Exact Phase Transitions and Approximate Algorithm of #CSP

Ping Huang[1], Minghao Yin[1*], Ke Xu[2*]

[1]School of Computer Science and Information Technology, Northeast Normal University, Changchun, P. R. China, 130117
[2]School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing, P. R. China, 100083
huangp218@nenu.edu.cn, ymh@nenu.edu.cn, kexu@nlsde.buaa.edu.cn

Introduction

In the past decade there has been a significant interest in the phase transition of NP-complete or NP-hard problems. However, it seems difficult to obtain the location of the exact phase transition point. (Xu and Li 2000) may be one of the few works that can prove the existence of phase transition and identified the phase transition points exactly. By introducing a revision of the standard CSP random model(Model B) in (Gent el. al. 2001), as we called Model RB, we can prove that the critical value of the phase transition point can be quantified. Moreover, Model RB provides a framework for generating asymptotically hard random constraint satisfaction problems and therefore has been widely used as benchmarks to evaluate the asymptotic behavior of CSP algorithms (Xu el. al. 2007). In this literature, we follow this line of research to study phase transition of counting the solutions of CSP instances following Model RB. Specifically, we consider a decision version of #CSP, called #CSP ($\geq d^{n/t}$). That is, deciding whether the instance has at least $d^{n/t}$ satisfying assignments. Note that #CSP($\geq d^{n/t}$) can be viewed as a generalization of #3SAT($\geq 2^{n/t}$), which was studied in (Bailey el. al. 2007). So #CSP($\geq d^{n/t}$) is at least PP-hard. The contribution of our work is as following:

1) We prove the existence of phase transition in Model RB for #CSP($\geq d^{n/t}$) can be guaranteed, and the threshold point can be precisely located rather than in the form of some loose but hard won bounds, for instance (Bailey el. al. 2007).
2) A careful analysis of phase transition can lead us to develop an approximate algorithm to estimate the solutions number in Model RB. Unlike other approximate algorithms, the accuracy of our algorithm increases with the increase of the problem scale.

Partially supported by the NSFC grants (60803102 and 60973033). Correspondence should be given to Minghao Yin (ymh@nenu.edu.cn) and Ke Xu (kexu@nlsde.buaa.edu.cn).

Model RB

In this section, we recall the concept of Model RB. For more details we refer to (Xu and Li 2000, 2006; Xu el. al. 2005, 2007).

Definition 1 (*Model RB*) *In Model RB, a class of random CSP instances is determined by parameters* (k, n, α, r, p) *where, for each instance:*
1) $k\geq 2$ denotes the arity of each constraint,
2) $n\geq 2$ denotes the number of variables,
3) $\alpha>0$ determines the domain size $d=n^\alpha$ of each domain,
4) $r>0$ determines the number $m=r\cdot n\cdot\ln n$ of constraints,
5) $1>p>0$ determines the number $t=pd^k$ of disallowed tuples of each relation.

The main difference between Model RB and Model B is that the domain size in Model RB grows with the number of variables. The generation of random CSP instance in Model RB is done as follows:

(1) Select m= $r\cdot n\cdot\ln n$ random constraints (with repetition), each one formed by randomly selecting k of n variables (without repetition).

(2) For each constraint select $t = pd^k$ incompatible tuples of values (without repetition). i.e., each constraint relation contains exactly $(1-p)d^k$ compatible tuples of values.

Phase transitions of #CSP($\geq d^{n/t}$)

Let Pr be the probabilistic distribution and let $X_{r,p}^{n,k,\alpha}$ denote the number of solutions of the instance generated following Model RB. The following theorems can be proved:

Theorem 1 If $p_{cr}=1-e^{-\frac{\alpha}{r}(1-\frac{1}{t})}$,where $\alpha>1/k$, r>0 are two constraints, k, α *and* r satisfy the inequality $k\cdot e^{-\alpha/r}\geq 1$, then

$$\lim_{n\to\infty}\Pr[X_{r,p}^{n,k,\alpha}\geq d^{n/t}]=1, when\ p<p_{cr}$$

$$\lim_{n\to\infty}\Pr[X_{r,p}^{n,k,\alpha}\geq d^{n/t}]=0, when\ p>p_{cr}$$

Theorem 1 uses constraint tightness as the parameter, when using constraint density as the parameter, we get a similar result.

Theorem 2 If $r_{cr}=-\frac{\alpha(1-1/t)}{\ln(1-p)}$, $\alpha>\frac{1}{k}$ and $0<p<1$ are two constraints satisfy $k \geq 1/(1-p)$,then

$$\lim_{n\to\infty} \Pr[X_{r,p}^{n,k,\alpha} \geq d^{n/t}] = 1 \quad \text{when } r<r_{cr}$$

$$\lim_{n\to\infty} \Pr[X_{r,p}^{n,k,\alpha} \geq d^{n/t}] = 0 \quad \text{when } r>r_{cr}$$

For the page limits, we omit the proof procedures of the theorems here. Note that when t goes to infinity, the problem of #CSP($\geq d^{n/t}$) reduces to the problem of deciding satisfiability of CSP, and then the theorems proposed in this paper reduce to theorem 1 and theorem 2 in (Xu and Li 2000) as well. So the theorems proposed in this paper can be viewed as a generalization of those proposed in (Xu and Li 2000). Moreover, the above two theorems hold when $n\to\infty$, but preliminary experiments have shown that exact phase transition do occur even if *n* is small.

On Estimating Solutions Number of CSPs

In this section we introduce how to evaluate solution numbers of CSPs in Model RB. According to theorem 1, when $p>1-e^{-\alpha/r}$, the number of solutions approaches 0. So we only need to consider conditions when $p<1-e^{-\alpha/r}$. According to the definition 1, it is easy to prove that the expected number of solutions E($X_{r,p}^{n,k,\alpha}$) of the instance can be calculated as:

$$\mathrm{E}(\, X_{r,p}^{n,k,\alpha}\,) = d^n (1-p)^{rn\ln n} = n^{\alpha n}(1-p)^{rn\ln n}\,.$$

Then the following theorem shows that E($X_{r,p}^{n,k,\alpha}$) can provide good lower and upper bounds for Model RB.

Theorem 3 Given a CSP instance *I* randomly generated following Model RB and *k, α and r* satisfy the inequality $ke^{-\alpha/r}\geq 1$ and $\alpha>1/k$. Let $X_{r,p}^{n,k,\alpha}$ be the number of solutions for *I*, E($X_{r,p}^{n,k,\alpha}$) denote the expected number of solution for *I*, δ be an arbitrary real number, and *Pr* denote the function for probability distribution. As the number of variable goes to infinity, if $p<1-e^{-\alpha/r}$ then

$$\underset{n\to\infty}{Lim}(Pr((1\text{-}\delta)\mathrm{E}(X_{r,p}^{n,k,\alpha}) < X_{r,p}^{n,k,\alpha} < (1+\delta)\mathrm{E}(X_{r,p}^{n,k,\alpha}))) \approx 1$$

α	r	p	n	[0.5,1.5]	[0.4,1.6]	[0.3,1.7]	[0.2,1.8]	[0.1,1.9]
0.7	2.3	0.2	13	23.33%	26.67%	26.67%	33.33%	36.67%
0.7	2.3	0.2	14	40.00%	43.33%	60.00%	70.00%	76.67%
0.7	2.3	0.2	15	36.67%	43.33%	50.00%	50.00%	50.00%
0.7	2.3	0.2	16	50.00%	70.00%	70.00%	76.67%	83.33%
0.8	1.5	0.3	7	46.67%	63.33%	76.67%	83.33%	83.33%
0.8	1.5	0.3	9	66.67%	76.67%	76.67%	83.33%	83.33%
0.8	1.5	0.3	11	63.33%	70.00%	76.67%	83.33%	83.33%
0.8	1.5	0.3	13	36.67%	40.00%	43.33%	53.33%	56.67%
0.9	2.1	0.3	11	6.67%	6.67%	13.33%	13.33%	13.33%
0.9	2.1	0.3	12	50.00%	70.00%	80.00%	86.67%	90.00%
0.9	2.1	0.3	13	66.67%	70.00%	70.00%	70.00%	73.33%
0.9	2.1	0.3	14	33.33%	40.00%	43.33%	46.67%	53.33%
1	2	0.35	11	46.67%	50.00%	63.33%	63.33%	66.67%
1	2	0.35	13	53.33%	60.00%	80.00%	83.33%	86.67%
1	2	0.35	15	66.67%	66.67%	70.00%	83.33%	86.67%
1	2	0.35	17	66.67%	70.00%	73.33%	80.00%	83.33%

Table 1 Accuracy of AE-count (for each point 300 instances are generated)

From above discussion, we can easily develop an approximate algorithm for #CSP in linear time, as we called AE-count. The only point where AE-count cannot work is when $p = p_{cr}$. It should be pointed out that the short coming of AE-count is that it is only capable of estimating random instances generated following Model RB, and can only return the estimate numbers of solutions rather enumerating the solutions. However, since we can generate large-scale CSP instances or SAT instances following Model RB, and simultaneously obtain an upper bound and lower bound of their number of solutions, we can use Model RB to generate benchmarks for counting algorithms. As can be seen in Table 1, even if the problems scales are relatively small, the estimates are found to be over 69% correct for δ*=0.9*, and the accuracy of the estimates grows with the increasing of problem scales. Note that we only make experiments on "SAT" region (former experiments in (Xu and Li 2000) have shown that no solutions exist in UNSAT region). So the real accuracy should be higher.

Conclusion

In this paper, we first present a probabilistic analysis of the problem #CSP. We show that for random instances generated following Model RB, exact phase transitions do exist for a decision version of #CSP, i.e. #CSP($\geq d^{n/t}$). Then, preliminary experimental results have confirmed the phase transition and threshold predicted by theory. Second, we present an accurate estimate of the number of solutions of random CSP instances generated following Model RB.

References

Baileya, D. D., Dalmaub, V., and Kolaitisa, P. G. 2007. Phase transitions of PP-complete satisfiability problems. *Discrete Applied Mathematics* 155: 1627-1639.

Bayardo Jr., R. J., and Pehoushek, J. D. 2000. Counting models using connected components. In *Proceedings of AAAI-00*, 157-162.

Gent I., MacIntyre E., Prosser P., Smith B. M., Walsh T.. 2001. Random Constraint Satisfaction: Flaws and Structure. *Constraints* 6(4): 345-372.

Gomes, C. P., Hoeve, W. J. V., Sabharwal, A., and Selman B. 2007. Counting CSP Solutions Using Generalized XOR Constraints. In *Proceedings of AAAI'07*, 204-209.

Gomes, C. P., Sabharwal A., Selman B. 2006. Model Counting: A New Strategy for Obtaining Good Bounds. In *Proceedings of AAAI 2006*, 54-61.

Sang, T., Bacchus F., Beame P., Kautz H.A. and Pitassi T. 2004. Combining Component Caching and Clause Learning for Effective Model Counting. *SAT 2004*.

Xu, K., Boussemart, F.; Hemery, F.; and Lecoutre C. 2007. Random constraint satisfaction: Easy generation of hard (satisfiable) instances. *Artificial Intelligence* 171(8-9)..

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

Xu, K., and Li, W. 2006. Many Hard Examples in Exact Phase Transitions, *Theoretical Computer Science* 355: 291-302.

Wei W., Selman B. 2005. A New Approach to Model Counting. In *Proceedings of SAT*, 324-339.