

随机正则(k, r)-SAT 问题的可满足临界^{*}

周锦程^{1,2}, 许道云¹, 卢友军¹

¹(贵州大学 计算机科学与技术学院, 贵州 贵阳 550025)

²(黔南民族师范学院 数学与统计学院, 贵州 都匀 558000)

通讯作者: 许道云, E-mail: dyxu@gzu.edu.cn



摘要: 研究 k -SAT 问题实例中每个变元恰好出现 $r=2s$ 次, 且每个变元对应的正、负文字都出现 s 次的严格随机正则(k, r)-SAT 问题. 通过构造一个特殊的独立随机实验, 结合一阶矩方法, 给出了严格随机正则(k, r)-SAT 问题可满足临界值的上界. 由于严格正则情形与正则情形的可满足临界值近似相等, 因此得到了随机正则(k, r)-SAT 问题可满足临界值的新上界. 该上界不仅小于当前已有的随机正则(k, r)-SAT 问题的可满足临界值上界, 而且还小于一般的随机 k -SAT 问题的可满足临界值. 因此, 这也从理论上解释了在相变点处的随机正则(k, r)-SAT 问题实例通常比在相应相变点处同规模的随机 k -SAT 问题实例更难满足的原因. 最后, 数值分析结果验证了所给上界的正确性.

关键词: 随机正则(k, r)-SAT 问题; 可满足临界值; 相变现象; 计算复杂性

中图法分类号: TP301

中文引用格式: 周锦程, 许道云, 卢友军. 随机正则(k, r)-SAT 问题的可满足临界. 软件学报, 2016, 27(12): 2985–2993. <http://www.jos.org.cn/1000-9825/5129.htm>

英文引用格式: Zhou JC, Xu DY, Lu YJ. Satisfiability threshold of the regular random (k, r)-SAT problem. Ruan Jian Xue Bao/ Journal of Software, 2016, 27(12): 2985–2993 (in Chinese). <http://www.jos.org.cn/1000-9825/5129.htm>

Satisfiability Threshold of the Regular Random (k, r)-SAT Problem

ZHOU Jin-Cheng^{1,2}, XU Dao-Yun¹, LU You-Jun¹

¹(College of Computer Science and Technology, Guizhou University, Guiyang 550025, China)

²(School of Mathematics and Statistics, Qiannan Normal University for Nationalities, Duyun 558000, China)

Abstract: This article studies the strictly regular (k, r)-SAT problem by restricting the k -SAT problem instances, where each variables occurs precisely $r = 2s$ times and each of the positive and negative literals occurs precisely s times. By constructing a special independent random experiment, the study derives an upper bound on the satisfiability threshold of the strictly regular random (k, r)-SAT problem via the first moment method. Based on the fact that the satisfiability threshold of the strictly regular and the regular random (k, r)-SAT problems are approximately equal, a new upper bound on the threshold of the regular random (k, r)-SAT problem is obtained. This new upper bound is not only below the current best known upper bounds on the satisfiability threshold of the regular random (k, r)-SAT problem, but also below the satisfiability threshold of the uniform random k -SAT problem. Thus, it is theoretically explained that in general the regular random (k, r)-SAT instances are harder to satisfy at their phase transition points than the uniform random k -SAT problem at the corresponding phase transition points with same scales. Finally, numerical results verify the validity of our new upper bound.

Key words: regular random (k, r)-SAT problem; satisfiability threshold; phase transition phenomena; computational complexity

* 基金项目: 国家自然科学基金(61262006, 61463044, 61462001); 贵州省科技厅联合基金(LKQS201313)

Foundation item: National Natural Science Foundation of China (61262006, 61463044, 61462001); Science and Technology Foundation of Guizhou Province of China (LKQS201313)

收稿时间: 2016-07-05; 采用时间: 2016-08-18; jos 在线出版时间: 2016-10-11

CNKI 网络优先出版: 2016-10-12 16:27:01, <http://www.cnki.net/kcms/detail/11.2560.TP.20161012.1627.029.html>

给定含有 N 个布尔变元的合取范式(conjunctive normal form,简称 CNF)公式 F ,可满足性问题(the satisfiability problem,简称 SAT 问题)是指:是否存在一组对所有 N 个布尔变元的真值指派 $\sigma \in \{0,1\}^N$,使得公式 F 的取值为 TRUE.在 SAT 问题中,限制每个子句长度为 k 的 SAT 问题称为 k -SAT 问题.该问题是判定由 N 个布尔变元 $\{v_1, v_2, \dots, v_N\}$ 、 M 个子句 $\{C_1, C_2, \dots, C_M\}$ 构成的合取范式 $F=C_1 \wedge C_2 \wedge \dots \wedge C_M$ 的可满足性问题.其中,每个子句 C_i 是由 k 个不同文字构成的析取范式 $C_i=(\ell_{i1} \vee \ell_{i2} \vee \dots \vee \ell_{ik})$,而文字 ℓ 是某个变元 v 或其否定形式 $\neg v$.当 $k \geq 3$ 时, k -SAT 问题是第一个被证明了的 NP-complete 问题^[1].因此在最坏情形下,通常认为该问题没有有效的求解算法.此外,研究者以 SAT 问题的 NP-complete 性质为种子,利用多项式规约转换等技术,已证明了大量的组合优化问题都是 NP-complete 问题^[2].因此,SAT 问题,特别是 k -SAT 问题,仍然是当前理论计算机科学研究领域的一个核心问题.

随机 k -SAT 问题作为 k -SAT 问题的子集合,其在 k -SAT 问题的典型计算复杂性研究中发挥着重要作用.在随机 k -SAT 问题中,当变元规模 $N \gg 1$ 时,一个重要的结构参数是子句个数 M 与变元个数 N 的比值 α (也称约束密度).已有的理论和实验研究结果表明:该参数不仅能够影响到公式的判定难度,还与公式的可满足性密切相关^[3,4].具体地,当 $N \gg 1$ 时,随着 α 的逐渐增大,存在某个与 k 相关的临界值点(threshold point) $\alpha_s(k)$,当随机公式 F 的约束密度满足 $\alpha > \alpha_s(k)$ 时,高概率地 F 是不可满足的,而当随机公式 F 的约束密度满足 $\alpha < \alpha_s(k)$ 时,高概率地 F 是可满足的,这种现象称为随机 k -SAT 问题的相变(phase transition)现象,而临界值点 $\alpha_s(k)$ 被称为随机 k -SAT 问题的相变点.此外,统计物理学中的一阶复本对称破缺(one step replica symmetry breaking,简称 1RSB)理论研究表明:在紧邻相变点 $\alpha_s(k)$ 前的某个位置,随机 k -SAT 问题开始呈现簇集相变(clustering threshold)^[5],从簇集相变点 $\alpha_d(k)$ 处开始, k -SAT 问题的解空间将会突然分裂成数目众多的解集簇,而这些解集簇之间相距甚远,且在解集簇内部,大量变元被凝固^[6-8].因此,若仅仅通过翻转某个解集簇中一个解的少量变元的赋值,不太可能将其转化为另一个解集簇中的解,所以对于临界值点 $\alpha_s(k)$ 附近的实例,现有的 k -SAT 求解算法均无法高效地求解,即便是采用当前求解 SAT 问题最为有效的概观传播(survey propagation,简称 SP)算法^[9,10],其在求解 $\alpha_s(k)$ 附近的 k -CNF 实例时也往往容易失效.另外,在远离相变点 $\alpha_s(k)$ 的两侧,绝大部分实例都是易于判定的.因此,研究 SAT 问题的相变现象将有利于更深入地认识 NP-complete 问题的难解本质和设计更为有效的 SAT 问题求解算法.然而,要找出该问题的精确相变点却是非常困难的.

当前已知的具有精确相变点的 SAT 问题主要包括:2-SAT^[11],Regular 2-SAT^[12], k -NAESAT^[13], k -XORSAT^[14] 和 Regular NAE-SAT^[15] 等几种具有特殊规则结构的 SAT 子类.此外,文献[16]采用 1RSB 理论预言随机 k -SAT 问题的相变点 $\alpha_s(k)$ 为 $\alpha_s(k)=2^k \ln 2 - (\ln 2 + 1)/2 + o_k(1)$.文献[17,18]分别通过寻找随机 k -SAT 问题中解的聚类,结合矩方法证明:当 k 充分大时, $\alpha_s(k)$ 的渐近值与文献[16]的预测是相吻合的.

此外,为使 SAT 问题的研究更为具体,研究者通过对 SAT 问题的结构加以某些限制,从而得到具有一定规则结构并保留 NP-complete 性质的公式子类.如:限制子句长度为 k 的 k -SAT 问题;在 k -SAT 问题的基础上提出的每个变元至多出现 s 次的 k -SAT 问题^[19];特别地,文献[12]提出的每个变元恰好出现 r 次且每个变元正、负出现的期望次数至多相差 1 次的平衡(k,r)-SAT 问题,因其比一般 k -SAT 问题更难计算而受到研究者广泛关注^[20-24].在此基础上,文献[20]研究了每个变元恰好出现 r 次且每个变元正、负出现次数至多相差 1 次的正则(k,r)-SAT 问题,并通过矩方法给出了该问题可满足临界的上下界.本文将在文献[20]的基础上,给出该问题可满足临界值的改进上界.

第 1 节介绍相关工作的研究现状.第 2 节给出随机正则(k,r)-SAT 问题的相关定义及严格随机正则(k,r)-SAT 问题的实例产生模型.第 3 节通过构造特殊的独立随机实验,结合一阶矩方法,给出严格随机正则(k,r)-SAT 问题可满足临界值的上界.由于严格正则情形与正则情形的可满足临界值近似相等,从而得到了随机正则(k,r)-SAT 问题可满足临界值的新上界.第 4 节通过数值分析结果验证所给上界的正确性.最后总结全文的工作.

1 随机正则(k,r)-SAT 问题可满足临界的研究现状

当前,在随机正则(k,r)-SAT 问题的相变性质研究方面,文献[12]首先证明了($3,r$)-SAT 问题的可满足临界值点 $\alpha_{rs}(3)$ 满足 $2.46 \leq \alpha_{rs}(3) \leq 3.78$.文献[21]在敌手可满足性问题(adversarial satisfiability problem)的研究中表明:

针对随机(3,r)-CNF 公式,当 $r > 11$ 时,信念传播(belief propagation,简称 BP)算法在公式实例对应的因子图上收敛并会导致矛盾.即:当 $r > 11$ 时,随机(3,r)-CNF 实例是高概率地不可满足的;当 $r = 10$ 和 $r = 11$ 时,BP 算法在其对应的因子图上不收敛且 SP 算法收敛于非平凡的固定点;而当 $r < 10$ 时,BP 算法收敛于固定点,即高概率地,一个随机(3,r)-CNF 实例是可满足的.因此,随机正则(3,r)-SAT 问题的相变点可能发生在 $r = 10$ 或 $r = 11$ 处.文献[22]通过引入一种特殊的树形结构,并将正则因子图转换到树型结构上进行研究,证明了若 r 为偶数时,簇集相变点可能发生在 $r = 8$ 或者 $r = 10$ 处.进一步,文献[23]首先通过将正则因子图转换成 Cayley 树,然后再将其转换为相应的 Bethe 晶格(lattice)进行研究,并证明了当 r 可以不取整数时,簇集相变点 $\alpha_{rd}(3) \sim 3.23$,可满足相变点 $\alpha_{rs}(3) \sim 3.6$.文献[24]结合概率生成函数的系数近似技术,采用一阶矩方法严格证明了当 $r > 11$ 时,随机生成的正则(3,r)-CNF 公式是高概率地不可满足的,并且通过实验分析表明,该问题的可满足相变点恰好发生在 $r = 11$ 处.另外,文献[24]的实验研究还表明,随机正则(3,r)-SAT 问题在相变点 $r = 11$ ($\alpha_{rs}(3) \sim 3.6667$) 处的随机实例比通常的同变元规模的均匀随机 3-SAT 问题在相变点 $\alpha_s(3) \sim 4.2667$ 处^[9]的随机实例更难求解.此外,文献[20]通过一阶矩和二阶矩方法证明了存在某个常量 k_0 ,当 $k \geq k_0$ 时,该问题的可满足临界值点 $\alpha_{rs}(k)$ 的上下界满足: $2^k \ln 2 - (k+1) \ln 2 / 2 - 1 \leq \alpha_{rs}(k) \leq 2^k \ln 2$,其上下界之间的间隙为 $(k+1) \ln 2 / 2 + 1$.

本文首先考虑随机正则(k,r)-SAT 问题中每个变元出现次数 $r = 2s$ 次且每个变元的正、负出现次数都恰为 s 次的情形,我们称这种情形下的 k -SAT 问题为严格随机正则(k,r)-SAT 问题.通过构造特殊的独立随机实验,结合一阶矩方法,我们将证明存在某个常数 k_1 ,当 $k \geq k_1$ 时,该问题可满足临界值的上界 $\alpha_{ru}(k)$ 为

$$\alpha_{ru}(k) = 2^k \ln 2 - (k-2) \ln 2 / 2.$$

当严格随机正则(k,r)-CNF 公式 F 的约束密度满足 $\alpha_r \geq \alpha_{ru}(k)$ 时,高概率地 F 是不可满足的.进一步,由于严格正则和正则情形的可满足临界值是近似相等的,由此,结合文献[20]所给出的此问题的下界,我们得到了随机正则(k,r)-SAT 问题可满足临界值点 $\alpha_{rs}(k)$ 的更为紧致的界,即: $2^k \ln 2 - (k+1) \ln 2 / 2 - 1 \leq \alpha_{rs}(k) \leq 2^k \ln 2 - (k-2) \ln 2 / 2$,且上下界之间仅间隙一个常数 $1 + 3 \ln 2 / 2$.

2 问题描述与实例生成模型

2.1 基本概念

正则(k,r)-SAT 问题^[20]是限制 k -SAT 问题中每个变元恰好出现 r 次,且每个变元的正、负出现次数至多相差 1 次的 k -SAT 问题子类.为了便于相关的理论分析,我们考虑每个变元恰好出现 $r = 2s$ ($s \in \mathbb{Z}^+$) 次且每个变元的正、负出现次数都恰为 s 次的 k -SAT 问题,并称其为严格正则(k,r)-SAT 问题.由严格正则(k,r)-SAT 问题的定义易知,严格(k,r)-CNF 公式的子句约束密度 α_r 满足 $\alpha_r = M/N = r/k = 2s/k$.严格随机正则(k,r)-CNF 公式是均匀地从所有的严格正则(k,r)-CNF 公式上选取的随机实例,其对应的因子图^[25]可以用双正则二部图来表示,其中,二部图中的一侧由公式中的子句集构成,而另一侧则由公式中的变元集构成.通常用矩形框表示子句节点,用实心圆点来表示变元节点.若某个变元 v_i 在子句 C_j 中正出现,则用实边连接 v_i 与 C_j ;否则,采用虚边连接.

图 1 给出了严格正则(3,6)-CNF 公式 $F = C_1 \wedge C_2 \wedge \dots \wedge C_6$ 的双正则二部图,其中, $C_1 = (\neg v_1 \vee v_2 \vee \neg v_3)$, $C_2 = (\neg v_1 \vee \neg v_2 \vee v_3)$, $C_3 = (v_1 \vee \neg v_2 \vee v_3)$, $C_4 = (\neg v_1 \vee v_2 \vee v_3)$, $C_5 = (v_1 \vee \neg v_2 \vee \neg v_3)$, $C_6 = (v_1 \vee v_2 \vee \neg v_3)$,即: F 中每个变元恰好出现 6 次,且每个变元对应的正、负文字恰好出现 3 次.

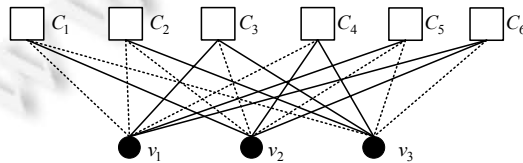


Fig.1 Bipartite graph representation of the strictly regular (3,6)-CNF formula F

图 1 严格正则(3,6)-CNF 公式 F 的双正则二部图表示

2.2 严格随机正则(k,r)-SAT问题实例生成模型

对于给定的约束密度 α 、子句长度 k 和变元规模 N ,一般的均匀随机 k -SAT问题实例的生成模型是指:从所有 $C_N^k \cdot 2^k$ 个长度为 k 的可能子句中独立并且均匀地随机选取 $M=\alpha N$ 个子句构成的随机 k -CNF公式.

在严格正则(k,r)-SAT问题中,由于子句约束密度 $\alpha_r=r/k$,因此我们需要给定的参数包括:变元出现次数 r (r 为正偶数)、子句长度 k 和变元规模 N .在随机正则(k,r)-SAT问题实例生成模型中,文献[24]所提出的生成随机($3,r$)-SAT问题难解实例的SRR模型极易扩展为生成严格随机(k,r)-SAT问题实例模型,但由于SRR模型为了防止在生成随机($3,r$)-SAT问题难解实例过程中产生非法子句,从而对实例的生成过程进行了一定的干预,这将会增加我们分析随机公式的难度.因此,本文采用文献[12]所提出的格局模型来生成严格随机正则(k,r)-CNF公式 F ,具体生成算法如下:

Input:变元出现次数 r ,子句长度 k 和变元规模 N ;

Step 1. 依次对文字集 $L=\{v_1, \neg v_1, \dots, v_N, \neg v_N\}$ 中的每个正文字 v_i 和负文字 $\neg v_i$,分别创建 $s=\frac{r}{2}$ 个拷贝

$$v_i^1, v_i^2, \dots, v_i^s, \neg v_i^1, \neg v_i^2, \dots, \neg v_i^s, \text{其中}, i \in [N].$$

Step 2. 对Step 1中的所有 rN 个文字随机生成一个投影 $\pi: L \times [s] \rightarrow [M] \times [k]$.

Step 3. 置第 i 个子句中的第 j 个文字 $F_{ij}=\pi(i,j)$,其中, $i \in [M], j \in [k]$.

Step 4. 输出公式 F .

事实上,该模型所生成的随机公式中,可能会出现某个变元在某个子句中出现多次的非法情形.此外,该模型生成的随机公式还有可能产生重复子句,但文献[12]的研究表明:存在常数 $\delta > 0$,当 $N \rightarrow \infty$ 时, $\Pr(F \text{ is legal}) \rightarrow \delta$,且若合法的严格随机正则公式的可满足临界值存在,则它与相应的格局公式的可满足临界值相同.因此,为证明严格随机正则公式的可满足临界值的界,我们仅需要证明相应的格局公式的可满足临界值的界即可.

3 随机正则(k,r)-SAT问题可满足临界

下面给出第3.1节中将用到的独立随机变元和的局部极限准则(the local limit law)^[26].

引理 1(局部极限准则). 令 Z_1, \dots, Z_N 是支撑 $\mathbb{Z}_{\geq 0}$ 上的 N 个独立随机变元,且 $G(z)$ 为其概率生成函数.令 $\mu=E[Z_i], \sigma^2=Var[Z_i]$,如果 $G(z)$ 是非周期的全函数且对所有的 $T_0 < \alpha < T_\infty$,当 $N \rightarrow \infty$ 时,有 $T_x = \lim_{z \rightarrow x} zG(z)/G'(z)$,则有:

$$\Pr\left[\sum_{i=1}^N Z_i = \alpha N\right] = \frac{1}{\sqrt{2\pi N\xi}} G(\zeta)^N \cdot \zeta^{-\alpha N - 1} (1 + o(1)) \quad (1)$$

其中, ζ 和 ξ 为下面两个方程的解:

$$\frac{\zeta G'(\zeta)}{G(\zeta)} = \alpha, \xi = \frac{d^2}{dz^2} (\ln G(z) - \alpha \ln(z))|_{z=\zeta} \quad (2)$$

3.1 严格随机正则(k,r)-SAT问题可满足临界值的上界

本节将构造一个特殊的独立随机实验,并结合一阶矩方法来证明严格随机正则(k,r)-SAT问题可满足临界值的上界.

若 Z 为非负整数随机变量且其期望为 $E[Z]$,则对任意的实数 $c > 0$,由马尔科夫不等式 $\Pr(Z \geq c) \leq \frac{E[Z]}{c}$ 知随机变元 $Z \geq 1$ 的概率至多为 $E[Z]$,此即一阶矩方法,即:

$$\Pr(Z \geq 1) \leq E[Z] \quad (3)$$

若 S 表示严格随机正则(k,r)-CNF公式 F 的解空间,则 $\Omega=|S|$ 为 F 的可满足解的总数目,则由公式(3)有:

$$\Pr(F \text{ is SAT}) = \Pr(\Omega > 0) \leq E[\Omega] \quad (4)$$

对于 F 中变元集 $V=\{v_1, v_2, \dots, v_N\}$ 的任意指派 $\sigma(v_1, v_2, \dots, v_N)$,令 \mathcal{A} 表示事件:指派 $\sigma(v_1, v_2, \dots, v_N)$ 满足公式 F ;令 \mathcal{B} 表示事件:指派 $\tau=1=(1, 1, \dots, 1)$ 满足公式 F .因为严格正则公式中的每个正、负文字都恰好出现 s 次,所以对变元

集 \$V\$ 的任何指派 \$\sigma\$ 都会使得恰有 \$\frac{1}{2}\$ 的文字取值为 TRUE,因此,任何 \$\sigma\$ 能成为一个可满足解的概率都是相同的,即有 \$\Pr(\mathcal{A})=\Pr(\mathcal{B})=\Pr(F \text{ is SAT by } \sigma)=\Pr(F \text{ is SAT by } \mathbf{1})\$ 成立.由于 \$N\$ 个变元的所有指派数共有 \$2^N\$ 个,所以结合公式(3)所给出的一阶矩方法有:

$$\Pr(F \text{ is SAT}) \leq E[\Omega] = \sum_{\sigma} \Pr(F \text{ is SAT by } \sigma) = 2^N \Pr(F \text{ is SAT by } \mathbf{1}) \quad (5)$$

为计算 \$\Pr(F \text{ is SAT by } \mathbf{1})\$,我们首先考虑这样的随机实验:在不考虑每个文字所对应的具体变元,而仅仅区分每个文字的正、负情形下,随机地分配正、负文字到所有 \$M\$ 个子句中.由于在指派 \$\tau=\mathbf{1}\$ 下所有正文字 \$v_i^1, v_i^2, \dots, v_i^s\$ 的取值均为 TRUE,而所有负文字 \$\neg v_i^1, \neg v_i^2, \dots, \neg v_i^s\$ 的取值均为 FALSE,其中, \$i \in [M]\$.因此, \$F\$ 是可满足的当且仅当每个子句中至少存在一个正文字.

我们用独立随机变元 \$F_{ij}(i \in [M], j \in [k])\$ 来表示第 \$i\$ 个子句中的第 \$j\$ 个文字的正、负取值,使得 \$F_{ij} \in \{\text{POSTTIVE}, \text{NEGATIVE}\}\$ 且 \$\Pr(F_{ij}=\text{POSTTIVE})=p\$,其中, \$F_{ij}=\text{POSTTIVE}\$ 表示第 \$i\$ 个子句中的第 \$j\$ 个文字为正文字,相应地, \$F_{ij}=\text{NEGATIVE}\$ 表示第 \$i\$ 个子句中的第 \$j\$ 个文字为负文字.如果我们令函数 \$f(s)\$ 为 \$E[\Omega]\$ 的熵密度,即:

$$f(s) = \frac{1}{N} \lim_{N \rightarrow \infty} \ln E[\Omega] \quad (6)$$

则有如下的引理成立:

引理 2. 严格随机正则 \$(k,r)\$-CNF 公式中可满足解的期望数 \$E[\Omega]\$ 的熵密度 \$f(s)\$ 为

$$f(s) \sim \ln 2 + \frac{2s}{k} \ln(2p) - s \ln(4p(1-p)) \quad (7)$$

其中, \$p=p(k)\$ 为方程 \$(1-p)^k + 2p - 1 = 0\$ 的解.

证明:为计算 \$f(s)\$,我们首先计算 \$\Pr(F \text{ is SAT by } \mathbf{1})\$,令指示变元 \$I_{F_{ij}=\text{POSITIVE}}\$ 定义为

$$I_{F_{ij}=\text{POSITIVE}} = \begin{cases} 1, & \text{if } F_{ij} = \text{POSITIVE} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

我们考虑如下两个事件:

- \$\mathcal{C}\$:对严格随机正则公式 \$F\$ 中的任何子句 \$i \in [M]\$,存在某个 \$j \in [k]\$,使得 \$F_{ij}=\text{POSTTIVE}\$;
- \$\mathcal{D}\$:严格随机正则公式 \$F\$ 中的正文字总数恰为 \$sN\$ 个,即: \$\sum I_{F_{ij}=\text{POSITIVE}} = sN\$.

由事件 \$\mathcal{C}\$ 和事件 \$\mathcal{D}\$ 的定义得知, \$\Pr(F \text{ is SAT by } \mathbf{1})\$ 的值与随机实验中事件 \$\mathcal{D}\$ 发生下事件 \$\mathcal{C}\$ 发生的概率是相等的,因此有:

$$\Pr(F \text{ is SAT by } \mathbf{1}) = \Pr(\mathcal{C} | \mathcal{D}) = \frac{\Pr(\mathcal{C}) \cdot \Pr(\mathcal{D} | \mathcal{C})}{\Pr(\mathcal{D})} \quad (9)$$

因为正则公式中的任何子句至少含有 1 个正文字的概率为 \$1-(1-p)^k\$,所以由事件 \$\mathcal{C}\$ 的定义得知: \$M\$ 个子句中每个子句都至少含有 1 个正文字的概率 \$\Pr(\mathcal{C})\$ 为

$$\Pr(\mathcal{C}) = (1-(1-p)^k)^M \quad (10)$$

另外,在事件 \$\mathcal{D}\$ 中,由于随机变量 \$I_{F_{ij}=\text{POSITIVE}}\$ 为伯努利随机变量,因此, \$\sum I_{F_{ij}=\text{POSITIVE}}\$ 的取值服从 \$B(2sN, p)\$ 的二项分布,所以有:

$$\Pr(\mathcal{D}) = \Pr\left(\sum I_{F_{ij}=\text{POSITIVE}} = sN\right) = C_{2sN}^{sN} p^{sN} (1-p)^{sN} \quad (11)$$

由 \$\frac{M}{N} = \frac{2s}{k}\$ 及 Stirling 公式 \$N! \sim \sqrt{2N\pi} \cdot N^N e^{-N}\$ 有:

$$\Pr(\mathcal{D}) = C_{2sN}^{sN} p^{sN} (1-p)^{sN} = \frac{(2sN)!}{(sN)!(sN)!} \cdot p^{sN} (1-p)^{sN} \sim \frac{2^{2sN}}{\sqrt{sN\pi}} \cdot p^{sN} (1-p)^{sN} \quad (12)$$

进一步,我们采用具有 \$M\$ 个随机变量的序列 \$(Y_i)_{i \in [M]}\$ 来统计随机公式 \$F\$ 中每个子句中的正文字数.由于在条件 \$\mathcal{C}\$ 的限制下,每个变量 \$Y_i=j\$ 发生的概率与事件 \$B(k,p)\$ 且 \$j \geq 1\$ 发生的概率相等,其中, \$B(k,p)\$ 表示参数为 \$k,p\$ 的二项分布,令 \$Y = \sum_{i=1}^M Y_i\$,则此时只需选择适当的 \$p\$ 使得 \$Y=E(Y|\mathcal{C})=sN\$ 成立,则有事件 \$\mathcal{D}\$ 发生,即:

$$E(Y | \mathcal{C}) = M \cdot E[Y_i] = \frac{2sN}{k} \cdot \sum_{j=1}^k \frac{j \cdot \Pr\{Y_i = j\}}{\Pr\{Y_i \geq 1\}} = 2sN \cdot \frac{p}{1 - (1-p)^k} = sN \tag{13}$$

由公式(13)得知, $p=p(k)$ 应满足方程:

$$(1-p)^k + 2p - 1 = 0 \tag{14}$$

对于 M 个独立随机变元的和 $\sum_{i=1}^M Y_i$, 由引理 1 知, 存在常数 $\delta_k(s) > 0$, 使得:

$$\Pr(\mathcal{D} | \mathcal{C}) = \Pr\left(\sum_{i=1}^M Y_i = sN | \mathcal{C}\right) \sim \frac{\delta_k(s)}{\sqrt{2\pi N}} \tag{15}$$

结合公式(9)、公式(10)、公式(12)、公式(14)和公式(15)有:

$$\begin{aligned} f(s) &= \lim_{N \rightarrow \infty} \frac{1}{N} \ln E[\mathcal{L}] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} (2^N \cdot \Pr(F \text{ is SAT by } \mathbf{1})) \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \left(2^N \cdot \frac{\Pr(\mathcal{C}) \cdot \Pr(\mathcal{D} | \mathcal{C})}{\Pr(\mathcal{D})} \right) \\ &\sim \ln 2 + \frac{2s}{k} \ln(1 - (1-p)^k) - s \ln(4p(1-p)) \\ &= \ln 2 + \frac{2s}{k} \ln(2p) - s \ln(4p(1-p)). \end{aligned}$$

引理 2 证毕. □

进一步, 对于每个固定的 $k \geq 3$, 由公式(14)知 $p=p(k)$ 为常数, 若能够证明函数 $f(s)$ 是关于 s 的严格单调递减函数, 则当 $f(s)=0$ 时, 便可得到严格随机正则 (k, r) -SAT 问题可满足临界值的上界 $\alpha_{rn}(k)$, 由此, 我们给出如下的定理:

定理 3. 存在某个常数 k_1 , 当 $k \geq k_1$ 时, 若严格随机正则 (k, r) -CNF 公式 F 的约束密度 $\alpha_r = \frac{2s}{k}$ 满足 $\alpha_r > \alpha_{rn}(k) = 2^k \ln 2 - \frac{k-2}{2} \ln 2 + o_k(1)$ 时, 则随机公式 F 是高概率地不可满足的.

证明: 由引理 2 得知 $f'(s) = \frac{2}{k} \ln(2p) - \ln(4p(1-p))$, 由 $p \in (0, 1)$ 及 $p=p(k)$ 满足方程 $(1-p)^k = 1-2p$ 易知 $0 < 1-2p < 1$

且 $p < \frac{1}{2}$. 令 $t = \frac{1}{1-2p}$, 则公式(14)可化为 $\left(\frac{t+1}{2t}\right)^k = \frac{1}{t}$, 即 $k = \ln(t) / \ln\left(\frac{2t}{t+1}\right)$, 下面我们证明 $f'(s) < 0$.

由 $\frac{2}{k} \ln(2p) - \ln(4p(1-p)) = \frac{2}{k} \ln\left(\frac{2t-2}{2t}\right) - \ln\left(\frac{t^2-1}{t^2}\right)$, 即, 我们需要证明 $\frac{\ln(1-t^2)}{\ln(1-t^{-1})} < \frac{2}{k}$ 成立. 因此, 我们只需证明

公式(16)和公式(17)成立即可:

$$\frac{\ln(1-t^2)}{\ln(1-t^{-1})} < t^{-1} \tag{16}$$

$$k < 2t \tag{17}$$

令 $x=t^{-1}$, 则有 $0 < x < 1$. 由公式(16)得知 $\ln(1-x^2) - x \ln(1-x) > 0$, 令 $g(x) = \ln(1-x^2) - x \ln(1-x)$, 则在 $x \in (0, 1)$ 上有 $g'(x) = \frac{x}{1+x} - \ln(1-x) > 0$ 成立, 所以 $g(x)$ 是 $x \in (0, 1)$ 上的严格单调递增函数. 因此当 $x \in (0, 1)$ 时, 有 $\ln(1-x^2) - x \ln(1-x) > g(0) = 0$ 成立, 即公式(16)成立.

进一步证明公式(17)也成立. 由 $t = \frac{1}{1-2p}$ 得知: 为了证明 $k = \frac{\ln(t)}{\ln\left(\frac{2t}{t+1}\right)} < 2t$, 即 $\frac{\ln(1-2p)}{\ln(1-p)} < \frac{2}{1-2p}$ 成立, 我

们只要证明 $(1-2p)\ln(1-2p) - 2\ln(1-p) > 0$ 即可.

令 $h(p)=(1-2p)\ln(1-2p)-2\ln(1-p)$, 则对任意的 $k \geq 3$ 及 $p < \frac{1}{2}$ 有 $(1-p)^k \leq (1-p)^3$, 由 $(1-p)^k + 2p - 1 = 0$ 得知 $1-2p \leq (1-p)^3$, 所以有 $(1-p)^3 + 2p - 1 \geq 0$, 即 $p \geq \frac{3-\sqrt{5}}{2} > \frac{1}{4}$. 因此, 结合 $p < \frac{1}{2}$ 得知:

$$h'(p) = (1-2p)\ln(1-2p) - 2\ln(1-p) = -2\ln(1-2p) - \frac{2p}{1-p} > 0 \quad (18)$$

故 $h(p)$ 是 $p \in \left(\frac{1}{4}, \frac{1}{2}\right)$ 上的严格单调递增函数, 因此有 $h(p) > h\left(\frac{1}{4}\right) = \frac{3\ln 2}{2} > 0$, 即 $(1-2p)\ln(1-2p) - 2\ln(1-p) > 0$ 成立, 所以有 $f'(s) < 0$. 因此, $f(s)$ 是关于 s 的严格单调递减函数, 易知其有唯一的零点.

因为当 k 大于某个常数 k_1 时, 有 $p = p(k) \sim 2^{-1} - 2^{-k}$, 则由 $f(s) = \ln 2 + \frac{2s}{k} \ln(2p) - s \ln(4q(1-p)) = 0$ 有:

$$\alpha_{rs}(k) = \frac{2s}{k} = \frac{\ln 2}{\frac{k}{2} \ln(1-2^{-2k}) - \ln(1-2^{-k})} \quad (19)$$

由 $\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} + \dots + (-1)^{n-1} \frac{x^n}{n} + o(x^n)$, 对公式(19)进行简化得 $\alpha_{rs}(k) \sim 2^k \ln 2 - \frac{k-2}{2} \ln 2 + o_k(1)$. 因此, 当 k 较大且随机公式 F 满足 $\alpha_r(k) = \frac{2s}{k} > \alpha_{rs}(k)$ 时, 由于 $\lim_{N \rightarrow \infty} \frac{1}{N} \ln E[\Omega] < 0$ 成立, 所以高概率地随机公式 F 是不可满足的. 因此, 我们得到了 $\alpha_{rs}(k)$ 的一个上界, 即 $\alpha_{rs}(k) \leq \alpha_{rs}(k) = 2^k \ln 2 - \frac{k-2}{2} \ln 2 + o_k(1)$ 成立.

定理 3 证毕. □

3.2 随机正则(k,r)-SAT 问题可满足的临界

随机正则(k,r)-SAT 问题仅在 r 取奇数时有别于严格随机正则(k,r)-SAT 问题, 而当 r 取奇数时, 随机正则(k,r)-CNF 公式中每个变元的正、负出现次数恰好相差 1 次. 由于 $\alpha_{rs}(k)$ 的确切值不一定恰好发生在 r 取奇数的情形, 另外, 即便这种情形发生, 此时的约束密度 $\alpha_{rs}(k) = \frac{r}{k}$ 的绝对误差仅为 $\left| \frac{2\lfloor r/2 \rfloor + 1}{k} - \frac{2\lfloor r/2 \rfloor}{k} \right| = \frac{1}{k}$, 由于 k 较大时, $\frac{1}{k}$ 的值相对于整个临界值点 $\alpha_{rs}(k)$ 的值来说影响甚微, 因此可以认为严格正则和正则情形的可满足临界值是近似相等的. 此外, 文献[21]中分别对 r 取偶数和奇数情形的推导表明: 当 k 较大时, r 取偶数或取值为一个与之相差 1 的奇数时, 其对应的可满足临界值是近似相等的. 由此, 结合文献[20]中二阶矩方法给出的下界 $\alpha_{r,i} = 2^k \ln 2 - \frac{k+1}{2} \ln 2 - 1$, 我们得到了随机正则(k,r)-SAT 问题可满足临界值 $\alpha_{rs}(k)$ 的一个更紧致界, 即:

$$2^k \ln 2 - \frac{k+1}{2} \ln 2 - 1 \leq \alpha_{rs}(k) \leq 2^k \ln 2 - \frac{k-2}{2} \ln 2 \quad (20)$$

且上下界之间仅相差一个常数 $1 + \frac{3\ln 2}{2}$.

4 数值结果

当 k 较小时, 我们可以直接通过计算得到引理 2 中 $p(k)$ 的数值解, 并由 $f(s) = 0$ 来直接计算其相应的可满足临界值点的数值上界 $\alpha_{rru}(k)$. 表 1 给出了 $3 \leq k \leq 11$ 时, $p(k)$ 的数值解及近似解 $2^{-1} - 2^{-k}$ 、数值上界 $\alpha_{rru}(k)$ 、渐近上界 $\alpha_{ru}(k) = 2^k \ln 2 - (k-2) \ln 2 / 2$ 以及渐近上界 $\alpha_{ru}(k)$ 与数值上界 $\alpha_{rru}(k)$ 的差 $\alpha_{ru}(k) - \alpha_{rru}(k)$. 从表 1 所给出的计算结果可知: 数值解上界 $\alpha_{rru}(k)$ 与渐近上界 $\alpha_{ru}(k) = 2^k \ln 2 - (k-2) \ln 2 / 2$ 的值较为接近, 且随着 k 不断增大(例如当 $k \geq 9$ 时), 由于 $2^{-1} - 2^{-k}$ 能够较好地近似 $q(k)$ 的数值结果, 因此, $2^k \ln 2 - (k-2) \ln 2 / 2$ 也能很好地近似 $\alpha_{ru}(k)$ 的值.

当 $k=3$ 时, 文献[23]证明了: 如果 r 可以不取整数时, 随机正则(k,r)-SAT 问题的临界值 $\alpha_{rs}(3) \sim 3.6$. 文献[24]预言: r 取整数时, 该问题的临界值 $\alpha_{rs}(3) \sim 3.6667$. 而我们所给的数值上界结果 $\alpha_{rru}(3) = 3.7822$ 已经非常接近 $\alpha_{rs}(3)$ 的

值.此外,文献[23]的研究表明:在接近相变点处的随机正则 (k,r) -SAT 问题实例,其解空间的熵密度比接近相应相变点处同规模的均匀随机 k -SAT 问题实例的熵密度要小.由于均匀随机 k -SAT 问题在 k 较大时的可满足临界值点 $\alpha_s(k)$ 为 $2^k \ln 2 - (\ln 2 + 1)/2 + o_k(1)$,因此我们给出的随机正则 (k,r) -SAT 问题的可满足临界值的上界 $\alpha_{ru}(k)$ 明显在 $\alpha_s(k)$ 的左侧,这也进一步从理论上解释了在相变点处的随机正则 (k,r) -SAT 问题实例为什么比在相应相变点处同规模的随机 k -SAT 问题实例更难满足.

Table 1 Numerical calculation results

表 1 数值计算结果

k	$p(k)$	$2^{-1} - 2^{-1-k}$	α_{nru}	α_{ru}	$\alpha_{ru} - \alpha_{nru}$
3	0.381 97	0.437 50	3.782 2	4.505 5	0.723 3
4	0.456 31	0.468 75	9.107 6	9.704 1	0.596 5
5	0.481 21	0.484 38	19.934 5	20.447 8	0.513 3
6	0.491 34	0.492 19	41.825 5	42.282 0	0.456 5
7	0.495 86	0.496 09	85.879 1	86.296 8	0.417 7
8	0.497 98	0.498 05	174.281 4	174.673 0	0.391 6
9	0.499 01	0.499 01	351.447 4	351.772 2	0.324 8
10	0.499 51	0.499 51	705.984 1	706.317 0	0.332 9
11	0.499 75	0.499 75	1 415.415 0	1 415.753 1	0.338 1

5 结束语

描述解空间 S 大小 $|\Omega|$ 的一个重要统计物理学量是其熵密度 $\ln|\Omega|$,因此,我们通过构造特殊的独立随机实验,结合一阶矩方法,采用 $\ln E[|\Omega|]$ 来近似 $\ln|\Omega|$ 并给出了随机正则 (k,r) -SAT 问题可满足临界值的一个新的上界.此外,由统计物理中的1RSB理论可知^[27,28]:当一个随机正则公式 F 的约束密度 α_r 满足 $\alpha_r < \alpha_{rs}(k)$ 时, F 的解空间被分解成了若干分割清晰的聚类,每个聚类中包含着指数级可满足指派中的一小部分解,此时有 $\ln E[|\Omega|] \sim \ln|\Omega|$ 成立.相比之下,当 $\alpha_r > \alpha_{rs}(k)$ 时,1RSB 预言在接近相变点附近,有限数量的聚类开始起主导作用,即,有限多个聚类覆盖了几乎整个解空间,所以解空间 S 中,解的分布的不均匀性将变得较为显著,在这种情形下得到的 $\ln E[|\Omega|]$ 的值将大于 $\ln|\Omega|$.因此在这个区域内, $\ln E[|\Omega|]$ 不能很好地近似 $\ln|\Omega|$,所以本文通过一阶矩方法得到的上界值应该比 $\alpha_{rs}(k)$ 的实际值要稍微偏大一些.因此,进一步的工作是如何利用统计物理学的相关知识,找到 $\alpha_{rs}(k)$ 的确切值.

References:

- [1] Cook SA. The complexity of theorem-proving procedures. In: Proc. of the 3rd Annual ACM Symp. on Theory of Computing. ACM Press, 1971. 151–158. [doi: 10.1145/800157.805047]
- [2] Johnson DS. The NP-completeness column: An ongoing guide. Journal of Algorithms, 1981,2(4):393–405. [doi: 10.1016/0196-6774(81)90037-7]
- [3] Cook SA, Mitchell DG. Finding hard instances of the satisfiability problem: A survey. In: Proc. of the Satisfiability Problem: Theory and Applications: DIMACS Workshop. American Mathematical Soc., 1997. 1–17.
- [4] Friedgut E, Bourgain J. Sharp thresholds of graph properties, and the k -sat problem. Journal of the American Mathematical Society, 1999,12(4):1017–1054. [doi: 10.1090/S0894-0347-99-00305-7]
- [5] Krzakala F, Montanari A, Ricci-Tersenghi F, Semerjian G, Zdeborová L. Gibbs states and the set of solutions of random constraint satisfaction problems. Proc. of the National Academy of Sciences, 2007,104(25):10318–10323. [doi: 10.1073/pnas.0703685104]
- [6] Semerjian G. On the freezing of variables in random constraint satisfaction problems. Journal of Statistical Physics, 2008,130(2):251–293. [doi: 10.1007/s10955-007-9417-7]
- [7] Achlioptas D, Ricci-Tersenghi F. Random formulas have frozen variables. SIAM Journal on Computing, 2009,39(1):260–280. [doi: 10.1137/070680382]
- [8] Achlioptas D, Coja-Oghlan A, Ricci-Tersenghi F. On the solution-space geometry of random constraint satisfaction problems. Random Structures & Algorithms, 2011,38(3):251–268. [doi: 10.1002/rsa.20323]
- [9] Mézard M, Zecchina R. Random k -satisfiability problem: From an analytic solution to an efficient algorithm. Physical Review E, 2002,66(5):056126. [doi: 10.1103/PhysRevE.66.056126]
- [10] Mézard M, Parisi G, Zecchina R. Analytic and algorithmic solution of random satisfiability problems. Science, 2002,297(5582):812–815. [doi: 10.1126/science.1073287]
- [11] Goerdt A. A threshold for unsatisfiability. In: Proc. of the Int'l Symp. on Mathematical Foundations of Computer Science. Berlin, Heidelberg: Springer-Verlag, 1992. 264–274. [doi: 10.1007/3-540-55808-X_25]

- [12] Boufkhad Y, Dubois O, Interian Y, Selman B. Regular random k -SAT: Properties of balanced formulas. *Journal of Automated Reasoning*, 2005,1(35):181–200. [doi: 10.1007/s10817-005-9012-z]
- [13] Coja-Oghlan A, Panagiotou K. Catching the k -NAESAT threshold. In: *Proc. of the 44th Annual ACM Symp. on Theory of Computing*. ACM Press, 2012. 899–908. [doi: 10.1145/2213977.2214058]
- [14] Ricci-Tersenghi F, Weigt M, Zecchina R. Simplest random k -satisfiability problem. *Physical Review E*, 2001,63(2):026702-1. [doi: 10.1103/PhysRevE.63.026702]
- [15] Ding J, Sly A, Sun N. Satisfiability threshold for random regular NAE-SAT. *Communications in Mathematical Physics*, 2016, 341(2):435–489. [doi: 10.1007/s00220-015-2492-8]
- [16] Mertens S, Mézard M, Zecchina R. Threshold values of random k -SAT from the cavity method. *Random Structures & Algorithms*, 2006,28(3):340–373. [doi: 10.1002/rsa.20090]
- [17] Coja-Oghlan A, Panagiotou K. The asymptotic k -SAT threshold. *Advances in Mathematics*, 2016,288:985–1068. [doi: 10.1016/j.aim.2015.11.007]
- [18] Ding J, Sly A, Sun N. Proof of the satisfiability conjecture for large k . In: *Proc. of the 47th Annual ACM on Symp. on Theory of Computing*. ACM Press, 2015. 59–68. [doi: 10.1145/2746539.2746619]
- [19] Hoory S, Szeider S. Computing unsatisfiable k -SAT instances with few occurrences per variable. *Theoretical Computer Science*, 2005,337(1):347–359. [doi: 10.1016/j.tcs.2005.02.004]
- [20] Rathi V, Aurell E, Rasmussen LK, Skoglund M. Bounds on threshold of regular random k -SAT. *Lecture Notes in Computer Science*, 2010,6175:264–277. [doi: 10.1007/978-3-642-14186-7_22]
- [21] Castellana M, Zdeborová L. Adversarial satisfiability problem. *Journal of Statistical Mechanics: Theory and Experiment*, 2011, 2011(3):P03023. [doi: 10.1088/1742-5468/2011/03/P03023]
- [22] Krishnamurthy S, Sahoo S. Balanced K -satisfiability and biased random k -satisfiability on trees. *Physical Review E*, 2013,87(4): 042130. [doi: 10.1103/PhysRevE.87.042130]
- [23] Krishnamurthy S. Exact satisfiability threshold for k -satisfiability problems on a Bethe lattice. *Physical Review E*, 2015,92(4): 042144. [doi: 10.1103/PhysRevE.92.042144]
- [24] Zhou JC, Xu DY, Lu YJ, Dai CK. Strictly regular random $(3,s)$ -SAT model and its phase transition phenomenon. *Journal of Beijing University of Aeronautics and Astronautics (in Chinese with English abstract)*. <http://www.cnki.net/kcms/detail/11.2625.V.20160413.1537.007.html>
- [25] Kschischang FR, Frey BJ, Loeliger HA. Factor graphs and the sum-product algorithm. *IEEE Trans. on Information Theory*, 2001, 47(2):498–519. [doi: 10.1109/18.910572]
- [26] Flajolet P, Sedgewick R. *Analytic Combinatorics*. Cambridge University Press, 2009.
- [27] Coja-Oghlan A, Zdeborová L. The condensation transition in random hypergraph 2-coloring. In: *Proc. of the 23rd Annual ACM-SIAM Symp. on Discrete Algorithms*. SIAM, 2012. 241–250. [doi: 10.1137/1.9781611973099.22]
- [28] Ding J, Sly A, Sun N. Maximum independent sets on random regular graphs. *arXiv preprint arXiv: 1310.4787*, 2013.

附中文参考文献:

- [24] 周锦程,许道云,卢友军,代寸宽.严格随机正则 $(3,s)$ -SAT 模型及其相变现象.北京航空航天大学学报. <http://www.cnki.net/kcms/detail/11.2625.V.20160413.1537.007.html>



周锦程(1981—),男,贵州开阳人,博士生,副教授,CCF 会员,主要研究领域为计算复杂性,可满足性问题,算法设计与分析.



卢友军(1985—),男,博士生,CCF 学生会会员,主要研究领域为计算复杂性,复杂网络.



许道云(1959—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为可计算分析,计算复杂性,可满足性问题.