Solutions for Hard and Soft Constraints Using Optimized Probabilistic Satisfiability

Marcelo Finger[1,2], Ronan Le Bras[1], Carla P. Gomes[1], Bart Selman[1]

[1]Department of Computer Science, Cornell University

[2]*On leave from:* Department of Computer Science
Institute of Mathematics and Statistics
University of Sao Paulo, Brazil

July 2013

Topics

Next Topic

Motivation

- Practical problems combine real-world hard constraints with soft constraints
- Soft constraints: preferences, uncertainties, flexible requirements
- We explore probabilistic logic as a mean of dealing with combined soft and hard constraints

Goals

- Aim: Combine **<u>L</u>ogic** and **<u>P</u>robabilistic** reasoning to deal with Hard (**L**) and Soft (**P**) constraints
- Method: develop optimized Probabilistic Satisfiability (**oPSAT**)
- Application: Demonstrate effectiveness on a real-world reasoning task in the domain of Materials Discovery.

An Example

Summer course enrollment

m students and k summer courses.

Potential team mates, to develop coursework. Constraints:

Hard Coursework to be done alone or in pairs.
Students must enroll in at least one and at most three courses.
There is a limit of ℓ students per course.

Soft Avoid having students with no teammate.

An Example

Summer course enrollment

m students and k summer courses.

Potential team mates, to develop coursework. Constraints:

Hard Coursework to be done alone or in pairs.
Students must enroll in at least one and at most three courses.
There is a limit of ℓ students per course.

Soft Avoid having students with no teammate.

In our framework: P(student with no team mate) "minimal" or bounded

Combining Logic and Probability

- Many proposals in the literature
  - Markov Logic Networks [Richardson & Domingos 2006]
  - Probabilistic Inductive Logic Prog [De Raedt *et. al* 2008]
  - Relational Models [Friedman *et al* 1999], etc

Combining Logic and Probability

- Many proposals in the literature
  - Markov Logic Networks [Richardson & Domingos 2006]
  - Probabilistic Inductive Logic Prog [De Raedt *et. al* 2008]
  - Relational Models [Friedman *et al* 1999], etc
- Our choice: **Probabilistic Satisfiability (PSAT)**
  - Natural extension of Boolean Logic
  - Desirable properties, e.g. respects Kolmogorov axioms
  - Probabilistic reasoning free of independence presuppositions

Combining Logic and Probability

- Many proposals in the literature
  - Markov Logic Networks [Richardson & Domingos 2006]
  - Probabilistic Inductive Logic Prog [De Raedt *et. al* 2008]
  - Relational Models [Friedman *et al* 1999], etc
- Our choice: **Probabilistic Satisfiability (PSAT)**
  - Natural extension of Boolean Logic
  - Desirable properties, e.g. respects Kolmogorov axioms
  - Probabilistic reasoning free of independence presuppositions
- What is PSAT?

Next Topic

Is PSAT a Zombie Idea?

An idea that refuses to die!

A Brief History of PSAT

- Proposed by [Boole 1854], *On the Laws of Thought*
- Rediscovered several times since Boole
  - De Finetti [1937, 1974], Good [1950], Smith [1961]
  - Studied by Hailperin [1965]
  - Nilsson [1986] (re)introduces PSAT to AI
  - PSAT is NP-complete [Georgakopoulos *et. al* 1988]
  - Nilsson [1993]: "complete impracticability" of PSAT computation
  - Many other works; see Hansen & Jaumard [2000]

Or a Wild Amazonian Flower?

Awaits special conditions to bloom!

(Linear programming + SAT-based techniques)

The Setting

- Formulas $\alpha_1, \ldots, \alpha_\ell$ over logical variables $\mathcal{P} = \{x_1, \ldots, x_n\}$
- Propositional valuation $v : \mathcal{P} \to \{0, 1\}$

The Setting

- Formulas $\alpha_1, \ldots, \alpha_\ell$ over logical variables $\mathcal{P} = \{x_1, \ldots, x_n\}$
- Propositional valuation $v : \mathcal{P} \to \{0, 1\}$
- A **probability distribution over propositional valuations**

$$\pi : V \to [0, 1]$$

$$\sum_{i=1}^{2^n} \pi(v_i) = 1$$

THE SETTING

- Formulas $\alpha_1, \ldots, \alpha_\ell$ over logical variables $\mathcal{P} = \{x_1, \ldots, x_n\}$
- Propositional valuation $v : \mathcal{P} \to \{0, 1\}$
- A **probability distribution over propositional valuations**

$$\pi : V \to [0, 1]$$

$$\sum_{i=1}^{2^n} \pi(v_i) = 1$$

- Probability of a formula α according to π

$$P_\pi(\alpha) = \sum \{\pi(v_i) | v_i(\alpha) = 1\}$$

The PSAT Problem

- Consider ℓ formulas $\alpha_1, \ldots, \alpha_\ell$ defined on n atoms $\{x_1, \ldots, x_n\}$
- A PSAT problem Σ is a set of ℓ restrictions

$$\Sigma = \{P(\alpha_i) \substack{\leq \\ = \\ \geq} p_i | 1 \leq i \leq \ell\}$$

- Probabilistic Satisfiability: is there a π that satisfies Σ?

THE PSAT PROBLEM

- Consider ℓ formulas $\alpha_1, \ldots, \alpha_\ell$ defined on n atoms $\{x_1, \ldots, x_n\}$
- A PSAT problem Σ is a set of ℓ restrictions

$$\Sigma = \{P(\alpha_i) \substack{\leq \\ = \\ \geq} p_i | 1 \leq i \leq \ell\}$$

- Probabilistic Satisfiability: is there a π that satisfies Σ?
- In our framework, $\ell = m + k$, $\Sigma = \Gamma \cup \Psi$:

HARD $\Gamma = \{\alpha_1, \ldots, \alpha_m\}$, $P(\alpha_i) = 1$ (clauses)
SOFT $\Psi = \{P(s_i) \leq p_i | 1 \leq i \leq k\}$

s_i atomic; p_i given, learned or minimized

Example continued

Only one course, three student enrollments: x, y and z
Potential partnerships: p_{xy} and p_{xz}, mutually exclusive.
Hard constraint

$$P(x \wedge y \wedge z \wedge \neg(p_{xy} \wedge p_{xz})) = 1$$

Soft constraints

$$\begin{array}{lcl} P(x \wedge \neg p_{xy} \wedge \neg p_{xz}) & \leq & 0.25 \\ P(y \wedge \neg p_{xy}) & \leq & 0.60 \\ P(z \wedge \neg p_{xz}) & \leq & 0.60 \end{array}$$

Example continued

Only one course, three student enrollments: x, y and z
Potential partnerships: p_{xy} and p_{xz}, mutually exclusive.
Hard constraint

$$P(x \wedge y \wedge z \wedge \neg(p_{xy} \wedge p_{xz})) = 1$$

Soft constraints

$$\begin{array}{lcl} P(x \wedge \neg p_{xy} \wedge \neg p_{xz}) & \leq & 0.25 \\ P(y \wedge \neg p_{xy}) & \leq & 0.60 \\ P(z \wedge \neg p_{xz}) & \leq & 0.60 \end{array}$$

(Small) solution: distribution π

$$\begin{array}{ll} \pi(x, y, z, \neg p_{xy}, \neg p_{xz}) = 0.1 & \pi(x, y, z, p_{xy}, \neg p_{xz}) = 0.4 \\ \pi(x, y, z, \neg p_{xy}, p_{xz}) = 0.5 & \pi(v) = 0 \text{ for other 29 valuations} \end{array}$$

Solving PSAT

- Algebraic formulation for $\Gamma(\bar{s}, \bar{x}) \cup \{P(s_i) = p_i | 1 \leq i \leq k\}$: find $A_{(k+1)\times 2^n}$ a $\{0,1\}$-matrix, $\pi_{2^n \times 1} \geq 0$ such that

$$A\pi = \begin{bmatrix} 1 \\ p \end{bmatrix}, \qquad \text{1st line: } \sum \pi_j = 1$$

if $\pi_j > 0$ then column A^j is Γ-consistent.

Solving PSAT

- Algebraic formulation for $\Gamma(\bar{s}, \bar{x}) \cup \{P(s_i) = p_i | 1 \leq i \leq k\}$: find $A_{(k+1)\times 2^n}$ a $\{0,1\}$-matrix, $\pi_{2^n \times 1} \geq 0$ such that

$$A\pi = \begin{bmatrix} 1 \\ p \end{bmatrix}, \qquad \text{1st line: } \sum \pi_j = 1$$

  if $\pi_j > 0$ then column A^j is Γ-consistent.
- Solved by linear program (exponentially sized)

$$\begin{array}{ll} \text{minimize} & c'\pi \\ \text{subject to} & A\pi = p \text{ and } \pi \geq 0 \end{array}$$

  c: cost vector, $c_j = 1$ if A^j is Γ-inconsistent; $c_j = 0$ otherwise
  Solution when $c'\pi = 0$ (may not be unique)

Solving PSAT

- Algebraic formulation for $\Gamma(\bar{s}, \bar{x}) \cup \{P(s_i) = p_i | 1 \leq i \leq k\}$:
  find $A_{(k+1)\times 2^n}$ a $\{0,1\}$-matrix, $\pi_{2^n \times 1} \geq 0$ such that

  $$A\pi = \begin{bmatrix} 1 \\ p \end{bmatrix}, \qquad \text{1st line: } \sum \pi_j = 1$$

  if $\pi_j > 0$ then column A^j is Γ-consistent.
- Solved by linear program (exponentially sized)

  $$\begin{array}{ll} \text{minimize} & c'\pi \\ \text{subject to} & A\pi = p \text{ and } \pi \geq 0 \end{array}$$

  c: cost vector, $c_j = 1$ if A^j is Γ-inconsistent; $c_j = 0$ otherwise
  Solution when $c'\pi = 0$ (may not be unique)
- **Theorem**: $\Gamma \cup \{P(s_i) = p_i | 1 \leq i \leq k\}$ is P-satisfiable $\Longrightarrow$ there is π with at most $k+1$ values $\pi_j > 0$ (PSAT is NP-complete)

SAT-based column generation

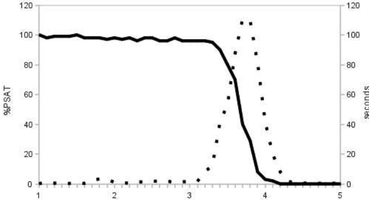
- Goal: implicit representation of exponential-sized system
- Simplex algorithm: at each iteration i, store $A^{(i)}_{(k+1)\times(k+1)}$.

  Compute new column $s^{(i)}$ with a SAT-formula $\Gamma \cup \Delta$ such that:
  - Column generated $s^{(i)}$ is Γ-consistent
  - Cost does not increase: inequality over $\{0, 1\}$-variables, converted to a SAT formula Δ
- PSAT instance is P-unsat if $\Gamma \cup \Delta$ is unsat.

SAT-based column generation

- Goal: implicit representation of exponential-sized system
- Simplex algorithm: at each iteration i, store $A^{(i)}_{(k+1)\times(k+1)}$.

  Compute new column $s^{(i)}$ with a SAT-formula $\Gamma \cup \Delta$ such that:
  - Column generated $s^{(i)}$ is Γ-consistent
  - Cost does not increase: inequality over $\{0,1\}$-variables, converted to a SAT formula Δ
- PSAT instance is P-unsat if $\Gamma \cup \Delta$ is unsat.
- PSAT Phase-transition [Finger & De Bona 2011]



The Interface Between Logic/SAT and Linear Algebra

- From the simplex method, *reduced cost* of inserting a column $s = [1\ s_1 \dots s_k]'$ into A:

$$c_s - c_A' A^{-1} s \leq 0$$

- c_s is the cost of the new column; $c_s = 0$ for PSAT.
- c_A: (column) vector of costs of the columns of A
- Use [Warners 98] method to convert inequality to SAT formula Δ
- Also, s must be $\Gamma(s; x)$-consistent
- SAT-solver: obtain v s.t. $v(\Gamma \cup \Delta) = 1$
- $v(s)$ is the new column. Apply *simplex merge* to insert $v(s)$ in A, generating A^* s. t. $A^* \pi^* = p, \pi^* \geq 0$

Example of PSAT solution

Add variables for each soft violation: s_x, s_y, s_z.

$$\Gamma = \left\{ \begin{array}{c} x,\ y,\ z,\ \neg p_{xy} \vee \neg p_{xz}, \\ (x \wedge \neg p_{xy} \wedge \neg p_{xz}) \rightarrow s_x,\ (y \wedge \neg p_{xy}) \rightarrow s_y,\ (z \wedge \neg p_{xz}) \rightarrow s_z \end{array} \right\}$$

$$\Psi = \{\ P(s_x) = 0.25,\ P(s_y) = 0.6,\ P(s_z) = 0.6\ \}$$

Iteration 0:

$$\begin{array}{c} \\ s_x \\ s_y \\ s_z \end{array} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.4 \\ 0 \\ 0.35 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.25 \\ 0.60 \\ 0.60 \end{bmatrix}$$

$$\begin{aligned} \mathrm{cost}^{(0)} &= 0.4 \\ b^{(0)} &= [1\ 0\ 1\ 0]' : \mathrm{col}\ 3 \end{aligned}$$

Example of PSAT solution

Add variables for each soft violation: s_x, s_y, s_z.

$$\Gamma = \left\{ \begin{array}{c} x,\ y,\ z,\ \neg p_{xy} \vee \neg p_{xz}, \\ (x \wedge \neg p_{xy} \wedge \neg p_{xz}) \to s_x,\ (y \wedge \neg p_{xy}) \to s_y,\ (z \wedge \neg p_{xz}) \to s_z \end{array} \right\}$$

$$\Psi = \{\ P(s_x) = 0.25,\ P(s_y) = 0.6,\ P(s_z) = 0.6\ \}$$

Iteration 1:

$$\begin{array}{c} \\ s_x \\ s_y \\ s_z \end{array} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.05 \\ 0.35 \\ 0.35 \\ 0.25 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.25 \\ 0.60 \\ 0.60 \end{bmatrix}$$

$$\begin{aligned} \mathrm{cost}^{(1)} &= 0.05 \\ b^{(1)} &= [1\ 1\ 0\ 1]' : \mathrm{col}\ 1 \end{aligned}$$

Example of PSAT solution

Add variables for each soft violation: s_x, s_y, s_z.

$$\Gamma = \left\{ \begin{array}{c} x,\ y,\ z,\ \neg p_{xy} \vee \neg p_{xz}, \\ (x \wedge \neg p_{xy} \wedge \neg p_{xz}) \to s_x,\ (y \wedge \neg p_{xy}) \to s_y,\ (z \wedge \neg p_{xz}) \to s_z \end{array} \right\}$$

$$\Psi = \{\ P(s_x) = 0.25,\ P(s_y) = 0.6,\ P(s_z) = 0.6\ \}$$

Iteration 2:

$$\begin{array}{c} \\ s_x \\ s_y \\ s_z \end{array} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} 0.05 \\ 0.35 \\ 0.40 \\ 0.20 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.25 \\ 0.60 \\ 0.60 \end{bmatrix}$$

$$\text{cost}^{(2)} = 0$$

Next Topic

Optimizing PSAT solutions

- Solutions to PSAT are not unique
- First optimization phase: determines if constraints are solvable
- Second optimization phase to obtain a distribution with desirable properties.
- A different objective (cost) function

Minimizing Expected Violations

- Idea: minimize the expected number of soft constraints violated by each valuation, $S(v)$

$$E(S) = \sum_{v_i | v_i(\Gamma)=1} S(v_i)\pi(v_i)$$

- **Theorem:** Every linear function of a model (valuation) has constant expected value for any PSAT solution
- In particular, $E(S)$ is constant, no point in minimizing it
- Any other **model linear function** is not a candidate for minimization

Minimizing Variance

- Idea: penalize high S, minimize $E(S^2)$
- Lemma: The distribution that minimizes $E(S^2)$ also minimizes variance, $\mathrm{Var}(S) = E((S - E(S))^2)$
- **oPSAT** is a second phase minimization whose objective function is $E(S^2)$
- Problem: computing a SAT formula that decreases cost is harder than in PSAT
- oPSAT needs a more elaborate interface logic/linear algebra

oPSAT Cost Minimization Strategy

- Reduced cost: $c_s - c_A A^{-1} s \leq 0$
- In PSAT, $c_s = 0$
- In oPSAT, $c_s \in \{0, 1, 4, \ldots, k^2, (k+1)^2\}$
- Strategy: $k + 2$ iterations of optimization, one for each possible number of soft violations
- At the end, we obtain a distribution with minimal variance

oPSAT Cost Minimization Strategy

- Reduced cost: $c_s - c_A A^{-1} s \leq 0$
- In PSAT, $c_s = 0$
- In oPSAT, $c_s \in \{0, 1, 4, \ldots, k^2, (k+1)^2\}$
- Strategy: $k + 2$ iterations of optimization, one for each possible number of soft violations
- At the end, we obtain a distribution with minimal variance

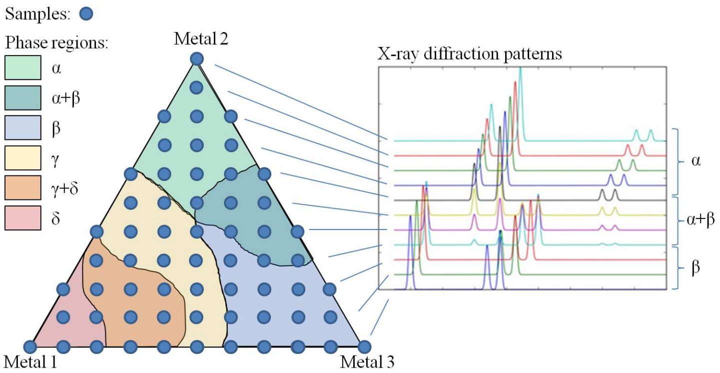
$$
\begin{array}{c} \\ s_x \\ s_y \\ s_z \end{array}
\begin{bmatrix} 1 & 1 & \mathbf{1} & 1 \\ 1 & 0 & \mathbf{0} & 0 \\ 0 & 0 & \mathbf{1} & 1 \\ 1 & 1 & \mathbf{0} & 1 \end{bmatrix}
\cdot
\begin{bmatrix} 0.25 \\ 0.15 \\ 0.40 \\ 0.20 \end{bmatrix}
=
\begin{bmatrix} 1 \\ 0.25 \\ 0.60 \\ 0.60 \end{bmatrix}
$$

$$E(S^2) = 2.35$$

If one valuation has to be chosen, choose one with maximal probability

Next Topic

Problem Definition



Problem Modeling

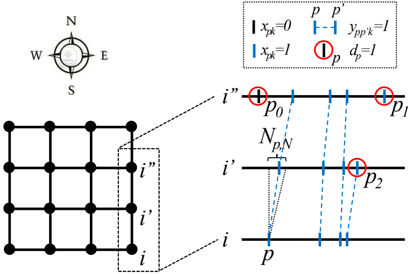
Find an association of **peak angles to phases**, respecting structural constraints, such that the **probability of sample point defect is limited**.

Some structural constraints:

- At most 3 phases per sample point i
- Shifting factors between peaks in neighbor sample points must be within $[S_{min}, S_{max}]$: potential edges
- 2 directions in Grid: NS, EW; at most one edge in each direction; connected peaks in same phase
- Peak with no edge = *peak defect*. Sample point with peak defect = **point defect** (soft violation)
- Sample points are embedded into a connected graph $\mathcal{G}_K$ per phases K

Variables in oPSAT Encoding

- $x_{p,k}$, peak p belongs to phase k
- $z_{i,k}$, sample point i has a peak in phase k, $z_{i,k} = \bigvee_{p \in G(i)} x_{p,k}$
- $y_{pp'k}$, p is paired with p' in k, $y_{pp'k} \rightarrow x_{pk} \wedge x_{p'k}$
- Shift direction of phase k: $D_{1k} \in \{N, S\}$ and $D_{2k} \in \{E, W\}$
- d_p, peak p is not paired to any other peak, **defect**.
- d_i, sample point containing defect, **soft constraint**



Formulas in oPSAT Encoding

- A peak is assigned to at most one phase, $\sum_k x_{pk} \leq 1$
- An unassigned peak is considered unmatched, $(\bigvee_k x_{pk}) \vee d_p$
- Non-defective peaks are paired with a neighboring peak, $x_{pk} \rightarrow \left(\bigvee_{p'} y_{pp'k}\right)$
- If two adjacent samples share a phase, each peak of one must be paired with a peak of the other.
- Relaxed form of convex connectivity: if any two samples involve a given phase, there should be a sample in between them that involves this phase as well.

Implementation

- Implementated in C++
- Linear Solver uses *blas* and *lapack*
- SAT-solver: minisat
- PSAT formula (DIMACS extension) generated by C++ formula generator
- $P(d_p) \leq 2\% \Longrightarrow P(d_i) \leq 1 - (1 - P(d_p))^{L_i}$
- Input: Peaks at sample point
- Output: oPSAT most probable model
- Compare with SMT implementation in SAT 2012
- **psat.sourceforge.net**

Experimental Results

Dataset					SMT	oPSAT	
System	*P*	*L**	*K*	*#Peaks*	**Time**(s)	**Time**(s)	**Accuracy**
Al/Li/Fe	28	6	6	170	346	5.3	84.7%
Al/Li/Fe	28	8	6	424	10076	8.8	90.5%
Al/Li/Fe	28	10	6	530	28170	12.6	83.0%
Al/Li/Fe	45	7	6	651	18882	121.1	82.0%
Al/Li/Fe	45	8	6	744	46816	128.0	80.3%

- The accuracy of SMT is 100%
- P: n. of sample points; L^*: the average n, of peaks per phase
- K: n. of basis patterns; *#Peaks*: overall n. of peaks
- #Aux variables $> 10\,000$

Next Topic

Conclusions and the Future

- oPSAT can be effectively implemented to deal with hard and soft constraints
- Can be successfully applied to non-trivial problems of materials discovery with acceptable precision and superior run times than existing methods
- Other forms of logic-probabilistic inference are under investigation