Renormalization as a function of clause lengths for solving random k-SAT formulae.

Olivier Dubois and Gilles Dequen

LIP6, CNRS-Université Paris 6, 4 place Jussieu, 75252 Paris cedex 05, France.
LaRIA, Université de Picardie Jules Verne, CURI, 5 rue du moulin neuf, 80000 Amiens, France.
e-mail:Olivier.Dubois@lip6.fr, dequen@laria.u-picardie.fr

Introduction

In [2] we presented a new branch variable selection heuristic for solving random 3-SAT formulae within a DPL-type procedure. This heuristic relied on the notion of *backbone* of a SAT formula, a concept recently introduced in a statistical physics study [3]. We now present the generalization of this heuristic for solving random k-SAT formulae. This generalization rests on two principles. The first is inspired by an analogy with the well-known renormalization principle in statistical physics studies of phase transitions. The second depends on recent work, also within a statistical physics framework, investigating the dynamics of solving a SAT formula by means of a DPL-type procedure [1]. We shall show that using this heuristic in a DPL-type procedure quite markedly improves its effectiveness in solving random k-SAT formulae. Implementing this heuristic within *cnfs* [2] yields a complete solver that has the best current performance on all of 3-SAT, 4-SAT, 5-SAT, 6-SAT . . .

Renormalization

The h heuristic as described in [2] may be seen, in terms of phase transition physics, as some way of estimating how any variable appearing in a 3-SAT formula correlates with as many of the other variables of the formula as possible. A strong correlation between a variable x and a subset A of the other variables of the formula may for instance translate into the fact that any assignment of 0s or 1s to the variables of A determines or fixes the value of x, lest some clause be contradicted. When restricted to 3-SAT formula solution, this correlation estimate may be computed through only the length-2 clauses containing the variable evaluated by the h heuristic, the unit propagation being taken into account [2]. If, however, a k-SAT formula is to be solved, the correlation of a variable with the other variables of the formula, at some node in the solving tree, may have to be estimated through clauses of varying lengths, possibly ranging from $k-1$ to 2. This brings about strong disparities between the correlations of different occurrences of each variable with the other variables. Intuitively, if some occurrence of a variable is located in a clause of length 2, it will generally be more strongly correlated to the other variables of the formula than an occurrence appearing in a clause of higher length. Moreover, if an occurrence of a variable is associated within a clause with variables having very few occurrences, (e.g., pure literals), it generally has, on the contrary, weaker correlation to the other variables of the formula than an occurrence sharing a clause with variables having very many occurrences. The heuristic computation must take account of these disparities by a normalization aiming to reduce to a single scale all estimates of the correlations of occurrences of a variable with the other variables of the formula. Accuracy of this correlation requires that it be redefined at every node of the solution tree. The technique just described actually draws its inspiration from the renormalization principle in the statistical physics study of phase transitions; roughly stated, this consists in dividing a lattice into blocks of strongly correlated elements which by virtue of sharing the same correlation scale, have broadly similar behaviours. We present a very simple renormalization computation using the weighted means of the numbers of occurrences of the variables considered in the computation of the generalized h heuristic.

Formula-solving hardness diagram

The dynamic analysis of solving of a 3-SAT formula, as done by S. Cocco and R. Monasson in [1], shows that it is essential in the process of solution to control the evolution of the proportion of clauses of length 2 in a formula. Indeed, this proportion characterizes the hardness of solving a formula, and its trajectory on the phase diagram in the course of solution, may cross domains that correspond to either hard-to-solve or easy-to-solve formulae (see Cocco & Monasson [1]). We may imagine an extrapolation of this analysis to k-SAT formulae. Denoting by $p_2, p_3, p_4, \ldots, p_i, \ldots, p_k$ the proportions of clauses of lengths $2, 3, 4, \ldots, i, \ldots k$, in a k-SAT formula in the course of solution, the proportions p_i in the formula constitute a vector the trajectory of which can be plotted in a k-dimensional diagram. The analysis in [1] suggests that the trajectory of the p_i vector is just as crucial for the solving time of a formula, and that it is therefore as essential that this trajectory be kept under control. An analysis of the p_i trajectory along the lines of [1] is surely extremely complex. We merely show experimentally that in the course of solution, the nonzero p_i with the smallest index is a dominant factor with regard to the solving time of a k-SAT formula. As a consequence, we show how the h heuristic can be optimized.

Experimental results.

We implement the generalized h heuristic within the program *cnfs*[1] [2]. We have shown in [2] that solution performance on hard random 3-SAT formulae was considerably improved with *cnfs* as compared to the current best solvers. In like manner, we show that with the generalized heuristic presented here, *cnfs* significantly improves solving time on random 4-SAT, 5-SAT, 6-SAT ... formulae with respect to existing solvers. As a result, the branch

[1] `http://www.laria.u-picardie.fr/~dequen/sat/`

variable selection heuristic presented here for DPL-type procedures, has comparatively the best behaviour simultaneously on 3-SAT, 4-SAT, 5-SAT, 6-SAT ...

References

[1] S. Cocco and R. Monasson, *Trajectories in phase diagrams, growth processes and computational complexity: how search algorithms solve the 3-Satisfiability problem*, Phys. Rev. Lett. **86** (2001), 1654–1658.

[2] O. Dubois and G. Dequen, *A Backbone Search Heuristic for Efficient Solving of Hard 3-SAT Formulae*, Proc. of the 17th Internat. Joint Conf. on Artificial Intelligence (Seattle), August 4–10 2001, pp. 248–253.

[3] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky, *Determining Computational Complexity from Characteristic 'Phase Transitions'*, Nature 400 (1999), 133–137.