![Scientific Research Publishing]

Complexity Analysis of Mixed Constraint Satisfaction Problems

Xiaomei Shi

College of Science, University of Shanghai for Science and Technology, Shanghai, China
Email: shixiaomeiusst@163.com

Abstract

This paper analyzes the resolution complexity of a random CSP model named model RB^{mix}, the instance of which is composed by constraints with different length. For model RB^{mix}, the existence of phase transitions has been established and the threshold points have been located exactly. By encoding the random instances into CNF formulas, it is proved that almost all instances of model RB^{mix} have no tree-like resolution proofs of less than exponential size. Thus the model RB^{mix} can generate abundant hard instances in the threshold. This result is of great significance for algorithm testing and complexity analysis in NP-complete problems.

Keywords

Model RB^{mix}, Tree-Like Resolution, Complexity, Hard Instances

1. Introduction

Constraint satisfaction problem (CSP) is originated from artificial intelligence. It is a very important branch in the field of artificial intelligence, and it is a problem widely studied in computer science, information science, discrete mathematics and other interdisciplines. At present, a number of problems restricting the development of computer science, automatic control, system engineering and other disciplines can be modeled as CSPs. At the same time, CSPs are widely used in related application fields such as resource allocation, pattern recognition, temporal reasoning and image recognition.

CSP is composed of a set of variables and a set of constraints. Each variable takes a value from the corresponding non-empty domain. The number of elements in the domain can be fixed, or change with the number of variables. Each constraint has a corresponding incompatible assignment set to restrict the values of variables appearing in the constraint. The constraint set randomly selected

constitutes a random instance of CSP. Given a random instance, if there exists an assignment to the variables satisfying all the constraints in the instance simultaneously, we call this assignment a solution. Usually, most of the CSPs are NP-complete problems. A lot of theoretical research and experimental results show that for many NP-complete problems we can define a control parameter (such as constraint density, constraint tightness, etc.). There is a critical value of the control parameter, below which the probability of an instance being satisfiable goes to 1, and above which it goes to 0 as the number of variables approaches infinity. People call this phenomenon the phase transition. The critical value is a satisfiability phase transition point.

Random k-SAT (k-satisfiability. Every clause of random k-SAT is randomly and independently generated. Each clause contains just k different variables) problem is a typical CSP with a fixed domain (The variable is Boolean, *i.e.* 0 and 1 or T and F), and the first NP-complete problem proved by Cook. Research shows that when $k = 2$, 2-SAT is in P class [1] [2]. We can get the satisfiability phase transition point $\alpha = 1$ (α represents the ratio of the number of constraints and the number of variables). When $k \geq 3$, it belongs to NP-complete problem [3]. Although the phase transition phenomenon has been established, it is hard to get the accurate phase transition point rigorously. The currently known upper and lower bound of the phase transition region of the 3-SAT problem is 4.4898 [4] and 3.52 [5], respectively. It is shown by Mertens *et al.* that the satisfiability phase transformation of 3-SAT problem occurs near 4.2667 [6] experimentally. In order to study the transition behavior between P and NP-complete problem, Monasson *et al.* [7] proposed the model $(2 + p)$-SAT mixing with 2-clause and 3-clause. Here, $p \in [0,1]$, and in m clauses of the model there are $(1 - p)m$ 2-clauses and pm 3-clauses. Obviously, when $p = 0$, it is a 2-SAT problem, and when $p = 1$, it is a 3-SAT problem. Later, It is rigorously proved by Achlioptas *et al.* [8] that when $p < 2/5$ the model had transition characteristics similar to those of 2-SAT problem, and we could get the exact satisfiability phase transition point. When $p \geq 2/5$, the model has transition characteristics similar to those of 3-SAT problem, and we can't get the exact phase transition point. This means that phase transition characteristics of model $(2 + p)$-SAT are determined by the ratio of 2-clauses and 3-clauses.

Model RB (Revised B) is a nontrivial random CSP with growing variable domain. Specifically, in order to overcome the trivial gradual unsolvability of model B in the standard model CSP [9] [10] (that is the binary model CSP), Xu *et al.* proposed a new CSP model, *i.e.* model RB [11]. This model is a modification of model B in terms of domain of variables and the number of constraints, which avoids model B's disadvantage of failure to generate hard instances. In [11], it has been strictly proved that phase transitions do exist for model RB as the number of variables approaches infinity. Moreover, the phase transition points are also known exactly. Later, Xu *et al.* theoretically proved that random instances generated by model RB in phase transition region had the exponential tree-like resolution complexity, which meant that these instances were almost all

hard [12]. Later, an experiment also proved that near the phase transition threshold the hardness of finding solutions grows exponentially with the problem size [13]. Therefore, hard instances generated based on model RB are widely applied in all kinds of algorithm competitions such as CSP and SAT [14] as benchmarks, in order to test the performance of the algorithm. Since it was proposed, in a relatively short period of time model RB drew the attention of some scholar and was studied by them. Zhao *et al.* analyzed how the phase transition region of model RB became an exact phase transition point under the finite-size scaling effect, and gave the lower bound of scaling window [15]. Based on model RB, Shen *et al.* proposed a new model CSP, *i.e.* model d-k-CSP, and proved that the model also had the exact satisfiability phase transition phenomenon, and could generate a large number of hard instances [16] [17] [18]. In terms of the algorithm, Zhao *et al.* introduced the cavity method in statistical physics, and successively proposed two kinds of random CSP solved by the message passing algorithm guided by Belief Propagation and Reinforced Belief Propagation algorithm generated by model RB with a large domain [19] [20] [21]. Recently, Yuan *et al.* proposed RSA (Revised Simulated Annealing Algorithm) and GSA (Genetic-simulated Annealing Algorithm) to solve random instances of model RB [22].

Inspired by model $(2+p)$-SAT, Zhao *et al.* proposed a random CSP model mixed with constraints with different lengths (The constraint length is defined as the number of variables contained in the constraints), *i.e.* model RB^{mix}. Interestingly, the phase transition behavior of model $(2+p)$-SAT is determined by the proportion of 2-clauses and 3-clauses, while the phase transition phenomenon of model RB^{mix} has nothing to do with the number of constraints with different lengths. It has been proved that model RB^{mix} mixed with constraints with different lengths has the exact satisfiability phase transition phenomenon similar to that of model RB, and the phase transition points can be located exactly. We know that if the new model CSP has the exact satisfiability phase transition phenomenon and can generate a large number of hard instances, the model has important significance for testing of CSP algorithm. For model RB^{mix}, we use the resolution method (its general form is for logic formula) to encode random instances generated by model RB^{mix} into the conjunctive normal form in SAT problem, *i.e.* CNF formula. By using five lemmas, we prove a famous theorem about the resolution length. Therefore, it is proved that the random instances generated by model RB^{mix} almost have no resolution complexity for which the resolution length is less than the exponential size. This shows that model RB^{mix} is hard for algorithms based on resolution, so model RB^{mix} can generate a large number of hard instances in the phase transition region. It is proved in this paper that almost all instances of model RB^{mix} have no tree-like resolution proofs of less than exponential size. Thus, we not only introduce new families of random CSP instances hard to solve, which can be useful in the experimental evaluation of CSP algorithms, but also propose a CSP model with both many hard instances and exact phase transitions.

2. Model RB^mix

The random instance of model RB^{mix} is composed of the variable set $\{x_1, x_2, \cdots, x_n\}$ and the constraint set $\{C_1, C_2, \cdots, C_m\}$, where $m = rn \ln n$ ($r > 0$ is called the constraint density). Each variable x_i takes values from its corresponding domain D_i. Here, $|D_i| = d = n^\alpha$ ($\alpha > 0$ is a constant), which means that the domain of variables grows polynomially with the variable number n. We divide m constraints into ℓ groups. Each group contains $m_i = r_i n \ln n$ constraints. Here, $\sum_{i=1}^{\ell} r_i = r$. The constraints in each group are generated in the following way.

Step 1: m_i k_i-ary constrains are selected randomly and repeatably ($k_i \geq 2$). Each k_i-ary constrain is composed of k_i different variables selected randomly from n variable sets. Here, k_i is called the constraint length.

Step 2: For each k_i-ary constrain, pd^{k_i} assignments are selected randomly from d^{k_i} assignments without repetition to form the corresponding incompatible assignment set and limit values of k_i variables. Here, $p \in (0,1)$ is called constraint tightness.

If random instances generated by model RB^{mix} have a group of assignments of n variables, so that all k_i-ary constrains are satisfied at the same time, then the group of assignments is a solution. If the assignments of k_i variables do not belong to an incompatible assignment set, this constraint is satiable. Otherwise, it is unsatisfiable.

Thus, model RB^{mix} is a promotion of model RB on the constraints. In model RB, all constraints are k-ary constraints, which means that the constraint length is fixed. Model RB^{mix} is composed of constraints with different lengths. Apparently, model RB is a special circumstance of model RB^{mix} when $\ell = 1$. Therefore, the constraint composition of model RB^{mix} is more generally representative.

Assume that $\Pr(\text{SAT})$ is the probability of random instance satisfaction of model RB^{mix}. Let $\kappa = \min\{k_1, k_2, \cdots, k_\ell\}$. Then we've proved that the following two theorems hold.

Theorem 1 Assume that $r_s = -\dfrac{\alpha}{\ln(1-p)}$. If $\alpha > \dfrac{1}{\kappa}$ and $p \leq 1 - \dfrac{1}{\kappa}$, then:

$$\lim_{n \to \infty} \Pr(\text{SAT}) = \begin{cases} 1 & \text{if } r < r_s \\ 0 & \text{if } r > r_s \end{cases}. \tag{1}$$

Theorem 2 Assume that $p_s = 1 - e^{-\frac{\alpha}{r}}$. If $\alpha > \dfrac{1}{\kappa}$ and $r \geq \dfrac{\alpha}{\ln k}$, then:

$$\lim_{n \to \infty} \Pr(\text{SAT}) = \begin{cases} 1 & \text{if } p < p_s \\ 0 & \text{if } p > p_s \end{cases}. \tag{2}$$

From the above two theorems, we can know that under the condition that the domain of variables is not too small ($\alpha > 1/\kappa$), as long as the constraint tightness is low enough or the constraint density is big enough ($r \geq \alpha/\ln k$), when $n \to \infty$, with the continuous increase of r or p, The probability of a random instance generated by model RB^{mix} being satisfiable suddenly changes from 1 to 0,

which means that model RB$^{\text{mix}}$ has the satisfiability phase transition phenomenon, and it can get the accurate phase transition point. Therefore, model RB$^{\text{mix}}$ is a mixed random CSP with accurate phase transition and a large domain.

3. Main Results

Assume that I is a random instance generated by model RB$^{\text{mix}}$. We can get the following theorem.

Theorem 3 If $p > p_s$, then I almost has no resolution proof of the length less than $2^{\Omega(n)}$. Here, $\Omega(n) = \{f(n): \text{Constant } c > 0, f(n) \geq cn\}$.

As we know, for the unsatisfiable random instance I, there must be the shortest resolution proof which can reason out empty clauses. Its length is the lower bound of time consumed by any algorithm based on the resolution principle. Therefore, from Theorem 3 we can know that the solution algorithm of unsatisfiable instances of model RB$^{\text{mix}}$ has the complexity of exponential size. Thus in the phase transition region model RB$^{\text{mix}}$ can generate a large number of hard instances.

4. Proof of Theorem 3

This paper uses the resolution method to analyze the complexity of model RB$^{\text{mix}}$. The general form is CNF formula for SAT problem. The so-called CNF formula refers to the conjunction expression ($\wedge$) of clause (disjunctive form $\vee$ of positive literal l or negative literal $\neg l$). Therefore, we first must encode the random instance I of model RB$^{\text{mix}}$ according to a certain rules into CNF formula, then the resolution complexity of I is the complexity of the corresponding CNF formula.

In model RB$^{\text{mix}}$, assume that the domain of each variable x_i is $D_i = \{1, 2, \cdots, d\}$. Define the new proposition variable x_{ij} (Boolean). If the value of x_i is j ($j \in D_i$), then $x_{ij} = \text{T}$. The following three types of clauses are required to encode I into CNF formula.

(1) Clauses of the domain: Each variable x_i is taken from its domain D_i, *i.e.* $x_{i1} \vee x_{i2} \vee \cdots \vee x_{id}$;

(2) A value clause at most: Ensure that each variable takes a value from its domain at most at a time, *i.e.* $\neg x_{ij_1} \vee \neg x_{ij_2}$, $j_1, j_2 \in D_i$ and $j_1 \neq j_2$;

(3) Conflicted clause: It is used to remove the assignment in the uncoordinated set, so that the constraint is satisfied. For example, $\{3, 5\}$ is an element in the uncoordinated assignment set limiting the value of x_1 and x_2. Then the conflicted clause is $\neg x_{13} \vee \neg x_{25}$.

Definition 1 (Length of the clause) The number of variables appearing in Clause C is called length of C, written as $\omega(C)$.

Definition 2 (Length of CNF) The maximum length of all the clauses in F of CNF formula F is called the length of F, *i.e.* $\omega(\text{F}) = \max\limits_{C \in \text{F}} \{\omega(C)\}$.

Definition 3 (Resolution) The clause sequence $\pi = C_1, C_2, \cdots, C_t$ ($C_t = C$) is called a resolution in which C is derived from F. If and only if each C_i belongs to F, or C_i is derived from C_j and C_k ($1 \leq j, k \leq i$) in accordance with

the rule $C_j = A \lor x$ and $C_k = B \lor \neg x \Rightarrow C_i = A \lor B$ (among them A and B are clauses, x is propositional variable), t is its length.

Definition 4 (Derived length) The length for C is derived from F is the minimum length of π, written as $\omega(\mathrm{F}\,|-C)$, *i.e.*

$$\omega(\mathrm{F}\,|-C) = \min_{\pi}\left\{\max_{C \in \pi}\left\{\omega(C)\right\}\right\}.$$

Definition 5 (Resolution) The resolution in which an empty clause (expressed with $\square$) is derived from F in CNF formula is called the resolution of F, written as $\mathrm{Res}(\mathrm{F})$, *i.e.* $\mathrm{Res}(\mathrm{F}) = \min_{\pi}\{t : \pi = C_1, C_2, \cdots, C_t\}$. In which, $C_t = \square$. Then $\mathrm{Res}(\mathrm{F})$ is the resolution complexity of formula F.

We consider the clause in π as a node. If C_i is derived from C_i and C_k, the edge is used to connect C_i and C_j, C_i and C_k. If the resulting graph G_π is a tree, π is called the tree-like resolution.

For the unsatisfiable formula F, there must be the shortest resolution proof deriving $\square$. Its length is the lower bound of time consumed by any algorithm based on the resolution principle.

Ben-Sasson and Wigderson gave the following theorem used to prove the lower bound of the resolution length of CNF formula [23].

Theorem 4 Assume that F is a CNF formula with n variables, then we have:

$$\mathrm{Res}(\mathrm{F}) \geq 2^{\omega(F|-\square)-\omega(F)}. \tag{3}$$

We want to use Theorem 4 to prove the important conclusion Theorem 3 in this paper. We need to use the following five lemmas.

Before giving Lemma 1, we first give the following definition.

Definition 6 (Subproblem) Assume that I is an instance of a CSP, then the instance composed of subsets of constraint set in I is called a subproblem of I.

Lemma 1 Assume that I is a random instane generated by model $\mathrm{RB}^{\mathrm{mix}}$, then $\exists 0 < c < 1$, so that almost every subproblem with the size of $s \leq cn$ in I has $m_0 = \beta s \ln n$ ($\beta = -\alpha/6\kappa \ln(1-p)$) constraints at most. In m_0 constraints, the length of $m_0^i = \beta_i s \ln n$ constraints is k_i, in which $i = 1, 2, \cdots, \ell$ and $\beta = \sum_{i=1}^{\ell} \beta_i$. Here, the subproblem with the size of s refers to s variables contained in the subproblem of I.

Proof Assume that A = {I has at least $m_0 = \beta s \ln n$ constraints and a subproblem with the size of $s \leq cn$}. We prove that $\lim_{n\to\infty} \mathrm{Pr}(A) = 0$.

Let $\kappa' = \max\{k_1, k_2, \cdots, k_\ell\}$, then we have

$$
\begin{aligned}
\mathrm{Pr}(A) &\leq \sum_{1 \leq s \leq cn} C_n^s C_m^{m_0} \prod_{i=1}^{\ell} \left(\frac{C_s^{k_i}}{C_n^{k_i}}\right)^{m_0^i} \\
&\leq \sum_{1 \leq s \leq cn} \left(\frac{en}{s}\right)^s \left(\frac{ern \ln n}{\beta s \ln n}\right)^{\beta s \ln n} \left(\frac{s}{n}\right)^{\sum_{i=1}^{\ell} k_i \beta_i s \ln n} \\
&\leq \sum_{1 \leq s \leq cn} \left(\frac{en}{s}\right)^s \left(\frac{ern}{\beta s}\right)^{\beta s \ln n} \left(\frac{s}{n}\right)^{\kappa' \beta s \ln n} \\
&\leq \sum_{1 \leq s \leq cn} \left(e\left(\frac{er}{\beta}\right)^{\beta \ln n} \left(\frac{s}{n}\right)^{(\kappa'-1)\beta \ln n - 1}\right)^s.
\end{aligned}
\tag{4}
$$

For the sufficiently large n, here exists a constant $c_1 > 0$, so that

$$e\left(\frac{er}{\beta}\right)^{\beta \ln n} = en^{\beta(1+\ln r - \ln \beta)} < n^{c_1}. \tag{5}$$

Let $c < \frac{1}{2}e^{-\frac{2+c_1}{(\kappa'-1)\beta}}$, then

$$e\left(\frac{er}{\beta}\right)^{\beta \ln n}\left(\frac{s}{n}\right)^{(\kappa'-1)\beta \ln n - 1} \le n^{c_1}c^{(\kappa'-1)\beta \ln n - 1} \le \frac{1}{n^2}. \tag{6}$$

thus

$$\Pr(A) \le \sum_{1 \le s \le cn}\left(\frac{1}{n^2}\right)^s \le \frac{cn}{n^2} = \frac{c}{n}. \tag{7}$$

Therefore, $\lim_{n\to\infty}\Pr(A) = 0$.

Before giving Lemma 2, we first give the following definitions.

Definition 7 (*i*-constraint assignment group) Assume that i constraints contain variable x. If assignments are made for variables except x in i constraints, the such assignment group is called *i*-constraint assignment group, written as T_{ix}.

Definition 8 (flawed *i*-constraint assignment) The assignment of given variable x is $\sigma \in D_x$ (D_x is the domain of x). If (σ, T_{ix}) makes at least one constraint in these i constraints unsatisfiable, variable x's assignment σ is called flawed for T_{ix}. Further, if any assignment of variable x is defective for T_{ix}, T_{ix} is called a defective *i*-constraint assignment group.

Lemma 2 Assume that I is a random instance generated by model RB^{mix}, then for $i \le 3\kappa\beta \ln n$ defective T_{ix} almost may not exist in I. In which, in these i constraints the length of $i_\tau = 3\kappa\beta_\tau \ln n$ constraints is k_τ. Here, $\sum_{\tau=1}^{\ell} i_\tau = i$.

Proof Assume that $A = \{$I has a defective *i*-constraint assignment group $T_{ix}\}$. In which, $i \le 3\kappa\beta \ln n$, which means that it is proved that $\lim_{n\to\infty}\Pr(A) = 0$.

Next, let $B = \{T_{ix}$ is not flawed$\}$, then we have

$$\Pr(\bar{B})\big|_{i \le 3\kappa\beta \ln n} \le \Pr(\bar{B})\big|_{i = 3\kappa\beta \ln n}. \tag{8}$$

$B_\sigma = \{\sigma$ is not flawed for T_{ix}, $\sigma \in D_x\}$. It's worth noting that for model RB^{mix} $D_x = \{1, 2, \cdots, d\}$, $d = n^\alpha$, so

$$\Pr(\bar{B}) = 1 - \Pr(B_1 \bigcup B_2 \bigcup \cdots \bigcup B_d)$$
$$= 1 - \sum_{\sigma=1}^{d}(-1)^{\sigma-1} C_d^\sigma \cdot \Pr(B_1 B_2 \cdots B_\sigma) \tag{9}$$

Here,

$$\Pr(B_1 B_2 \cdots B_\sigma) = \prod_{\tau=1}^{\ell}\left(\frac{C_{pd^{k_\tau}}^{d^{k_\tau}-\sigma}}{C_{d^{k_\tau}}^{pd^{k_\tau}}}\right)^{i_\tau}. \tag{10}$$

because

$$\frac{C_{pd^{k_\tau}}^{d^{k_\tau}-\sigma}}{C_{d^{k_\tau}}^{pd^{k_\tau}}} = \frac{\left(d^{k_\tau} - pd^{k_\tau}\right)}{d^{k_\tau}} \cdots \frac{\left(d^{k_\tau} - pd^{k_\tau} - \sigma + 1\right)}{\left(d^{k_\tau} - \sigma + 1\right)}, \tag{11}$$

And for $j = 0, 1, \cdots, \sigma$, we have

$$\frac{\left(d^{k_\tau} - pd^{k_\tau} - j + 1\right)}{\left(d^{k_\tau} - j + 1\right)} \leq 1 - p. \tag{12}$$

From $\beta = -\dfrac{\alpha}{6\kappa \ln(1-p)}$ in Lemma 1,

$$\Pr\left(B_1 B_2 \cdots B_\sigma\right) \leq \prod_{\tau=1}^{\ell}\left[\left(1-p\right)^\sigma\right]^{i_\tau} = \left(1-p\right)^{3\sigma\kappa\beta\ln n}$$

$$= \left(1-p\right)^{-\frac{\alpha\sigma\ln n}{2\ln(1-p)}} = n^{-\frac{\alpha\sigma}{2}}, \tag{13}$$

Thus,

$$\Pr\left(\overline{B}\right) \leq 1 + \sum_{\sigma=1}^{d}(-1)^\sigma C_d^\sigma \cdot n^{-\frac{\alpha\sigma}{2}} = \left(1 - n^{-\frac{\alpha}{2}}\right)^d \leq e^{-n^{-\frac{\alpha}{2}}}. \tag{14}$$

Because the maximum choices of i-constraint assignment group T_{ix} are as follows

$$nC_m^i d^{(k_1-1)i_1} d^{(k_2-1)i_2} \cdots d^{(k_\ell-1)i_\ell}$$

And $i \leq 3\kappa\beta\ln n$, so when n is sufficiently large, constant $c_2 > 0$. Let

$$nC_m^i d^{(k_1-1)i_1} d^{(k_2-1)i_2} \cdots d^{(k_\ell-1)i_\ell}$$

$$= nC_{rn\ln n}^i \cdot d^{\sum_{\tau=1}^{\ell}(k_i-1)i_\tau}$$

$$\leq n\left(\frac{ern\ln n}{3\kappa\beta\ln n}\right)^{3\kappa\beta\ln n} \cdot n^{\alpha(\kappa'-1)\cdot 3\kappa\beta\ln n} \tag{15}$$

$$< e^{c_2 \ln^2 n}$$

Thus, we have

$$\Pr(A) = \sum_{i=1}^{3\kappa\beta\ln n} nC_m^i d^{(k_1-1)i_1} d^{(k_2-1)i_2} \cdots d^{(k_\ell-1)i_\ell} \Pr\left(\overline{B}\right)$$

$$\leq e^{c_2 \ln^2 n} \cdot e^{-n^{\frac{\alpha}{2}}} \cdot 3\kappa\beta\ln n \tag{16}$$

So $\lim\limits_{n\to\infty} \Pr(A) = 0$.

Lemma 3 Assume that I is a random instance generated by model RB^{mix}, then almost any subproblem with the size of $s \leq cn$ of I can be satisfied.

Proof Assume that I has the unsatisfiable subproblem with the size of $s \leq cn$ in I is, and the minimum unsatisfiable subproblem is I_1. According to Lemma 1, I_1 contains $\beta s \ln n$ constraints at most. So I_1 at least contains one variable x with the maximum degree of $\kappa\beta\ln n$ (the degree of variable is defined as the number of constraints containing the variable). After x and constraints containing x are removed, the remaining subproblem I_2 can be satisfied (I_1 is the smallest unsatisfiable subproblem). Thus, I_2 has the assignment group T_{ix} ($i \leq \kappa\beta\ln n$), so that I_2 can be satisfied, but I_1 is unsatisfiable. In means that for $\forall\sigma \in D_x$ $\left(\sigma, T_{ix}\right)$ cannot make I_1 satisfiable, so T_{ix} is defective. Therefore, for $i \leq \kappa\beta\ln n$ I has a flawed T_{ix}, which is contradictory with Lemma 2, so the assumption is not established. Thus, Lemma 3 is proved.

Before giving Lemma 4, we need to give the following definition.

Definition 9 (Complexity) Assume that I is a random CSP instance. Encode I into CNF formula F. Assume that F's clause C is derived from sequence $\pi = C_1, C_2, \cdots, C_t$ ($C_t = C$). We can say that the number of variables contained in the smallest scale of subproblems deriving C is the complexity of the clause C, written as $\mu(C)$.

Lemma 4 Assume that I is a random instance generated by model RB$^{\text{mix}}$. Encode I into CNF formula F. Then almost any inversion of F has a clause which satisfies $cn/2 \leq \mu(C) \leq cn$.

Proof From Lemma 3 we can know that $\mu(\square) > cn$. We establish a binary tree with the node of clause. In which, $\square$ is the root node, the child nodes are two clauses with which we can reason out $\square$, until the leaf node. Obviously, the clause complexity of leaf node $\mu(C_{leaf}) \leq \kappa$. From root node to leaf node, because the complexity of each clause does not increase, $\exists$ clause C_0 makes $\mu(C_0) \geq cn$, and it can be derived that C_1 and C_2, two clauses of C_0 (child nodes) satisfy $C_1 \leq cn$ and $C_2 \leq cn$. From $\mu(C_1) + \mu(C_2) \geq \mu(C_0)$, In C_1 and C_2 a C satisfies $cn/2 \leq \mu(C) \leq cn$.

Lemma 5 $\lim\limits_{n \to \infty} \Pr\left(\omega(C) \geq \dfrac{cn}{6} \middle| \dfrac{cn}{2} \leq \mu(C) \leq cn \right) = 1$.

Proof Assume that I is a random instance generated by model RB$^{\text{mix}}$. Encode I into CNF formula F. Assume that I_1 is the smallest subproblem of $F|-C$, and $cn/2 \leq \mu(C)$. From Lemma 1 we can know that I_1 contains $\beta cn \ln n$ constraints at most. These $\beta cn \ln n$ constraints contain up to $\kappa\beta cn \ln n$ variables. So the maximum number of variables with the degree higher than $3\kappa\beta \ln n$ is $cn/3$. Next, we will prove that these variables are in clause C.

Assume that variable x whose degree is not greater than not $3\kappa\beta \ln n$ is not in clause C. Assume that I_2 is the remaing subproblem after x and constraints containing x are removed. Because I_1 is the smallest subproblem, clause C cannot be derived from the corresponding CNF formula of I_2. So any assignment of variable x does not meet the corresponding CNF formula of I_1, and T_{ix} is defective. This is contradictory with Lemma 2. So variables whose degree is not greater than $3\kappa\beta \ln n$ are all in clause C.

So $\omega(C) \geq cn/2 - cn/3 = cn/6$. The conclusion is established.

Finally, from Lemma 5 and Theorem 4, we can get Theorem 3. That is unsatisfiable random instances generated by model RB$^{\text{mix}}$ have no resolution proof whose complexity is less than the exponential size.

5. Conclusion

In this paper, we analyze the resolution complexity of a mixed constraint satisfaction problem named model RB$^{\text{mix}}$. It is theoretically proved that unsatisfiable random instances generated by model RB$^{\text{mix}}$ almost have no complexity whose resolution length is smaller than the exponential size. As a result, based on the resolution algorithm the random instance of model RB$^{\text{mix}}$ has complexity of exponential size. The conclusion shows that model RB$^{\text{mix}}$ with different length

constraints not only has accurate satisfiability phase transition phenomenon, but also can generate a large number of hard instances in the phase transition region. Therefore, model RB$^{\text{mix}}$ is a large-domain nontrivial random constraint satisfaction problem with mixed constraints and accurate phase transition. In the later work, we can also verify the hardness of model RB$^{\text{mix}}$ in the experiment, and further study the solving algorithm.

References

[1] Verhoeven, Y. (2002) Random 2-SAT and Unsatisfiability. *Information Processing Letters*, **72**, 119-123.

[2] Bollobás, B., Borgs, C., Chayes, J.T., Kim, J.H. and Wilson, D.B. (2001) The Scaling Window of the 2-SAT Transition. *Random Structures and Algorithms*, **18**, 201-256. https://doi.org/10.1002/rsa.1006

[3] Cook, S. (1971) The Complexity of Theorem-Proving Procedures. *Proceedings of the* 3*rd Annual ACM Symposium on Theory of Computing*, Shaker Heights, OH, 3-5 May 1971, 151-158. https://doi.org/10.1145/800157.805047

[4] Díaz, J., Kirousis, L., Mitsche, D., *et al.* (2009) On the Satisfiability Threshold of Formulas with Three Literals per Clause. *Theoretical Computer Science*, **410**, 2920-2934.

[5] Kaporis, A.C., Kirousis, L.M. and Lalas, E.G. (2002) The Probabilistic Analysis of a Greedy Satisfiability Algorithm. In: Möhring, R. and Raman, R., Eds., *Algorithms—ESA* 2002, *Lecture Notes in Computer Science*, Vol. 2461,. Springer, Berlin, Heidelberg, 574-586. https://doi.org/10.1007/3-540-45749-6_51

[6] Mertens, S., Mézard, M. and Zecchina, R. (2006) Threshold Values of Random K-SAT from the Cavity Method. *Random Structures & Algorithms*, **28**, 340-373. https://doi.org/10.1002/rsa.20090

[7] Monasson, R., Zecchina, R., Kirkpatrick, S., *et al.* (1996) Phase Transition and Search Cost in the (2 + p)-SAT Problem. *Proceedings of the* 4*th Workshop on Physics and Computation*, Boston, 229-232.

[8] Achlioptas, D., Kirousis, L.M., Kranakis, E., *et al.* (2001) Rigorous Results for Random (2 + p)-SAT. *Theoretical Computer Science*, **265**, 109-129.

[9] Gent, I.P., Macintyre, E., Prosser, P., Smith, B.M. and Walsh, T. (2001) Random Constraint Satisfaction: Flaws and Structure. *Constraints*, **6**, 345-372. https://doi.org/10.1023/A:1011454308633

[10] Smith, B.M. and Dyer, M.E. (1996) Locating the Phase Transition in Binary Constraint Satisfaction Problems. *Artificial Intelligence*, **81**, 155-181.

[11] Xu, K. and Li, W. (2000) Exact Phase Transitions in Random Constraint Satisfaction Problems. *Journal of Artificial Intelligence Research*, **12**, 93-103.

[12] Xu, K. and Li, W. (2006) Many Hard Examples in Exact Phase Transitions. *Theoretical Computer Science*, **355**, 291-302.

[13] Xu, K., Boussemart, F., Hemery, F., *et al.* (2000) Random Constraint Satisfaction Problems. *Journal of Artificial Intelligence Research*, **12**, 93-103.

[14] Cai, S., Su, K. and Chen, Q. (2010) EWLS: A New Local Search for Minimum Vertex Cover. In: Fox, M. and Poole, D., Eds., *Program Cochairs*: *Proceedings of the Twenty-fourth AAAI Conference on Artificial Intelligence* (*AAAI*-10), The AAAI Press, Menlo Park, 45-50.

[15] Zhao, C. and Zheng, Z. (2011) Threshold Behaviors of a Random Constraint Satis-

faction Problem with Exact Phase Transitions. *Information Processing Letters*, **111**, 985-988.

[16] Shen, J. (2011) Model Structure and Phase Transition Phenomenon of Constraint Satisfaction Problem. Doctoral Dissertation, Central China Normal University, Wuhan.

[17] Fan, Y. and Shen, J. (2011) On the Phase Transitions of Random k-Constraint Satisfaction Problems. *Artificial Intelligence*, **175**, 914-927.

[18] Shen, J. and Ren, Y. (2016) Bounding the Scaling Window of Random Constraint Satisfaction Problems. *Journal of Combinatorial Optimization*, **31**, 786-801.
https://doi.org/10.1007/s10878-014-9789-y

[19] Zhao, C., Zhou, H., Zheng, Z., *et al.* (2011) A Message-Passing Approach to Random Constraint Satisfaction Problems with Growing Domains. *Journal of Statistical Mechanics: Theory & Experiment*, **2011**, P02019.
https://doi.org/10.1088/1742-5468/2011/02/P02019

[20] Zhao, C. and Zheng, Z. (2012) A Belief Propagation Algorithm of Solving the Constraint Satisfaction Problem Based on the Variable Entropy. *Chinese Science: Information Science*, **42**, 1170-1180.

[21] Zhao, C., Zhang, P., Zheng, Z., *et al.* (2012) Analytical and Belief Propagation Studies of Random Constraint Satisfaction Problems with Growing Domains. *Physical Review E*, **85**, 016-106.

[22] Yuan, Z. and Zhao, C. (2017) Solve Large-Domain Constraint Satisfaction Problem with Two Kinds of Revised Simulated Annealing Algorithm. *Application Research of Computers*, 12.

[23] Ben-Sasson, E. and Wigderson, A. (2001) Short Proofs Are Narrow-Resolution Made Simple. *Journal of the ACM*, **48**, 149-169.
https://doi.org/10.1145/375827.375835