The Thirty-Second AAAI Conference
on Artificial Intelligence (AAAI-18)

Learning Abduction Under Partial Observability

Brendan Juba, Zongyi Li, Evan Miller
Dept. of Computer Science and Engineering
Washington University in St. Louis
{bjuba, zli, evan.a.miller} @wustl.edu

Abstract

Our work extends Juba's formulation of learning abductive reasoning from examples, in which both the relative plausibility of various explanations, as well as which explanations are valid, are learned directly from data. We extend the formulation to consider partially observed examples, along with declarative background knowledge about the missing data. We show that it is possible to use implicitly learned rules together with the explicitly given declarative knowledge to support hypotheses in the course of abduction. We observe that when a small explanation exists, it is possible to obtain a much-improved guarantee in the challenging *exception-tolerant* setting.

Introduction

Abduction is the task of inferring a plausible hypothesis to explain an observed or hypothetical condition. Although it is most prominently observed in scientific inquiry as the step of proposing a hypothesis to be investigated, it is also an everyday mode of inference. Simple tasks such as understanding stories (Hobbs et al. 1990) and images (Poole 1990) involve a process of abduction to infer an interpretation of the larger events, context, and motivations that are only partially depicted.

In this work, we consider a *PAC-learning* (Valiant 1984; 2000) formulation of the combined task of *learning to abduce*, introduced by Juba (2016). In this formulation, one is given a collection of examples drawn from the prior distribution (i.e., example jointly sampled values of attributes) together with a condition to explain, represented as a Boolean formula c on the attributes. The task is then to propose a formula h, which essentially must be a k-DNF for computational reasons, satisfying the following two criteria:

1. *Plausibility:* the probability that h is satisfied on the prior distribution must be at least some (given) minimum value $\mu > 0$
2. *Entailment:* the probability that the condition to explain c is satisfied, conditioned on the hypothesis h holding, is at least $1 - \epsilon$ for some given error tolerance $\epsilon > 0$.

Our works extend Juba's formulation of the abduction task to use partial examples and draw on declaratively specified background knowledge. We observe that by using a covering algorithm, it is possible to guarantee significantly better explanations when a small hypothesis (using relatively few terms) is adequate. Concretely, when some r-term k-DNF explanation on n attributes has an error rate of ϵ^*, we obtain an error rate of $\tilde{O}(r(\log\log n + \log k)\epsilon^*)$, in contrast to the bound obtained for the state-of-the-art algorithm of Zhang et al. (2017), which gave an error rate of $\tilde{O}(\sqrt{n^k}\epsilon^*)$ (but does not consider the effect of the size of the hypothesis).

Preliminaries

We work in a standard machine learning model in which the data consists of many *examples,* assigning boolean values to a variety of *attributes.* For example, if our data is about birds, each bird may correspond to an example and then there can be attributes such as: whether the bird has feathers or not, whether it eats bugs or not, and other properties.

Partial Observability In the real world, it is hard to require each example to contain all of the attributes. So, we want to make inferences with incomplete data. *Partial observability* means that some attributes of examples may be unknown. We represent this by allowing the value of each attribute to be 1 (true), 0 (false), or $*$ (unobserved). For instance, an example $\rho^{(i)}$ could be $[x_1 = 1,\ x_2 = *,\ \cdots,\ x_n = 0]$. (For convenience, we denote $\rho^{(i)}$ to be the ith example and ρ_i to be the ith coordinate of an example ρ.) In our abduction task, we say our partial examples are drawn from such a masked distribution $M(D)$.

Implicit Learning The main tool to deal with partial observability is *implicit learning*. Implicit learning means learning without producing explicit representations. Given a knowledge base (a set of formulas), and a query formula, we want to know if the knowledge base can derive the query formula. The main theorem of implicit learning says, as long as the formulas in a knowledge base are sufficiently observed in partial examples, we can determine whether the knowledge base can derive the query *without* explicitly constructing or representing the knowledge base.

Definition 1 (Witnessed Formula) *Given a partial example ρ, we say a formula ϕ is* witnessed *if $\phi|_\rho$ is 0 or 1.*

Where $\phi|_\rho$ means the formula ϕ restricted to example ρ. Informally speaking, we get the restricted formula by plugging in the observed value of the given example. For instance, let $\phi = x_1 \vee x_2$. In a partial example, $x_1 = 1; x_2 = *$. Then ϕ is witnessed (true) even x_2 is not observed. But it could be hard to determine the value when it gets complicated. Notice that each formula can be either witnessed true, witnessed false, or not witnessed.

Definition 2 (Proof System) *Given a knowledge base KB (a set of formulas), and a query formula ϕ, a* proof *is a finite sequence of formulas $\psi_1, \cdots, \psi_k$, such that:*
1. $\{\psi_1, \cdots, \psi_k\} \vdash \phi$.
2. $\forall i \in [1, k]$, *either* $\psi_i \in KB$ *or* $\{\psi_1, \cdots, \psi_{i-1}\} \vdash \psi_i$.
*Where "$\vdash$" means "can prove" or "*provable*".*

Each step of the proof $\{\psi_1, \cdots, \psi_{i-1}\} \vdash \psi_i$ corresponds to a relation $R_j(\psi_1, \cdots, \psi_{i-1}, \psi_i)$. A proof system *is a set of such relations $\{R_j\}_{j=0}^{\infty}$, i.e., such that whenever $R_j(\psi_1, \cdots, \psi_{i-1}, \psi_i)$ holds, $\{\psi_1, \cdots, \psi_{i-1}\} \vdash \psi_i$.*

DecidePAC Algorithm We have an algorithm that can tell whether a formula is provable or not, from the previous work (Juba 2013). Given knowledge base KB and partial examples $\{\rho^{(1)}, \cdots, \rho^{(m)}\}$ drawn from $M(D)$, for a query formula ϕ, DecidePAC can tell whether there is a proof of ϕ if the knowledge we need is witnessed sufficiently often: DecidePAC will *Accept* if there exists a proof of ϕ in from KB and formulas $\psi_1, \psi_2, \cdots$ that are simultaneously witnessed true with probability at least $1 - \epsilon + \gamma$ on $M(D)$; otherwise, if $[KB \Rightarrow \phi]$ is not $(1 - \epsilon - \gamma)$-valid, then DecidePAC will $reject$ formula ϕ.

Notice that there are three different concepts of being true: 1. $observed$ (or $witnessed$), 2. $provable$, and 3. $true$. We want to bridge from the witnessed values of examples to their ground truth, through logical inference.

Abduction under Partial Observability

Given a query or an event, abduction is the task of finding an explanation for the query or event. An explanation is a combination of some conditions that may have caused the query. For example, when the query is "Engine does not run," an explanation can be "No gas, or key is not turned."

We require the resulting explanation to satisfy two conditions, *"plausibility"* and *"entailment."* Entailment means that when the conditions in the explanation are true, the query should also often be true, or at least rarely false. Thus, the explanation is a (potential) cause of the query. Plausibility means the explanation is often true. In other words, for many examples, these conditions are observed. This suppresses unlikely explanations such as "A comet hits the car." which is a valid entailment, but not plausible.

Definition 3 (Partial Information Abduction) *Abduction is the following task: given any query formula c and independent partial examples $\{\rho^{(1)} \cdots \rho^{(m)}\}$ over a masked distribution $M(D)$, we want to find a k-DNF explanation h, such that the explanation h satisfies:*
1. $\Pr[\exists t \in h : t \text{ provable under } \rho] \geq \mu$ (Plausibility)
2. $\Pr\left[\begin{array}{r} \neg c \text{ provable} \\ \text{under } \rho \end{array} \middle| \begin{array}{r} \exists t \in h : t \text{ provable} \\ \text{under } \rho \end{array}\right] \leq \epsilon$ (Weak Entailment)

Implicit Abduction Algorithm

A k-DNF explanation is actually a disjunction of terms, $h = t_1 \vee t_2 \vee \cdots \vee t_r$. Each term represents a condition, or a possibility. Our goal is to find a formula that covers as many such conditions as possible while still being a potential cause of the query c.

We observe there is a natural correspondence between our k-DNF abduction task and set cover: each example of abduction is an element of the set cover problem, and each term is a set. We say a term covers an example when the term is provable in that example. The number of examples from the distribution is equivalent to its frequency or empirical probability with respect to the distribution M(D). If the resulting explanation consists of terms that are provable in most of examples, then we can conclude that our explanation is provable with high probability.

At a high level, our algorithm enumerate through all possible terms, check them with DecidePac to ensure "entailment", and then use greedy algorithm of set cover problem to find a collection satisfying "plausibility."

Theorem 4 (Implicit Abduction) *Given a query c, partial examples $\rho^{(1)}, \cdots, \rho^{(m)}$ from a masked distribution $M(D)$, and a restriction-closed proof system with Knowledge base KB:*

If there exists a r-term k-DNF $h^* = t_1^* \vee \cdots \vee t_r^*$ satisfying:
1. With probability at least $(1+\gamma)\mu$, $\exists t_i^* \in h^*$, such that t_i^* is provable under ρ from KB (Plausibility).
2. If some term t^* of h^* is provable, then $\neg c$ is only provable with probability at most $(1-\gamma)\epsilon$. (Weak Entailment)

Then, we can find a k-DNF h in polynomial time, such that with probability $1 - \delta$,
1. $\Pr[\exists t \in h \text{ provable under } \rho] \geq (1-\gamma)\mu$ *(Plausibility)*
2. $\Pr\left[\begin{array}{r} \neg c \text{ provable} \\ \text{under } \rho \end{array} \middle| \begin{array}{r} \exists t \in h \text{ provable} \\ \text{under } \rho \end{array}\right] < \tilde{O}(r(\log\log n + \log k + \log r + \log\log \frac{1}{\delta} + \log \frac{1}{\gamma})(1+\gamma)\epsilon))$ *(Weak Entailment).*

References

Hobbs, J.; Stickel, M.; Appelt, D.; and Martin, P. 1990. Interpretation as abduction. Technical Report 499, SRI, Menlo Park, CA.

Juba, B. 2013. Implicit learning of common sense for reasoning. In *Proc. 23rd IJCAI*, 939–946.

Juba, B. 2016. Learning abductive reasoning using random examples. In *Proc. 30th AAAI*, 999–1007.

Poole, D. 1990. A methodology for using a default and abductive reasoning system. *Int'l J. Intelligent Sys.* 5:521–548.

Valiant, L. G. 1984. A theory of the learnable. *Communications of the ACM* 18(11):1134–1142.

Valiant, L. G. 2000. Robust logics. *Artificial Intelligence* 117:231–253.

Zhang, M.; Mathew, T.; and Juba, B. 2017. An improved algorithm for learning to perform abduction. In *Proc. 31st AAAI*, 1257–1265.