A PROOFS OF THEOREMS 1 AND 2

Theorem 1. *The reduction from SSAT to POMDP guarantees that there exists a POMDP policy π for time steps 0 to $|X|/2-1$ and optimal action at time step $|X|/2$ with value function $V^\pi = \Pr(\phi)$ iff there exists a policy tree ϕ with satisfiability probability $\Pr(\phi)$.*

Proof. Consider a POMDP policy π (for time steps 0 to $|X|/2-1$), which defines a policy tree ϕ. Each branch yields a final (unnormalized) belief with mass

$$\hat{b}^\pi_{o_{1:|X|/2}}(prob) = b_0(prob)\Pr(o_{1:|X|/2}|prob,\pi) \tag{1}$$

Based on the properties of the reward function, the optimal expected reward of each branch at the last time step $|X|/2$ is

$$R(\hat{b}^\pi_{o_{1:|X|/2}}) = \max_a \sum_s \hat{b}^\pi_{o_{1:|X|/2}}(s)R(s,a) \tag{2}$$

$$= \begin{cases} \Pr(o_{1:|X|/2}|prob,\pi) & \text{if branch is satisfying} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Hence the value of a policy is

$$V^\pi = \sum_{o_{1:|X|/2}} R(\hat{b}^\pi_{o_{1:|X|/2}}) \tag{4}$$

$$= \sum_{o_{1:|X|/2}\text{ is satisfying}} \Pr(o_{1:|X|/2}|prob,\pi) \tag{5}$$

$$= Pr(\phi) \tag{6}$$

The above equation shows that the value of a policy is equal to the probability of satisfying the Boolean formula with the corresponding policy tree ϕ. □

Theorem 2. *In the reduction of POMDP to SSAT, there exists a satisfiable policy tree, ϕ, with probability $\Pr(\phi)$ iff there exists a POMDP policy, π, with value function $V^\pi = \Pr(\phi)$.*

Proof. Consider a base case policy tree of size 1. Let the policy tree be $\phi = \{x_a \equiv \hat{k}\}$ with clauses:

$$\bigwedge_{i\in\mathcal{S}} x_s \neq i \vee x_r \equiv \hat{k}|\mathcal{S}| + i \tag{7}$$

The probability of satisfiability of (7) is equivalent to

$$\Pr(\phi) = \sum_i \Pr(x_s \equiv i)\Pr(x_r \equiv \hat{k}|\mathcal{S}| + i)$$

$$= \sum_i b(i)r(i,\hat{k}) \tag{8}$$

by using the distributions for the randomized variables: $\Pr(x_s \equiv i) = b(i)$ and $\Pr(x_r \equiv k|\mathcal{S}| + i) = r(i,k), \forall i,k$. However, (8) corresponds exactly to the policy that takes action $a_1 = \hat{k}$ and has a value of $V^\pi = \sum_i b(i)r(i,\hat{k})$.

For the general case, we give a proof by induction. Assume we have a policy tree ϕ_h, policy π_h, and we know $\Pr(\phi_h) = V^{\pi_h}$. Given ϕ_{h+1} and π_{h+1} show that $\Pr(\phi_{h+1}) = V^{\pi_{h+1}}$.

Since we are given the policy tree, all the actions are known. Therefore, if we simplify first by making the assignments in ϕ_{h+1}, then only the randomized variables will remain in the quantifier prefix. Any subset of variables can now be re-ordered freely. Based on the number of randomized variables we introduced for horizon h and $h+1$, encoding the probability of satisfiability is:

$$\Pr(\phi_{h+1})$$

$$= \sum_{v_1,\cdots,v_{h+1}}^{2} \sum_{z_1,\cdots,z_h}^{|\mathcal{O}|} \sum_{s_1,\cdots,s_{h+1}}^{|\mathcal{S}|} \prod_{l=1}^{h+1} \Pr(x_p^l = v_l, x_s^l = i, x_o^l = z_l, x_r^l) \prod_{l=1}^{h} \Pr(x_\Omega^l, x_T^l | x_p^l = v_l, x_s^l = i, x_o^l = z_l) \tag{9}$$

To achieve Eq. 10, the distribution for x_p is just a uniform distribution that can be factored out as 2^{-h}. However, each x_p is controlling the length of the process, so it naturally controls how many terms contribute to the total sum if we re-arrange by horizon and then simplify. Note that given values for x_p, x_o, x_s the other variables are forced by unit propagation to a specific value.

$$= 2^{-(h+1)} \sum_{\hat{h}=1}^{h+1} \sum_{z_1,\cdots,z_{\hat{h}-1}}^{|\mathcal{O}|} \sum_{s_1,\cdots,s_{\hat{h}}}^{|\mathcal{S}|} \prod_{l=1}^{\hat{h}} \Pr(x_s^l = i, x_o^l = z_l, x_r^l) \prod_{l=1}^{\hat{h}-1} \Pr(x_\Omega^l, x_T^l | x_p^l = v_l, x_s^l = i, x_o^l = z_l) \tag{10}$$

Similarly, for the distribution x_o the constant, $|O|^{h-1}$, can be factored out in front and its value is used in the conditional distribution x_Ω.

$$= 2^{-(h+1)}|O|^{-h} \sum_{\hat{h}=1}^{h+1} \sum_{z_1,\cdots,z_{\hat{h}-1}}^{|\mathcal{O}|} \sum_{s_1,\cdots,s_{\hat{h}}}^{|\mathcal{S}|} \prod_{l=1}^{\hat{h}} \Pr(x_s^l = i, x_r^l) \prod_{l=1}^{\hat{h}-1} \Pr(x_\Omega^l, x_T^l | x_p^l = v_l, x_s^l = i, x_o^l = z_l) \tag{11}$$

the next variable x_s^l has uniform distribution for all $l > 1$ and the initial belief when $l = 1$. Therefore, we can simplify the equation by pulling out the constant factors again.

$$= 2^{-(h+1)}(|O| \cdot |S|)|^{-h} \sum_{\hat{h}=1}^{h+1} \sum_{z_1,\cdots,z_{\hat{h}-1}}^{|\mathcal{O}|} \sum_{s_1,\cdots,s_{\hat{h}}}^{|\mathcal{S}|} \Pr(x_s^1 = i) \prod_{l=1}^{\hat{h}} \Pr(x_r^l) \prod_{l=1}^{\hat{h}-1} \Pr(x_\Omega^l, x_T^l | x_p^l = v_l, x_s^l = i, x_o^l = z_l) \tag{12}$$

According to the distribution x_p, rewards x_r will only be given at the end of the process for each $\hat{h}$.

$$= 2^{-(h+1)}(|O| \cdot |S|)^{-h} \sum_{\hat{h}=1}^{h+1} \sum_{z_1,\cdots,z_{\hat{h}-1}}^{|\mathcal{O}|} \sum_{s_1,\cdots,s_{\hat{h}}}^{|\mathcal{S}|} \Pr(x_s^1 = i) \Pr(x_r^{\hat{h}})$$
$$\prod_{l=1}^{\hat{h}-1} \Pr(x_\Omega^l, x_T^l | x_p^l = v_l, x_s^l = i, x_o^l = z_l) \tag{13}$$

If we replace the distributions below with their definitions and replace constants with the proportional relation, we obtain

$$\propto \sum_{\hat{h}=1}^{h+1} \sum_{z_1,\cdots,z_{\hat{h}-1}}^{|\mathcal{O}|} \sum_{s_1,\cdots,s_{\hat{h}}}^{|\mathcal{S}|} b(s_1) \prod_{l=1}^{\hat{h}-1} \Omega_{s_{l+1},z_l}^{a_l} T_{s_l,s_{l+1}}^{a_l} r(s_{\hat{h}}, a_{\hat{h}}) \tag{14}$$

$$= \sum_{s_1}^{|\mathcal{S}|} b(s_1) \left(r(s_1, a_1) + \sum_{z_1}^{|\mathcal{O}|} \sum_{s_2}^{|\mathcal{S}|} \Omega_{s_2,z_1}^{a_1} T_{s_1,s_2}^{a_1} \Pr(\phi_h) \right) \tag{15}$$

where $\Pr(\phi_h) = r(s,a) + \sum_z^{|\mathcal{O}|} \sum_{s'}^{|\mathcal{S}|} \Omega_{s',z}^a T_{s,s'}^a \Pr(\phi_{h-1})$

Now consider the reverse. Given a policy, π_{h+1}, with value function $V^{\pi_{h+1}}$ there exists a satisfiable policy tree, ϕ_{h+1}, with satisfiability probability $\Pr(\phi_{h+1})$ such that $V^{\pi_{h+1}} = \Pr(\phi_{h+1})$. First, Bellman's equation for a $h+1$ horizon policy is:

$$V^{\pi_{h+1}} = \sum_s b^{h+1}(s) \left(r(s,a) + \sum_o \sum_{s'} \Omega_{s'o}^a T_{ss'}^a V^{\pi_h}(b_o^a) \right), \; a = \pi(b) \tag{16}$$

However, any $h+1$ horizon policy can be written as a linear combination of h horizon policies. Since we know $Pr(\phi_h) = V_h^\pi$ by the inductive step, we conclude, that (15) and (16) are equal. Therefore, the probability of satisfying a $h+1$ depth policy tree corresponds to the value function of a $h+1$ step policy. $\square$

B PROBLEM STATISTICS

We test the improvements to the watch literal rule on a variety of problems from 3 different benchmark types as shown in Table 1. The POMDP problems are from Cassandra's repository [?] and consist of two easy and two hard problems that have quite a large number of literals per clause and variable cardinality. The inference problems are from a prior probabilistic inference competition [?] and tend to be highly structured and contain a large number of variables and clauses.

Finally, the random benchmarks consist of a series of variables with alternating quantifiers in 3-SAT and 10-SAT forms that were generated by a procedure. Assume we are given V the number of variables, C the number of clauses, k the number of literals in a clause, t the number of values for each variable and p the probability for each variable to be existentially quantified ($1-p$ is the probability for each variable to be randomly quantified). We can generate a problem by first sampling the quantifier for each variable $Q(v_i)$ and if randomly quantified, draw its distribution from a uniform Dirichlet with dimension t. For each clause c_i where $i \in \{0, ..., C-1\}$ a variable is sampled uniformly from $\{1, ..., V\}$ and a value is sampled uniformly from $\{0, ..., t-1\}$ repeatedly to generate k literals for each clause.

Benchmark	Problem	#var	#clause	avg #value	avg #literal
RANDOM	fail-learn1	50	120	2.00	3.00
	pure1	50	120	2.00	3.00
	big1	30	450	2.00	10.00
	big2	15	60	4.00	10.00
POMDP	tiger.95_H10	157	304	2.31	5.60
	ejs7_H10	121	212	2.16	4.58
	query.s4_H2	657	27,868	42.68	160.40
	aloha.10_H3	1,094	18,637	17.14	64.39
INFERENCE	mastermind_04_08	6,319	14,670	2.00	2.90
	fs-29	327,787	803,068	2.00	2.74

Table 1: Basic information for each benchmark problem.