Information Processing Letters 111 (2011) 985–988

Contents lists available at ScienceDirect

Information Processing Letters

www.elsevier.com/locate/ipl

Threshold behaviors of a random constraint satisfaction problem with exact phase transitions ☆

Chunyan Zhao [a,b,*], Zhiming Zheng [a,b]

[a] School of Mathematics and Systems Science, Beihang University, Beijing 100191, China
[b] Key Laboratory of Mathematics, Informatics and Behavioral Semantics, Ministry of Education, China

ARTICLE INFO

Article history:
Received 13 June 2010
Received in revised form 6 July 2011
Accepted 6 July 2011
Available online 18 July 2011
Communicated by J. Torán

Keywords:
Constraint satisfaction problem
Model RB
Phase transition
Finite-size scaling
Threshold behavior
Computational complexity

ABSTRACT

We consider a random constraint satisfaction problem named model RB, which exhibits a sharp satisfiability phase-transition phenomenon when the control parameters pass through the critical values denoted by r_{cr} and p_{cr}. Using finite-size scaling analysis, we bound the width of the transition region for finite problem size n, which might be the first rigorous study on the threshold behaviors of NP-complete problems.

1. Introduction

Constraint satisfaction problems (CSPs) arise in the disciplines of theoretical computer science, discrete mathematics and statistical physics. Generally speaking, a CSP consists of n variables and m constraints. Each variable takes values from a nonempty domain D with size $|D| = d$. Each constraint is defined on a subset of k distinct variables and specifies the incompatible values for these variables. An assignment of the n variables satisfying the m constraints simultaneously is a solution of the CSP. Given a CSP instance, the problem is to find such an assignment or to show that it is unsatisfiable. Accumulative empirical evidences suggest that CSPs undergo phase transitions at some critical values. The k-satisfiability problem (k-SAT) is perhaps the most famous example of CSPs [1]. For $k = 2$, the problem is in P and it has been analytically shown that the threshold locates at $\alpha = 1$ ($\alpha \equiv m/n$) [2], which is further strengthened in [3] and [4]. To gain a better understanding of the changes of phase-transition region, finite-size scaling analysis, a method from statistical physics was introduced. For random 2-SAT, how the width of the transition region narrows with increasing problem size is shown in [5] and [6]. However, for random k-SAT ($k \geqslant 3$), which is NP-complete, rigorous analysis appears to be difficult to locate the exact values of critical points. Loose upper and lower bounds have been obtained. Experimental evidence, however, strongly suggests a threshold occurs at $\alpha \approx 4.2667$ [7] for random 3-SAT.

Model RB is a random CSP proposed by Xu and Li [8] to overcome the trivial unsatisfiability of standard CSP models. In model RB, first, we construct a constraint by randomly choosing k ($k \geqslant 2$) distinct variables among n ones, and then uniformly select pd^k ($d = n^\alpha$) incompatible values for the k variables. Next, we repeat this process to obtain $m = rn \ln n$ independently chosen constraints and their associated sets of incompatible values. We take the logical AND of the m constraints, which gives an instance

☆ This work has been supported by International Cooperative Program (Grant No. 2010DFR00700) and NSFC (Grant No. 60973033).

* Corresponding author at: School of Mathematics and Systems Science, Beihang University, Beijing 100191, China.

E-mail address: zhaocy@ss.buaa.edu.cn (C. Zhao).

0020-0190/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ipl.2011.07.006

of model RB. The probability of a random instance I being satisfiable is denoted by $\mathbf{Pr}(I \text{ is SAT})$, it is shown in [8] that

Theorem 1.1. *Let $r_{cr} = -\frac{\alpha}{\ln(1-p)}$. If $\alpha > \frac{1}{k}$, $0 < p < 1$ are two constants and $k \geqslant \frac{1}{1-p}$, then*

$$\lim_{n\to\infty} \mathbf{Pr}(I \text{ is SAT}) = 1, \quad \text{if } r < r_{cr}$$
$$\lim_{n\to\infty} \mathbf{Pr}(I \text{ is SAT}) = 0, \quad \text{if } r > r_{cr}$$

Theorem 1.2. *Let $p_{cr} = 1 - e^{-\frac{\alpha}{r}}$. If $\alpha > \frac{1}{k}$, $r > 0$ are two constants and $ke^{-\frac{\alpha}{r}} \geqslant 1$, then*

$$\lim_{n\to\infty} \mathbf{Pr}(I \text{ is SAT}) = 1, \quad \text{if } p < p_{cr}$$
$$\lim_{n\to\infty} \mathbf{Pr}(I \text{ is SAT}) = 0, \quad \text{if} p > p_{cr}$$

Therefore, the satisfiability phase transition of model RB is exactly established. Moreover, it has been shown both theoretically [9] and experimentally [10] that almost all instances of model RB are hard to solve at the thresholds. Since hard instances can be used as benchmarks for evaluating the performance of algorithms, the instances of model RB have been widely used in various algorithm competitions and many research papers in the past few years. So studies on the threshold behaviors of model RB can provide further understanding of the hardness of these instances at the thresholds.

In this paper, we use finite-size scaling analysis to characterize the threshold behaviors of model RB with finite problem size n, which give direct evidence for the broadening of the transition region due to finite-size effects.

2. Main results

Suppose δ is an arbitrarily small constant. We can obtain the following results under the same conditions of Theorem 1.1.

Theorem 2.1. *Let $r_u = r_{cr} + \Theta(\frac{1}{n\ln n})$, then*

$$\mathbf{Pr}(I \text{ is SAT}) \leqslant \delta, \quad \text{if } r \geqslant r_u.$$

Theorem 2.2. *Let $r_l = r_{cr} - \Theta(\frac{1}{n^{1-\nu}\ln n})$, then*

$$\mathbf{Pr}(I \text{ is SAT}) \geqslant 1 - \delta, \quad \text{if } r \leqslant r_l$$

where the constants implicit in Θ depend on δ and $0 < \nu < 1$ is a constant.

It is shown that the threshold of model RB scales as $\frac{1}{n\ln n}$ and $\frac{1}{n^{1-\nu}\ln n}$ from above and below respectively, which bounds the width of the transition region accordingly. Clearly $\Theta(\frac{1}{n\ln n})$ and $\Theta(\frac{1}{n^{1-\nu}\ln n})$ vanish as n approaches infinity. So the exact critical point r_{cr} can be identified by the fact that δ is arbitrarily small. Similar results about the control parameter p can also be obtained by analogy with Theorems 2.1 and 2.2. In this paper, we mainly focus on the results about the control parameter r.

3. Proof of the results

For a random instance I, let $S(I) = \{\sigma\colon \sigma \text{ satisfies } I\} \subseteq D^n$ be the set of solutions of I, where σ is an assignment of the n variables. Let $X = |S(I)|$ denote the number of solutions of I, we will prove Theorem 2.1 by Markov inequality, and use the second moment method to prove Theorem 2.2.

Proof of Theorem 2.1. We begin by calculating the expectation of X

$$\mathbf{E}(X) = d^n\,\mathbf{Pr}(\sigma \text{ satisfies } I) = n^{\alpha n}(1-p)^{rn\ln n} \tag{1}$$

Assume that $\mathbf{E}(X) \leqslant \delta$, we have

$$\big[\alpha + r\ln(1-p)\big]n\ln n \leqslant \ln\delta \tag{2}$$

Hence we can obtain

$$r \geqslant r_{cr} + \frac{\ln\delta}{n\ln n\ln(1-p)} \equiv r_{cr} + \Theta\left(\frac{1}{n\ln n}\right) \tag{3}$$

Therefore, using Markov inequality $\mathbf{Pr}(I \text{ is SAT}) \leqslant \mathbf{E}(X)$, we can immediately get that if $r \geqslant r_u$, then $\mathbf{Pr}(I \text{ is SAT}) \leqslant \delta$. □

Proof of Theorem 2.2. Let $\mathbf{Pr}(\sigma_1, \sigma_2 \text{ satisfy } I)$ denote the probability that a pair of assignments σ_1, σ_2 satisfy I simultaneously. So we can have

$$\mathbf{E}(X^2) = \sum_{t=0}^{n} d^n \binom{n}{t} (d-1)^{n-t}\,\mathbf{Pr}(\sigma_1, \sigma_2 \text{ satisfy } I) \tag{4}$$

where t denotes the number of variables for which σ_1 and σ_2 take the same values, and

$$\mathbf{Pr}(\sigma_1, \sigma_2 \text{ satisfy } I) = \left[(1-p)\frac{\binom{t}{k}}{\binom{n}{k}} + \frac{\binom{d^k-2}{pd^k}}{\binom{d^k}{pd^k}}\left(1 - \frac{\binom{t}{k}}{\binom{n}{k}}\right)\right]^{rn\ln n} \tag{5}$$

Explanation of (5): For each constraint in I, the probability of σ_1 and σ_2 being assigned the same values to the k variables is $\binom{t}{k}/\binom{n}{k}$, in this case, the probability of σ_1 and σ_2 satisfying the constraint is $1-p$; otherwise the probability is $1 - \binom{t}{k}/\binom{n}{k}$, and the satisfying probability is $\binom{d^k-2}{pd^k}/\binom{d^k}{pd^k}$.

By the condition of $\alpha > \frac{1}{k}$ in Theorem 1.1, we can rewrite (5) as

$$\mathbf{Pr}(\sigma_1, \sigma_2 \text{ satisfy } I) = \left[(1-p)A(t) + (1-p)^2\big(1 - A(t)\big)\right]^{rn\ln n} \times \left(1 + O\left(\frac{1}{n}\right)\right) \tag{6}$$

where $A(t) = (\frac{t}{n})^k + \frac{g(t)}{n}$ with $g(t) = \frac{k(k-1)}{2}(\frac{t}{n})^{k-1}(\frac{t}{n} - 1)$. Clearly $A(t) \leqslant (\frac{t}{n})^k$ for $g(t) \leqslant 0$ with $t \in [0, n]$.

Consequently, we use (1), (4) and (6) to have