CSE 513T Theory of Artificial Intelligence & Machine Learning Nov 1th, 2016

Lecture 17

Lecturer: Brendan Juba *Scribe: Haoyang Rong*

1 Today's Topics

1.1 Partial information & masking processes

1.2 Testable rules & implicit learning

1.3 Restriction-closed proofs

1.4 (??) algorithm for reasoning with partial examples

2 Classical reasoning

Setting where no rules of any attributes are assumed to be known, the purpose of reasoning is to "fill in" our information (note specifically, chaining) Reasoning with examples:a setting in which the settings of all attributes are known: the purpose of reasoning is just to clarify formulas as true or false

3 Partial information

Formally, we still assume that there is some probability distribution D on X, and we wish to know whether given query formula is $(1-\varepsilon)-valid$ on D (i.e., $\Pr_{x\in D}[\varphi(x)] \geq 1-\varepsilon$)

Now, we'll only have access to partial examples $\varrho \in \{0,1,*\}^n$, obtained by applying a masking function $m : \{0,1\}^n \to \{0,1,*\}^n$ to a complete $x \in D$

Def: A masking function $m : \{0,1\}^n \to \{0,1,*\}^n$ is any function s.t. $m(x)$ is consistent with x, i.e. whenever $m(x)_i = x_i$

The masking functions themselves will, in general, be produced by a random process: we say that a masking function-valued random variable M is a masking process.

We'll assume that the random draws of $m \in M$ is independent of the draws of $x \in D$. Nevertheless, the partial examples$\varrho^{(i)} = m^{(i)}(x^{(i)})$ $(i = 1, ..., m)$ may depend on $x\&m$, and in particular, which attributes are hidden can depend on X

Example 1: Suppose M produces a single function m that always hides $x_1, ..., x_{fracn2}$ and never hides $x_{fracn2+1}, ..., x_n$. This captures a setting in which we measure or record $x_{fracn2+1}, ..., x_n$. But we are interested in answering question queries about all of the attributes (our formula can still refer to all n attributes)

Example 2: "Lossy data" Suppose that for $mu \in (0,1)$, M_μ chooses a masking function m by tossing a $\mu - biased$ coin for each attribute, and producing a masking function that hides precisely the attributes for which the coin came up "heads". M_μ ranges are functions m $s.t.$ $\forall i$ $either$ $(\forall x,\ m(x)_i = *)$ or $(\forall x,\ m(x)_i = x_i)$ For functions m that mask s attributes, $M_\mu(m) = m^s(1-\mu)^{n-s}$.

Examples 3: "Limited visibility" Suppose the attributes encode (1) the position of a sensor in an environment. Then, we can capture the sensor only recording data at its current position by a masking process that uses the value if the first type of attribute to decide which of the second type of attributes to hide/reveal.

4 Testable rules

We wish to say that the various rules that are "demonstrably" $(1-\varepsilon)-valid$ w.r.t. the partial information provided by draws from $M(D)$ will be available for reasoning as part of an "implicitly learned KB".

We need to be careful in defining "demonstrably valid" -every tautolagous 3DNF is $(1-\varepsilon)-valid$ for every distribution, but determining this without knowledge of the attributes is believed to be intractable.

One simple test we can use: we apply the restriction corresponding to a partial example ϱ to a formula, and check to see if this evaluates the formula to "1" (recall for a $CNF = \varphi_1$, this occurs when ϱ sets at least one literal in every clause of φ to "1"), e.g., if $C = \neg puppy \wedge mischief$, then if ϱ either sets mischief to "1" or puppy to "0", then C is evaluated under ϱ Notice, this test can be performed in time linear in $|\varrho|$

Def: We say that a representation φ is $(1-\varepsilon)-testable$ w.r.t. $M(D)$ if for $\varrho \in M(D)$, $\Pr_{\varrho \in M(D)}[\varphi|\varrho = 1] \geq 1-\varepsilon$.

Example: If in $D|bird = 1$, $M(D|bird)$ produces ϱ in which $can_fly = s1, w.p. \geq 99\%$, then the literal "can_fly" is "99%-testable" w.r.t. $M(D|bird)$ and we'll seek to show that then "can_fly" is available as part of our implicit KB.

5 Restriction-closed proofs

Def: We say that a set of proofs (e.g. a set of space-bounded treelike resolution) P is closed under restrictions if whenever there is a proof $\pi = \pi_1, \pi_2, ..., \pi_l$ of φ from premises $\psi_1, ..., \psi_k$ in P, then for every partial assignment ϱ, there is a proof of $\phi|\varrho$ from $\psi_1|_\varrho, ..., \psi_k|_\varrho$ in P, given that none of these simplified to 0.

It's actually challenging to propose a "natural" set of proofs that violates this condition.

Claim: Chaining proofs of $size \leq l$ are closed under restrictions. Strategy: start from a chaining proof of $size \leq l$, apply the restriction ϱ to each line of this proof. The line is either "simplified" or evaluated. We'll just drop the evaluated parts.

Example: Recall our proof of "mischeif" from "$dog, young,\ (dog \wedge young) \Rightarrow puppy, puppy \Rightarrow mischief$. Suppose ϱ sets $dog = 1$ & $puppy = 1$, so the residual proof (from young, mischief) is young, mischief, I could drop "young", and still have a proof of mischief.

Proof: Consider the chaining proof $\pi = \pi_1, \pi_2, ..., \pi_?$ of φ from $\psi_1, ..., \psi_k$ and any partial assignment ϱ that does not set $\psi_1|_\varrho = 0$,..., or $\psi_k|_\varrho = 0$.

Consider the proof π (? seem to be π'?) obtained by

(1) Applying ϱ to each line of π

(2) Omitting any portion of the proof that is only connected to the conclusion (sink of the graph) by paths traversing lines π_i s.t. $\pi_i|_\varrho$ is a constant. Now π' only contains "facts" that are not set by ϱ, and only contains rules s.t. (1) the head is not fixed to a constant, and (2) no literal in the body is "fixed to 0" by ϱ.

We'll argue by induction on the steps of the surviving proof that this is a chaining proof of $\varphi|_\varrho$ from $\psi_1|_\varrho, ..., \psi_k|_\varrho$.

Base: "0th step" (empty proof) -trivial.

Inductively: we have a syntactically correct chaining derivation up to step $i-1$. Step $i - \pi_i$ was a fact, then since it was not omitted it must not have been set by ϱ, so the fact $\pi_i|_\varrho$ must be among $\psi_1|_\varrho, ..., \psi_k|_\varrho$, so we can assert it.

If π_i was a rule, then either $\pi_i|_\varrho$ is a fact (the head of the rule π_i) or $\pi_i|_\varrho$ is itself a rule. In the first case, we again have the fact $\pi_i|_\varrho$ among $\psi_1|_\varrho, ..., \psi_k|_\varrho$, and in the second case, we know that also there is a corresponding rule $\pi_i|_\varrho$ in $\psi_1|_\varrho, ..., \psi_k|_\varrho$.

IN the final case, in which it was derived by chaining, we know that the original proof had derived the literals in the body $l_1, ..., l_k$ on some previous steps; by IH (¡-what is this?), we have derived these same literals whenever they aren't set to constant. Because this portion would have been dropped if the head was set to a constant or if the rules were falsified. We also have that the restriction of the rule we used is also asserted at some previous step. Therefore we are allowed to apply chaining to derive π_i.

This completes the induction step. Note that π_i is no larger than original π