Complexity

There are several ways to compare the effectiveness of calculi. One is by implementation and benchmarking. One is by simulation, calculus *A* can simulate calculus *Y*, e.g., Resolution can sumulate CDCL. Another one is by best case proof length. It is shown that calculus *X* produces shorter proofs than calculus *Y* on some class of formulas. This is the topic of this chapter.

2.14.1 Definition (Exponential Proof Length for Analytic Calculi)

Let $\text{EPL}_1 = \{P_1, \neg P_1\}$. Then EPL_{i+1} is inductively defined by $\text{EPL}_{i+1} = \{P_{2^i+j} \vee C_j, \neg P_{2^i+j} \vee C_j \mid C_j \in \text{EPL}_i\}$, where $\text{EPL}_i = \{C_0, \ldots, C_{2^i-1}\}$.

Note that the number of clauses as well as the number of propositional variables grows exponentially in i for the EPL_i clause sets: $|\text{EPL}_i| = 2^i$ and $|\{P_i \mid P_i \in C_j, C_j \in \text{EPL}_i\}| = 2^i - 1$. For example, $\text{EPL}_2 = \{P_2 \vee P_1, \neg P_2 \vee P_1, P_3 \vee \neg P_1, \neg P_3 \vee \neg P_1\}$.

2.14.2 Theorem (EPL Proof Length *[CookReckhow79]*)

The minimal proof length of Tableau and DPLL is exponential in $|\text{EPL}_i|$, whereas the minimal proof length of Resolution and CDCL is linear in $|\text{EPL}_i|$.

2.14.3 Definition (Pigeon Hole Formulas $\mathrm{ph}(n)$)

For some given n and propositional variables $P_{i,j}$, where $1 \leq j \leq n$, $1 \leq i \leq n+1$, the corresponding pigeon hole formula (clause set) $\mathrm{ph}(n)$ is

$$\mathrm{ph}(n) = \bigwedge_{1\leq i\leq n+1} P_{i,1} \vee \ldots \vee P_{i,n} \quad \wedge \quad \bigwedge_{1\leq j\leq n} \quad \bigwedge_{\substack{1\leq i,k\leq n+1 \\ i<k}} \neg P_{i,j} \vee \neg P_{k,j}$$

The intuition behind a variable $P_{i,j}$ is that it is true iff pigeon i sits in hole j. Then the formulas $P_{i,1} \vee \ldots \vee P_{i,n}$ express that every pigeon has to sit in some hole and the formulas $\neg P_{i,j} \vee \neg P_{k,j}$ that a hole can host at most one pigeon. Since there is one more pigeon than holes, the formula is unsatisfiable.

Note that the number of clauses of a pigeon hole formula $\mathrm{ph}(n)$ grows cubic in n. The famous theorem on the pigeon whole formulas says that any resolution proof showing unsatisfiability of $\mathrm{ph}(n)$ has a length at least exponential in n, i.e., no resolution-based system can efficiently show unsatisfiability of a pigeon hole formula.

2.14.4 Theorem (Pigeon Hole Proof Length *[Haken85]*)

The length of any resolution refutation of $\mathrm{ph}(n)$ is exponential in n.

Computing Cost Optimal Models (OCDCL)

OCDCL States

$(\epsilon; N; \emptyset; 0; \top; \epsilon)$ start state for some clause set N

$(M; N; U; k; \bot; O)$ final state, where

- N has no model if $O = \epsilon$
- O is a cost optimal model if $O \neq \epsilon$

$(M; N; U; k; \top; O)$ intermediate model search state

$(M; N; U; k; D; O)$ backtracking state if $D \notin \{\top, \bot\}$

- O denotes the cost optimal model of N
- M, N, U, k, D are defined analogously to CDCL
- but OCDCL always terminates with $D = \bot$

OCDCL Rules

Propagate $(M; N; U; k; \top; O) \Rightarrow_{\text{OCDCL}} (ML^{C\vee L}; N; U; k; \top; O)$

provided $C \vee L \in (N \cup U)$, $M \models \neg C$, L is undefined in M

Decide $(M; N; U; k; \top; O) \Rightarrow_{\text{OCDCL}}$
$(ML^{k+1}; N; U; k+1; \top; O)$

provided L is undefined in M, contained in N

ConflSat $(M; N; U; k; \top; O) \Rightarrow_{\text{OCDCL}} (M; N; U; k; D; O)$

provided $D \in (N \cup U)$ and $M \models \neg D$

ConflOpt $(M; N; U; k; \top; O) \Rightarrow_{\text{OCDCL}} (M; N; U; k; \neg M; O)$

provided $O \neq \epsilon$ and $\text{cost}(M) \geq \text{cost}(O)$

OCDCL Rules (ctd.)

Skip $(ML^{C\lor L}; N; U; k; D; O) \Rightarrow_{\text{OCDCL}} (M; N; U; k; D; O)$

provided $D \notin \{\top, \bot\}$ and $\text{comp}(L)$ does not occur in D

Resolve $(ML^{C\lor L}; N; U; k; D \lor \text{comp}(L); O) \Rightarrow_{\text{OCDCL}}$
$(M; N; U; k; D \lor C; O)$

provided D is of level k

Backtrack $(M_1 K^{i+1} M_2; N; U; k; D \lor L; O) \Rightarrow_{\text{OCDCL}}$
$(M_1 L^{D\lor L}; N; U \cup \{D \lor L\}; i; \top; O)$

provided L is of level k and D is of level i

Improve $(M; N; U; k; \top; O) \Rightarrow_{\text{OCDCL}} (M; N; U; k; \top; M)$

provided $M \models N$, M is total, i.e., contains all atoms in N, and $O = \epsilon$ or $\text{cost}(M) < \text{cost}(O)$

?? Definition (Reasonable OCDCL Strategy)

An OCDCL strategy is *reasonable* if ConflSat is preferred over ConflOpt is preferred over Improve is preferred over Propagate which is preferred over the remaining rules.

2.15.3 Proposition (OCDCL Basic Properties)

Consider an OCDCL state $(M; N; U; k; D'; O)$ derived by a reasonable strategy from start state $(\epsilon, N, \emptyset, 0, \top, \epsilon)$. Then the following properties hold:

1. M is consistent.
2. If $O \neq \epsilon$ then O is consistent and $O \models N$.
3. If $D' \notin \{\top, \bot\}$ then $M \models \neg D'$.
4. If $D' \notin \{\top, \bot\}$ then (i) D' is entailed by $N \cup U$, or (ii) for any model $M' \models \{\neg D'\} \cup N \cup U$: $\text{cost}(M') \geq \text{cost}(O)$.
5. If $D' = \top$ and M contains only propagated literals then for each valuation $\mathcal{A}$ with $\mathcal{A} \models (N \cup U)$ it holds $\mathcal{A} \models M$.

2.15.3 Proposition (OCDCL Basic Properties (ctd.))

6. For all models M with $M \models N$: if $O = \epsilon$ or $\text{cost}(M) < \text{cost}(O)$ then $M \models (N \cup U)$.
7. If $D' = \bot$ then OCDCL terminates and there is no model M' with $M' \models N$ and $\text{cost}(M') < \text{cost}(O)$.
8. Each infinite derivation

$$(\epsilon; N; \emptyset; 0; \top; \epsilon) \Rightarrow_{\text{OCDCL}} (M_1; N; U_1; k_1; D_1; O_1) \Rightarrow_{\text{OCDCL}} \cdots$$

   contains an infinite number of Backtrack applications.
9. OCDCL never learns the same clause twice.

2.15.4 Lemma (OCDCL Normal Forms)

The OCDCL calculus with a reasonable strategy has only 2 normal forms:

- $(M; N; U; 0; \bot; O)$ where $O \neq \epsilon$, $O \models N$ and $\text{cost}(O)$ is optimal
- $(M; N; U; 0; \bot; \epsilon)$ where N is unsatisfiable

2.15.5 Lemma (OCDCL Termination)

OCDCL with a reasonable strategy terminates in a state $(M; N; U; 0; \bot; O)$.

2.15.6 Theorem (OCDCL Correctness)

OCDCL with a reasonable strategy starting from a state $(\epsilon; N; \emptyset; 0; \top; \epsilon)$ terminates in a state $(M; N; U; 0; \bot; O)$. If $O = \epsilon$ then N is unsatisfiable. If $O \neq \epsilon$ then $O \models N$ and for any other model M' with $M' \models N$ it holds $\mathrm{cost}(M') \geq \mathrm{cost}(O)$.

Improving OCDCL

Prune $(M; N; U; k; \top; O) \Rightarrow_{\text{OCDCL}} (M; N; U; k; \neg M; O)$
provided for all total trail extensions MM' of M it holds
$\text{cost}(MM') \geq \text{cost}(O)$

ConflOpt $(M; N; U; k; \top; O) \Rightarrow_{\text{OCDCL}} (M; N; U; k; \neg M; O)$
provided $O \neq \epsilon$ and $\text{cost}(M) \geq \text{cost}(O)$

Improving OCDCL

Prune $(M; N; U; k; \top; O) \Rightarrow_{\text{OCDCL}} (M; N; U; k; \neg M; O)$

provided for all total trail extensions MM' of M it holds $\text{cost}(MM') \geq \text{cost}(O)$

ConflOpt $(M; N; U; k; \top; O) \Rightarrow_{\text{OCDCL}} (M; N; U; k; \neg M; O)$

provided $O \neq \epsilon$ and $\text{cost}(M) \geq \text{cost}(O)$

The Max-SAT Problem

Given $N = N_H \uplus N_S$ where N_H are hard clauses
and N_S are soft clauses

Find $\mathcal{A} \models N_H$ with minimal cost $\sum_{\mathcal{A} \models \neg C}^{C \in N_S} \omega(C)$
where $\omega \colon N_S \mapsto \mathbb{R}^+$

Reducing Max-SAT to OCDCL

1. Introduce a fresh variable S_i for each $C_i \in N_S = \{C_1, \ldots, C_n\}$
2. Define $N'_S = \{S_i \vee C_i \mid C_i \in N_S\}$
3. Compute cost optimal model for $N' = N_H \uplus N'_S$ with cost function $\text{cost}(L) = \begin{cases} \omega(C_i) & \text{if } L = S_i \\ 0 & \text{otherwise} \end{cases}$

2.15.7 Theorem (Max-SAT Solution)

$\mathcal{A}$ is a Max-SAT solution for $N = N_H \uplus N_S$ with minimal value $c = \sum_{\mathcal{A} \models \neg C}^{C \in N_S} \omega(C)$ iff $(\epsilon; N'; \emptyset; 0; \top; \epsilon) \Rightarrow^*_{\text{OCDCL}} (M; N'; U; k; \bot; O)$ with a reasonable strategy where $N' = N_H \uplus N'_S$, and $\mathrm{cost}(O) = c$.

Optimization

1. Introduce a fresh variable S_i for each $C_i \in N_S = \{C_1, \ldots, C_n\}$
2. Define $N'_S = \{S_i \vee C_i \mid C_i \in N_S\} \cup \{\neg C_i \vee \neg S_i \mid C_i \in N_S\}$
3. Compute cost optimal model for $N' = N_H \uplus N'_S$ with cost function $\mathrm{cost}(L) = \begin{cases} \omega(C_i) & \text{if } L = S_i \\ 0 & \text{otherwise} \end{cases}$

Minimal Covering Models

Given $\mathcal{M}$ set of all models of the set of clauses N

Find $\mathcal{M}' \subseteq \mathcal{M}$ such that

- $|\mathcal{M}'|$ is minimal
- for each propositional variable P in N there is a model $M \in \mathcal{M}'$ with $M(P) = 1$

Reduction to OCDCL

Given N with variables $P_1, \ldots P_n$ and clauses $C_1, \ldots, C_m$

1. Define $N_j := \{C\{P_i \mapsto P_i^j \mid 1 \leq i \leq n\} \vee \neg Q_j \mid C \in N\}$
2. Define $N_+ := \{P_i^1 \vee \ldots \vee P_i^n \mid 1 \leq i \leq n\}$
3. Define $N_Q := \{\neg P_i^j \vee Q_j \mid 1 \leq i, j \leq n\}$
4. Find a minimal cost model of $(\cup_{j=1}^n N_j) \cup N_+ \cup N_Q$ with cost function $\mathrm{cost}(M) = \sum_{j=1}^n M(Q_j)$

Requires

- $O(n^2)$ additional variables
- $O(n \cdot \max(m, n))$ additional clauses

Note: n = upper bound of number of models (Algorithm 10)

Reduction to OCDCL

Given N with variables $P_1, \ldots P_n$ and clauses $C_1, \ldots, C_m$

1. Define $N_j := \{C\{P_i \mapsto P_i^j \mid 1 \leq i \leq n\} \vee \neg Q_j \mid C \in N\}$
2. Define $N_+ := \{P_i^1 \vee \ldots \vee P_i^n \mid 1 \leq i \leq n\}$
3. Define $N_Q := \{\neg P_i^j \vee Q_j \mid 1 \leq i, j \leq n\}$
4. Find a minimal cost model of $(\cup_{j=1}^n N_j) \cup N_+ \cup N_Q$ with cost function $\mathrm{cost}(M) = \sum_{j=1}^n M(Q_j)$

Requires

- $O(n^2)$ additional variables
- $O(n \cdot \max(m, n))$ additional clauses

Note: n = upper bound of number of models (Algorithm 10)

Reduction to OCDCL

Given N with variables $P_1, \ldots P_n$ and clauses $C_1, \ldots, C_m$

1. Define $N_j := \{C\{P_i \mapsto P_i^j \mid 1 \leq i \leq n\} \vee \neg Q_j \mid C \in N\}$
2. Define $N_+ := \{P_i^1 \vee \ldots \vee P_i^n \mid 1 \leq i \leq n\}$
3. Define $N_Q := \{\neg P_i^j \vee Q_j \mid 1 \leq i, j \leq n\}$
4. Find a minimal cost model of $(\cup_{j=1}^n N_j) \cup N_+ \cup N_Q$ with cost function $\text{cost}(M) = \sum_{j=1}^n M(Q_j)$

Requires

- $O(n^2)$ additional variables
- $O(n \cdot \max(m, n))$ additional clauses

Note: n = upper bound of number of models (Algorithm 10)