprus, CY-1678, Nicosia, Cyprus. Email: {yannis, antonis}@turing.cs.ucy.ac.cy
[2] This paper has appeared as part of an extended paper [3] studying the possible integration of Abduction and Induction in Machine Learning.

ECAI 96. *12th European Conference on Artificial Intelligence*
Edited by W. Wahlster
Published in 1996 by John Wiley & Sons, Ltd.

is used to build a possible framework of integration of abduction and learning.

3 Handling Incomplete Information

Abduction and induction both provide a method for handling incomplete information in the description T of a problem. The addition of the generated hypothesis H to T completes (partially) this description. The two forms of reasoning though are concerned with different types of incompleteness in the background theory T and have different aims in dealing with this. Informally, we can say that abduction **extracts** the hypothesis H from T, while induction **constructs** its hypothesis using information from T. In the case of abduction, while the theory T is in general incomplete it is typically assumed that it contains complete (with respect to the external world that T is trying to model) definitions for some of its relations and the observations O are on these complete predicates. In this sense, the incompleteness of T is isolated in the abducible predicates on which we typically do not have direct information. The hypotheses H generated by abduction consists of formulae, usually facts, on these abducible predicates that aim to give an account (or an explanation) for the truth of the observations relative to the given theory T.

Abduction is thus viewed more as a way of reasoning with incomplete knowledge rather than as a mechanism for completing the knowledge in T. Its primary aim is to extract from the given theory T an explanation for the observation O and then reason with T augmented with this explanation. Completing the theory T with the information in the explanation is often a secondary issue. Consider for example the abductive framework of event calculus [4], [9]. The general theory T contains the axiom

$$\begin{aligned} holds(P, T_2) \quad \leftarrow \quad & happens(E, T_1), \\ & T_1 < T_2, \\ & initiates(E, P) \end{aligned}$$

together with other (domain specific) rules that define the "initiates" predicate. This general axiom expresses the persistence of a property P from the time T_1 that it is initiated by an event E to a later time T_2. The predicate "happens" is not defined (fully) by the theory and is designated as an abducible predicate. The incompleteness of the theory resides in this predicate.

This abductive theory is not intended to be used by first completing fully or learning the definition of the incomplete predicate "happens", a task which is general impossible or inappropriate[3] and then use deduction to infer what properties hold at different times. On the contrary the natural way to use this theory is to first assimilate any new information (observation) that a property holds at a particular time point through the explanation of this in terms of the happening of some event that initiates this property at an earlier point of time and then reason with the theory augmented with this explanation.

In induction the theory T may not in general have any complete predicates. The observations O are typically on predicates that are not defined by the theory and its primary task is to generate definitions for these in order to model the domain more accurately, or, in other words, complete the theory T. Their addition to T helps T to define more fully the interrelations between the different entities used to describe the domain. Hence the form of the generated hypothesis H can be as general as the language of T allows and it is typically a general rule (or rules). The addition of H to T aims in making T more complete in its general description of the problem domain.

[3] Similarly, once we have or learned the general laws of motion in nature (e.g. Newton's laws) we are not concerned with learning the boundary conditions for each object (e.g. the initial position and velocity) in the world.

4 The Utility of Abduction and Induction

The discussion in the previous section suggests that the utility of abduction and induction is in general different. Abduction is used to generate a reason for the truth of the observation O in terms of hypotheses H, which are typically specific to the situation and individual objects at hand. Each separate observation will get its own explanation not necessarily related to the explanations of other observations, and not intended to account for other (latter) observations. On the other hand, induction is used when we want to synthesize the information conveyed by the observations into a hypothesis which can account for all the observations together in a common (uniform) way. Moreover this hypothesis will also be able to account for other new observations on different objects than those involved in the original observations. Thus, typically an abductive hypothesis consists of specific facts further describing the situation at hand relative to the given theory, while in induction a hypothesis consists of general rules pertaining to a whole class of situations.

It is important to note that we are arguing here for a restricted meaning of abduction where explanations are generated (extracted) directly from the given theory T *based only* on the general laws that can be found in T. This means that abductive explanations typically do not contain new general rules. In contrast to this restricted view of abduction the term abduction has been used in some cases (e.g. [5]) more liberally allowing the inclusion of new general laws in the explanation. We argue here that in these cases the reasoning involved has a *complex form* that can be broken down to a combination of more basic abduction steps, of the restricted nature described above, together with inductive generalization steps.

For example, we may say that, given the theory $T =$ *"all bananas from Barbados are not sweet"* we can explain the observation(s) $O =$ *"all bananas in the shop are not sweet"* by the abductive explanation that "all bananas in the shop are from Barbados" (this example is taken from [5]). But we can also see this instead of a single process of abduction as comprising of two processes as follows. First we have a simple form of abduction to explain a typical case of "a banana being not sweet" and then we use (again a simple form of) induction to synthesize these different explanations, for each different banana in the shop together resulting in the final rule "all bananas in the shop are from Barbados". In the same spirit the term abduction has been used to name the process of generating new scientific theories, a process which can have a very complex form that can involve several separate more basic mechanisms.

These complex forms of reasoning named as abduction are thus viewed here as combinations of simpler and more basic computational mechanisms that include abductive and inductive inference as described above. In other words, the allowance in our explanations of general rules is not seen as an extension of the type of possible explanations that can be generated by abduction but rather as a form of synthesis (or generalization) together of several basic abductive explanations. This then helps us to separate and distinguish abductive and inductive processes.

REFERENCES

[1] H. Ade and M. Denecker. Abductive Inductive Logic Programming. *IJCAI-95*.

[2] W. W. Cohen. Incremental Abductive EBL. *Machine Learning*, Vol. 15, pp. 5-24, 1994.

[3] Y. Dimopoulos and A. Kakas. Abduction and Learning. In *Advances in ILP*. L. De Raedt (ed), IOS Press, 1996.

[4] Eshghi K. Abductive Planning with Event Calculus. *Proceedings 5th International Conference on Logic Programming*, 1988
[5] R. S. Michalski. Inferential Theory of Learning as a Conceptual Basis for Multistrategy Learning. *Machine Learning*, Vol. 11, pp 111-151, 1993.
[6] D. Ourston and R. J. Mooney. Theory Refinement Combining Analytical and Empirical Methods. *Artificial Intelligence*, 66, pp. 311–344, 1994.
[7] C. S. Peirce. *Collected Papers of Charles Sanders Peirce*, Vol. 2, 1931-1958. Hartshorn at al. (eds), Harvard University Press.
[8] C. Thompson and R. Mooney. Inductive Learning for Abductive Diagnosis. *Proc. of AAAI-94*.
[9] Shanahan M.P. Prediction is Deduction but Explanation is Abduction. *Proceedings IJCAI-89*, 1989