Sharp Thresholds for ‘non-trivial’ Satisfiability of k-Horn Formulae

Paul E. Dunne & Trevor J.M. Bench-Capon
Department of Computer Science
University of Liverpool
Liverpool L69 7ZF
Great Britain

ped@uk.ac.liv.csc
http://www.csc.liv.ac.uk/≈ped/

Forsan et haec olim meminisse iuvabit

We consider the ‘non-trivial’ satisfiability problem on *Horn*-clause formulae, having exactly k literals per clause.

An instance of k-*HSAT* is a k-*CNF* formula, F, in which every clause contains *at most one* positive literal.

k-*HSAT*(F) is **true** if there is an assignment, α such that

a) $F(\alpha) = 1$, (α satisfies F)
b) *At least* $k - 1$ variables are set to 1 by α.

F is *useful* if k-*HSAT*(F).

α is *non-trivial* for F if (a) and (b) hold.

(**Note:** k-*HSAT* $\in P$).

Motivation

i) Horn-clause formulae are a standard representation in Knowledge-based systems (so-called ‘expert’ systems).

ii) ‘Satisfiability’ (of an unrestricted Horn formula) is equivalent to ‘consistency’ of the set of rules defining the clauses.

Non-trivial Assignments

The ‘pure’ satisfiability question for k-*HSAT* is trivial: any instantiation with at most $k-2$ variables set to 1 is a satisfying assignment.

Such cases are, however, uninteresting in gauging the utility of a set of rules, since only the (obvious) conclusion that the model in which no rule is activated is consistent.

Requiring at least $k-1$ variables to be 1 tests whether the set of rules is consistent when *at least one* rule is activated.

Random Instances of k-HSAT

$HC(n, k)$ is the set of all clauses over n variables containing exactly k literals at most one of which is un-negated.

F_p, is a random instance of k-$HSAT$ constructed by including each clause in $HC(n, k)$ with probability p: all such decisions being made independently.

Properties of random formulae in this model translate to the model where F is chosen uniformly at random from the set of all instances with exactly $p\,|HC(n, k)|$ clauses.

We denote by $\mu_p(k - HSAT_n)$ the probability that the n-variable random formula F_p is useful, i.e. has a non-trivial satisfying assignment.

Theorem: $\forall\ k \geq 2$

Let $p(n) = \dfrac{\log_e n}{n}$.

$$\lim_{n\to\infty} \mu_{\alpha p(n)}(k - HSAT_n) = 1 \quad \alpha < k - 1$$

$$\lim_{n\to\infty} \mu_{\alpha p(n)}(k - HSAT_n) = 0 \quad \alpha > k - 1$$

Corollary: Let $m(n,k) = n^{k-1} \log_e n$. If $F_{cm(n,k)}$ is chosen u.a.r. from instances of $k - HSAT_n$ with *exactly* $cm(n,k)$ clauses, then

$$\lim_{n\to\infty} P[F_{cm(n,k)} \text{ is useful}] = 1 \text{ if } c < \frac{(k-1)(k+1)}{k!}$$

$$\lim_{n\to\infty} P[F_{cm(n,k)} \text{ is useful}] = 0 \text{ if } c > \frac{(k-1)(k+1)}{k!}$$

The proofs of these results rely only on basic combinatorial tools.

Characterising useful Instances

F is useful for k-*HSAT* if and only:

$\exists\ B \subseteq \mathbf{X_n}$ such that $|B| \geq k - 1$ and

Q1) $\forall \{ y_1, \ldots, y_k \} \subseteq B$,

$$(\bar{y}_1 \vee \bar{y}_2 \vee \cdots \vee \bar{y}_k) \notin F.$$

Q2) $\forall \{ y_1, \ldots, y_{k-1} \} \subseteq B$, $\forall\ z \in \mathbf{X_n} - B$,

$$(z \vee \bar{y}_1 \vee \bar{y}_2 \vee \cdots \vee \bar{y}_{k-1}) \notin F.$$

$Q(B, F)$ denotes the relationship:

$B \subseteq \mathbf{X_n}$ with $|B| \geq k - 1$.

B satisfies Q1 and Q2 for F.

Proof: Suppose that $Q(B, F)$ for some B. Let β be the (partial) assignment

$$< x := 1 : \forall\, x \in B >$$

Then,

$F(\beta) \not\equiv 0$ from (Q1).
No clause of $F(\beta)$ is a single positive literal - (Q2).

Hence,

$$< x := 1 : \forall\, x \in B > \ , < x := 0 : \forall\, x \notin B >$$

is non-trivial, i.e. F is useful.

If F is useful and α a non-trivial assignment, let

$$B = \{\ x : x = 1 \ \mathit{under}\ \alpha\ \}$$

then $Q(B, F)$ is immediate. $\square$

Let $p(n) = \dfrac{\log_e n}{n}$.

$\forall\ k \geq 2$

$\forall\ \alpha > k - 1$

$$\lim_{n \to \infty} \mu_{\alpha p}(k - HSAT_n) = 0 \qquad \text{(a)}$$

$\forall\ \alpha < k - 1$

$$\lim_{n \to \infty} \mu_{\alpha p}(k - HSAT_n) = 1 \qquad \text{(b)}$$

Outline Proof of (a)

$$NTR(F) =_{def} \{ B \subseteq \mathbf{X_n} : Q(B, F) \}$$

$|NTR(F_p)|$ is a random variable. Obviously,

$\mu_{\alpha p}(k - HSAT_n) \leq E[\,|NTR(F_{\alpha p})|\,]$

Hence it suffices to show,

$$\lim_{n \to \infty} E[\,|NTR(F_{\alpha p})|\,] = 0$$

when $p = \log_e n/n$, $\alpha > k - 1$.

$$E[\,|NTR(F_{\alpha p})|\,] \leq \sum_{B \subseteq \mathbf{X_n}} P[\,Q(B, F_{\alpha p})\,]$$

$$= \sum_{r=k-1}^{n} \sum_{B : |B|=r} P[\,Q(B, F_{\alpha p})\,]$$

$$\leq \sum_{r=k-1}^{n} \binom{n}{r} (1 - \alpha p)^{\binom{r}{k} + \binom{r}{k-1}(n-r)}$$

$$\leq \sum_{r=k-1}^{n} \exp(-r \log_e n \, \gamma(r))$$

where $\gamma(r)$ is

$$\alpha \binom{r}{k-1} \left\{ \frac{1}{r} - \frac{1}{n} \right\} - 1 + \frac{\alpha}{nr} \binom{r}{k} + \frac{\log_e r - 1}{\log_e n}$$

For all $k \geq 2$ and any $\varepsilon > 0$.

If $\alpha = k - 1 + \varepsilon$, it can be shown that

$$\forall\ r : k - 1 \leq r \leq n \qquad \gamma(r) > \frac{\varepsilon}{2k}$$

(when n is sufficienctly large).

Hence,

$$\begin{aligned} E[\,|NTR(F_{\alpha p})|\,] &\leq \sum_{r=k-1}^{n} \exp(-r \log_e n\, \gamma(r)) \\ &\leq \sum_{r=k-1}^{n} \exp(-r\varepsilon/2k \log_e n) \\ &= \sum_{r=k-1}^{n} n^{-r\varepsilon/2k} \quad \underset{n}{\rightarrow} \quad 0 \end{aligned}$$

The proof of part (b) is less tortuous.

$\mu_{\alpha p}(k - HSAT_n)$ is

$$\geq P\,[\,\exists\, B : |B| = k - 1 \textit{ and } Q(B, F_{\alpha p})\,]$$

$$\geq 1 - P\,[\,\exists z\,(z \vee \bigvee_{i=1}^{k-1} \bar{x}_i) \in F_{\alpha p}\,]^{\binom{n}{k-1}}$$

$$\geq\ 1 - \exp\left(-\binom{n}{k-1} (1 - \alpha p)^{n-k+1} \right)$$

$$\geq\ 1 - \frac{1}{\Omega(\exp(n^{k-1-\alpha})} - o(1)$$

For $\alpha < k - 1$ this approaches 1. $\square$