CSE 513T Theory of Artificial Intelligence & Machine Learning March 5, 2015

Lecture 16

Lecturer: Brendan Juba *Scribe: Allen Lavoie*

1 Final projects

- Discuss ideas with Professor Juba shortly after Spring Break
- Figure out who you want to work with (can be same as for homework)
- Ideally projects would be something new, but have a backup plan (e.g. a survey of known results on a topic we didn't cover very well in class)
- Topics are anything you want. Some coding is fine, but not required.

2 Efficient algorithm for (resolution) proofs

- Focus on finding proofs where only size s working memory is required (space-s proof)
- From HW3, we know that "space-2 treelike resolution" subsumes chaining, so even small-space proofs can be interesting.
- We have efficient algorithms for finding space-bounded proofs.

Theorem 1 *There is an algorithm SpaceRES(s, φ) that returns a space-s treelike refutation of φ if one exists and runs in time $O(|\varphi|n^{2s-1})$ where $|\varphi|$ is the size of φ (e.g. in bits) and n is the number of attributes.*

Algorithm 1 SpaceRES(s, φ)

```
if φ = ⊥ ∧ ... then
    return ⊥
else if φ = 1 then
    return FAIL
else if s > 1 then                          ▷ To do anything, we need space s > 1
    for each literal l in φ do
        if SpaceRES(s − 1, φ|_{l=1}) = Π^l (≠ FAIL) then
            if SpaceRES(s, φ|_{l=0}) = Π^{¬l} (≠ FAIL) then
                return Π^l Π^{¬l} ⊥
            else
                return FAIL
return FAIL
```

2.1 Proof of SpaceRES correctness

We use the characterization of space-s treelike proofs as those with a rank $s-1$ graph. By induction on $s > n$:

Base case: $s = 1$ **(rank 0)** The only possible derivation that ends with '$\bot$' in a single step is the proof '$\bot$', which is only a proof if '$\bot$' was a premise. Likewise for $n = 0$: with no attributes, either $\varphi = \bot$ or $\varphi = 1$.

Assume the induction hypothesis for ($n' < n$ and $s' \leq s$) or ($n' = n$ and $s' < s$) Let some space-s treelike refutation of φ be given. One of the children of the final step uses space $s-1$ for the literal $\neg l$ derived on that branch. When we fix l to 1, then the proof gives us a space-$(s-1)$ derivation of $\neg\varphi|_{l=1} = \bot$. This follows immediately from the following claim:

Proposition 2 *(proof is part of HW4) Suppose that φ has a space-s treelike refutation and ρ is* any *partial assignment. Then $\varphi|_\rho$ also has a space-s treelike resolution refutation.*

So for at least one $\tilde{l}$, a call to SpaceRES($s-1$, $\varphi|_{\tilde{l}=1}$) returns a proof $\Pi^{\tilde{l}}$ (if it's a refutation of φ we'd be done, otherwise it derives $\neg\tilde{l}$). We don't know that $\tilde{l} = l$ (as in our given refutation). However, we assumed that there was a space-s refutation of φ, so by the claim above there is still a refutation of $\varphi|_{\tilde{l}=0}$. But $\varphi|_{\tilde{l}=0}$ has one less attribute, so by the induction hypothesis SpaceRES(s, $\varphi|_{\tilde{l}=0}$) finds a refutation of $\varphi|_{\tilde{l}=0}$.

This is either a refutation of φ already, or else it derives $\tilde{l}$ (since $\varphi|_{\tilde{l}=0} = \bot$), so applying cut to $\tilde{l}$ in this latter case (following $\Pi^{\tilde{l}}$ and $\Pi^{\neg\tilde{l}}$) is a refutation of φ, and since $\Pi^{\tilde{l}}$ had a graph of rank $\leq s-2$ and $\Pi^{\neg\tilde{l}}$ had a graph of rank $\leq s-1$, this refutation is of rank $\leq s-1$.

2.2 Proof of SpaceRES running time

The claimed running time bound is $C|\varphi|n^{2s-1}$. Base cases $s = 1$ or $n = 1$ take time $O(1)$ in either case, so we just need C large enough.

Inductive step For recursive calls, we spend time:

$$\leq 2nC|\varphi|(n-1)^{2(s-1)-1} + C|\varphi|(n-1)^{2s-1}$$

Since $s \geq 2$, $2(s-1)-1 \geq 1$ and $2s-1 \geq 3$. Then the time spent on recursive calls is at most:

$$\begin{aligned}
&\frac{n-1}{n}2nC|\varphi|n^{2(s-1)-1} + \left(\frac{n-1}{n}\right)^3 C|\varphi|n^{2s-1} \\
=&C|\varphi|\left[2\frac{n-1}{n^2} + \frac{n-1}{n}\left(\frac{n-1}{n}\right)^2\right]n^{2s-1} \\
=&C|\varphi|\left[\frac{2}{n} - \frac{2}{n^2} + \left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}+\frac{1}{n^2}\right)\right]n^{2s-1} \\
=&C|\varphi|\left[1 - \frac{1}{n^2} - \frac{1}{n}\left(\frac{n-1}{n}\right)^2\right]n^{2s-1} \\
=&C|\varphi|n^{2s-1} - \left[\frac{1}{n} + \left(\frac{n-1}{n}\right)^2\right]C|\varphi|n^{2s-2} \\
=&C|\varphi|n^{2s-1} - \left[1 - \frac{1}{n} + \frac{1}{n^2}\right]C|\varphi|n^{2s-2}
\end{aligned}$$

Since $n \geq 2$, we have that $-1/n + 1/n^2 \geq -1/4$. So as long as the body of SpaceRES takes time $\leq \frac{3C}{4}|\varphi|n^2$, the bound holds under the induction hypothesis (note: in class we used the looser bound $-1/n + 1/n^2 \geq -1/2$, but it doesn't make any difference for the analysis).

Achieving this running time is not difficult for each of at most $2n$ literals. We just need to prepare $\varphi|_{l=b}$, which takes time $O(|\varphi|)$. So overall as long as C is large enough, this is time $\leq \frac{3C}{4}|\varphi|n^2$.

2.2.1 Notes on running time

The refutations we find with SpaceRES only ever apply a cut rule once on an attribute on each branch. So if we label each step of the proof with the eliminated variable, the labels appear at most once on each branch, so the relationships between tree rank and tree size apply.

In particular, in order for a proof to have length l, it must use space at least:

$$s \geq \log\left(\frac{l+1}{2}\right) + 1$$

Corollary 3 *SpaceRES($\log l + 1$, φ) runs in time $|\varphi| 2^{O(\log n \log l)}$ and returns a treelike resolution refutation of φ if one of length $\leq l$ exists.*

It is open whether a similar algorithm with a polynomial time bound exists.

For general resolution, the fastest algorithms for finding proofs of length $l = \text{poly}(n)$ run in time $n^{O(\sqrt{n})}$, although CDCL SAT solvers seem to do much better in practice. It is open whether this can be shown formally.

To summarize, at least with our current knowledge, finding richer proofs comes at a substantial cost.

3 Reasoning with examples (PAC learning)

We suppose again that there is a distribution D over the instance space X, and that we have access to examples drawn independently from D.

Classical reasoning task Given a query representation φ, decide whether KB $\models \varphi$ (i.e. the knowledge base entails φ), based on the tacit assumption that the knowledge base is valid for the *set* of actual worlds "D" (different from the distribution D above, but having the same role).

PAC-semantics decide whether:

$$\Pr_{x \in D}[\varphi(x) = 1] \geq 1 - \epsilon$$

(On HW3, we'll show that entailment in classical settings carries over to PAC-semantics)

But it turns out that given enough examples, deciding whether φ is $(1-\epsilon)$-valid with respect to D (i.e. $\Pr_{x \in D}[\varphi(x) = 1] \geq 1 - \epsilon$) is easy. This relates to the golf game, Question 1 on Homework 1. By the Chernoff bound (looking that the fraction of times $\varphi(x^{(i)}) = 1$), we have that

$$\frac{1}{m} \sum_{i=1}^{m} \varphi(x^{(i)})$$

is within an additive γ of $\Pr[\varphi(x) = 1]$ with probability $\geq 1 - 2e^{-2m\gamma^2}$. So given $\frac{1}{2\gamma^2} \ln \frac{2}{\delta}$ examples, our estimate is within γ with probability $\geq 1 - \delta$.

We can also use an Occam's razor-style bound. We find that given

$$m \geq \frac{1}{2\gamma^2} \ln \frac{2 \cdot 2^{B+1}}{\delta} = \frac{\ln 2}{2\gamma^2}\left(B + 2 + \ln \frac{1}{\delta}\right)$$

every B-bit representation φ has $\frac{1}{m}\sum_{i=1}^{m} \varphi(x^{(i)})$ within $\pm\gamma$ of $\Pr[\varphi(x) = 1]$ with probability $1 - \delta$. The proof is standard Occam's razor-style, a union bound of 2^{B+1} bad events of probability $\frac{\delta}{2^{B+1}}$.

Since there are only 2^{n^3} 3-CNFs, these have $O(n^3)$-bit representations, so a set of $O\left(\frac{1}{\gamma^2}\left(n^3 + \ln \frac{1}{\delta}\right)\right)$ examples suffices to answer *all* 3-CNF queries.

Partial information One issue is that this requires complete information. In examples with partial information, we need classical techniques to find values of attributes we are not given.