EXPECTED PERFORMANCE OF BRANCHING RULES

Vašek Chvátal
Canada Research Chair in Combinatorial Optimization
Department of Computer Science and Software Engineering
Concordia University
Montréal, QC, Canada

First version written on 20 September 2006
Last updated on 28 September 2006

Problem SAT

A *truth assignment* is a mapping f that assigns 0 (interpreted as "false") or 1 (interpreted as "true") to each variable in its domain; we shall enumerate all the variables in this domain as $x_1, \ldots, x_n$. The *complement* $\overline{x}_i$ of each such variable x_i is defined by $f(\overline{x}_i) = 1 - f(x_i)$ for all truth assignments f; both x_i and $\overline{x}_i$ are called *literals*; if $u = \overline{x}_i$ then $\overline{u} = x_i$. A *clause* is a set of (distinct) literals and a *formula* (in a conjunctive normal form) is a family of (not necessarily distinct) clauses. A truth assignment *satisfies* a clause if it maps at least one of its literals to 1; the assignment *satisfies* a formula if and only if it satisfies each of its clauses. A formula is called *satisfiable* if it is satisfied by at least one truth assignment; otherwise it is called *unsatisfiable*. The problem of recognizing satisfiable formulas is known as the *satisfiability problem*, or SAT for short.

Solving SAT by implicit enumeration

Given a formula $\mathcal{F}$ and a literal v in $\mathcal{F}$, we let $\mathcal{F}|v$ denote the "residual formula" arising from $\mathcal{F}$ when $f(v)$ is set at 1: explicitly, this formula is obtained from $\mathcal{F}$ by

- removing all the clauses that contain v,
- deleting $\overline{v}$ from all the clauses that contain $\overline{v}$,
- removing both v and $\overline{v}$ from the list of literals.

Trivially, $\mathcal{F}$ is satisfiable if and only if at least one of $\mathcal{F}|v$ and $\mathcal{F}|\overline{v}$ is satisfiable. This observation leads to a recursive algorithm that, given any formula $\mathcal{F}$, returns either the message `SAT` or the message `UNSAT`; this algorithm, Implicit-Enumeration, is defined in Figure 1.

if $\mathcal{F} = \emptyset$ **then** return `SAT` **end**
if $\mathcal{F}$ includes the empty clause **then** return `UNSAT` **end**
choose a literal v of $\mathcal{F}$;
if Implicit-Enumeration($\mathcal{F}|v$) =`SAT` **then** return `SAT` **end**
if Implicit-Enumeration($\mathcal{F}|\overline{v}$) =`SAT` **then** return `SAT` **end**
return `UNSAT`;

Figure 1: Implicit-Enumeration($\mathcal{F}$)

Optimal branching rules

Efficiency of Implicit-Enumeration depends on its implementation of the instruction

choose a literal v of $\mathcal{F}$;

implementations of this instruction are called *branching rules*. For each formula $\mathcal{F}$, let Opt($\mathcal{F}$) denote the number of leaves in the recursion tree of Implicit-Enumeration($\mathcal{F}$) minimized over all branching rules.

Random formulas

We will consider the "fixed-clause-length model" of randomly generating formulas in a conjunctive normal form: A clause C is called *ordinary* if there is no variable x such that $x \in C$ and $\overline{x} \in C$; in a *random formula of m clauses of length k over n variables,* clauses are independent random variables $C_1, \ldots, C_m$ such that each C_i is distributed uniformly over all the $\binom{n}{k}2^k$ ordinary clauses of size k with variables coming from a fixed set of size n.

For every integer k greater than 1, there are positive constants a_k and b_k such that, with probability $1 - o(1)$ as $n \to \infty$, a random formula with $(1 + o(1))cn$ clauses of size k over n variables is satisfiable whenever $c < a_k$ and unsatisfiable whenever $c > b_k$. The existence of a_k proportional to $2^k/k$ was proved in the special case $4 \leq k \leq 40$ by Chao and Franco [3] and without this restriction by Chvátal and Reed [5]; Achlioptas and Peres [1] established $a_k = 2^k \ln 2 - O(k)$ as $k \to \infty$. The existence of b_k equal to $2^k \ln 2$ is an observation apparently made first by Franco and Paull [6].

Expected performance of optimal branching rules

For every choice of positive integers k, n, m, let $f_k(n, m)$ denote the expected value of $\textsc{Opt}(\mathcal{F})$ in a random formula of m clauses of length k over n variables.

Theorem 1 *For every choice of an integer k and a real number c such that $k \geq 3$ and $c > b_k$, we have*

$$(1+\varepsilon)^n \;<\; f_k(n, (1+o(1))cn) \;<\; 2^{(1+o(1))\gamma n}$$

for some positive ε and for

$$\gamma = \frac{k-1}{k}\left(\frac{2^k \ln 2}{ck}\right)^{1/(k-1)}. \tag{1}$$

For instance, if $k = 3$ and $c = 6$, then $\gamma = \frac{4}{9}\sqrt{\ln 2} \approx 0.37002$, and so

$$f_3(n, 6n) \;<\; 1.29238^n \quad \text{for all sufficiently large } n.$$

The lower bound in Theorem 1 follows from a theorem of Chvátal and Szemerédi [4], whose proof gives a procedure for computing values of ε as a function of k and c. (These are nothing to write home about.)

Proof of the upper bound: Given k, m, n, set

$$\beta = \frac{k-1}{k}\left(\frac{2^k \ln 2}{k} \cdot \frac{n}{m}\right)^{1/(k-1)};$$

next, given a random formula of m clauses of length k over n variables, branch at random (that is, choose each new v from the uniform distribution on the literals whose truth values have not been set yet) and, for each $s = 0, 1, \ldots, n$, let $\tau(s)$ denote the expected number of internal nodes of the recursion tree on level s. In this notation,

$$f_k(n, m) = 1 + \sum_{s=0}^{n} \tau(s)$$

and

$$\tau(s) = 2^s \left(1 - \frac{\binom{s}{k}}{\binom{n}{k}} (\tfrac{1}{2})^k \right)^m.$$

In particular, $\tau(s) = 2^s$ whenever $s < k$; if $s \geq k$, then

$$\frac{\binom{s}{k}}{\binom{n}{k}} > \left(\frac{s-k+1}{s}\right)^k \cdot \left(\frac{s}{n}\right)^k = (1-\delta(s))\left(\frac{s}{n}\right)^k$$

with

$$\delta(s) = 1 - \left(1 - \tfrac{k-1}{s})\right)^k$$

and so

$$\begin{aligned}\tau(s) &< 2^s\left(1-(1-\delta(s))\left(\frac{s}{2n}\right)^k\right)^m \\ &< \exp\left(\left(\frac{s}{n} - \frac{(1-\delta(s))}{2^k \ln 2}\cdot\frac{m}{n}\cdot\left(\frac{s}{n}\right)^k\right)\cdot n\ln 2\right) \\ &\leq 2^{(1+\varepsilon(s))\beta n}\end{aligned}$$

with

$$\varepsilon(s) = \left(\frac{1}{1-\delta(s)}\right)^{1/(k-1)} - 1 \ .$$

Since $\varepsilon(s)$ is a decreasing function of s, we conclude that

$$f_k(n,m) < \sum_{s<\beta n/2} 2^s + n \cdot 2^{(1+\varepsilon(\beta n/2))\beta n},$$

which completes the proof.

Since branching at random on a random formula $\mathcal{F}$ is equivalent to branching on $\mathcal{F}$ in any deterministic way that disregards all information about $\mathcal{F}$ (for instance, always branching on the candidate x_i that has the smallest subscript i), the upper bound applies to any such branching rule. Significant results on almost sure performance of such branching rules on random formulas have been obtained by Beame, Karp, Pitassi, and Saks [2]. In particular, their Theorem 6.3 implies that, for every choice of an integer k and a function m of n such that $k \geq 3$ and $m/n > b_k$, we have

$$f_k(n,m) \ < \ 2^{O(\alpha n)} n^{O(1)} \quad \text{with} \quad \alpha = \left(\frac{n}{m}\right)^{1/(k-2)} .$$

Problem: Improve the bounds on $f_k(n, (1+o(1))cn)$.

References

[1] D. Achlioptas and Y. Peres, "The threshold for random k-SAT is $2^k \log 2 - O(k)$", *J. Amer. Math. Soc.* **17** (2004), 947 – 973.

[2] P. Beame, R. Karp, T. Pitassi, and M. Saks, "The efficiency of resolution and Davis-Putnam procedures", *SIAM J. Comput.* **31** (2002), 1048 – 1075.

[3] M.-T. Chao and J. Franco, "Probabilistic analysis of a generalization of the unit-clause literal section heuristic for the k-satisfiability problem," *Inform. Sci.* **51** (1990), 289 – 314.

[4] V. Chvátal and E. Szemerédi, "Many hard examples for resolution", *J. Assoc. Comput. Mach.* **35** (1988), 759 – 768.

[5] V. Chvátal and B. Reed, "Mick gets some (the odds are on his side)", in: *Proceedings of the 33rd Annual Symposium on Foundations of Computer Science,* IEEE Computer Society Press, Washington, 1992, pp. 620 – 627.

[6] J. Franco and M. Paull, "Probabilistic analysis of the Davis-Putnam procedure for solving the satisfiability problem," *Discrete Appl. Math.*, **5** (1983), 77 – 87.