软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2021,32(9):2629−2641 [doi: 10.13328/j.cnki.jos.006049]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

取定 *s* 的严格 *d*-正则随机(3,2*s*)-SAT 问题的可满足临界*

王永平[1,2], 许道云[1]

[1](贵州大学 计算机科学与技术学院,贵州 贵阳 550025)

[2](贵州财经大学 数统学院,贵州 贵阳 550025)

通讯作者: 许道云, E-mail: dyxu@gzu.edu.cn

摘 要: 3-CNF 公式的随机难解实例生成对于揭示 3-SAT 问题的难解实质和设计满足性测试的有效算法有着重要意义.对于整数 $k>2$ 和 $s>0$,如果在一个 k-CNF 公式中每个变量正负出现次数均为 s,则称该公式是严格正则(k,2s)-CNF 公式.受严格正则(k,2s)-CNF 公式的结构特征启发,提出每个变量正负出现次数之差的绝对值均为 d 的严格 d-正则(k,2s)-CNF 公式,并使用新提出的 SDRRK2S 模型生成严格 d-正则随机(k,2s)-CNF 公式.取定整数 $5<s<11$,模拟实验显示,严格 d-正则随机(3,2s)-SAT 问题存在 SAT-UNSAT 相变现象和 HARD-EASY 相变现象.因此,立足于 3-CNF 公式的随机难解实例生成,研究了严格 d-正则随机(3,2s)-SAT 问题在 s 取定时的可满足临界.通过构造一个特殊随机实验和使用一阶矩方法,得到了严格 d-正则随机(3,2s)-SAT 问题在 s 取定时可满足临界值的一个下界.模拟实验结果验证了理论证明所得下界的正确性.

关键词: 3-CNF 公式;随机难解实例生成;正则子类;严格 d-正则随机(3,2s)-SAT 问题;可满足临界

中图法分类号: TP301

中文引用格式: 王永平,许道云.取定 s 的严格 d-正则随机(3,2s)-SAT 问题的可满足临界.软件学报,2021,32(9):2629−2641. http://www.jos.org.cn/1000-9825/6049.htm

英文引用格式: Wang YP, Xu DY. Satisfiability threshold of strictly d-regular random (3,2s)-SAT problem for fixed s. Ruan Jian Xue Bao/Journal of Software, 2021,32(9):2629−2641 (in Chinese). http://www.jos.org.cn/1000-9825/6049.htm

Satisfiability Threshold of Strictly *d*-regular Random (3,2*s*)-SAT Problem for Fixed *s*

WANG Yong-Ping[1,2], XU Dao-Yun[1]

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

[2](School of Mathematics and Statistics, Guizhou University of Finance and Economics, Guiyang 550025, China)

Abstract: Generating random hard instances of the 3-CNF formula is an important factor in revealing the intractability of the 3-SAT problem and designing effective algorithms for satisfiability testing. Let $k>2$ and $s>0$ be integers, a k-CNF formula is a strictly regular (k,2s)-CNF one if the positive and negative occurrence number of every variable in the formula are s. On the basis of the strictly regular (k,2s)-CNF formula, the strictly d-regular (k,2s)-CNF formula is proposed in which the absolute value of the difference between positive and negative occurrence number of every variable is d. A novel model is constructed to generate the strictly d-regular random (k,2s)-CNF formula. The simulated experiments show that the strictly d-regular random (3,2s)-SAT problem has an SAT-UNSAT phase transition and a HARD-EASY phase transition when the parameter $5<s<11$ is fixed, and that the latter is related to the former. Hence, the satisfiability threshold of the strictly d-regular random (3,2s)-SAT problem is studied when the parameter s is fixed. A lower bound of the satisfiability threshold is obtained by constructing a random experiment and using the first moment method. The subsequent simulated experiments verify well the lower bound proved.

Key words: 3-CNF formula; generating random hard instances; subclass with regular structure; strictly d-regular random (3,2s)-SAT problem; satisfiability threshold

* 基金项目: 国家自然科学基金(61762019, 61862051)
Foundation item: National Natural Science Foundation of China (61762019, 61862051)
收稿时间: 2019-03-22; 修改时间: 2019-11-28; 采用时间: 2020-04-03; jos 在线出版时间: 2020-05-26

设 N 是正整数,$x_1,x_2,\ldots,x_N$ 是布尔变量.对于任取的 $x\in\{x_1,x_2,\ldots,x_N\}$,称$\neg x$ 和 x 为文字,其中,$\neg x$ 是 x 的否定.有限个文字的析取构成一个子句,而有限个子句的合取构成一个合取范式(conjunctive normal form,简称 CNF)公式.设 F 是一个 CNF 公式,如果存在一个真值指派$\sigma\in\{0,1\}^N$ 使得 F 取值为真,则称 F 是可满足的,并称σ是 F 的一个解.如果 F 没有解,则称 F 是不可满足的.例如,$F=(x_1\vee\neg x_2)\wedge(x_3\vee\neg x_4\vee\neg x_5)$是一个包含两个子句的 CNF 公式,而 $\sigma=(0,0,1,1,1)$(这意味$(x_1,x_2,x_3,x_4,x_5)=(0,0,1,1,1)$)是 F 的一个解(这里,$\sigma(\neg x)=\neg\sigma(x)$).

习惯地,将判定一个 CNF 公式是否可满足称为求解该公式.设整数 $k>2$,如果一个 CNF 公式的每个子句恰好包含 k 个文字,则称该公式是 k-CNF 公式.求解任意一个 k-CNF 公式的问题称为 k-SAT 问题.1973 年,Levin 证明了 3-SAT 问题是 NP-完全的[1].这说明:如果 P≠NP,则 3-SAT 问题不存在多项式时间算法.因此,3-CNF 公式的随机难解实例生成,对于揭示 3-SAT 问题的难解实质和设计满足性测试的有效算法有着重要意义.

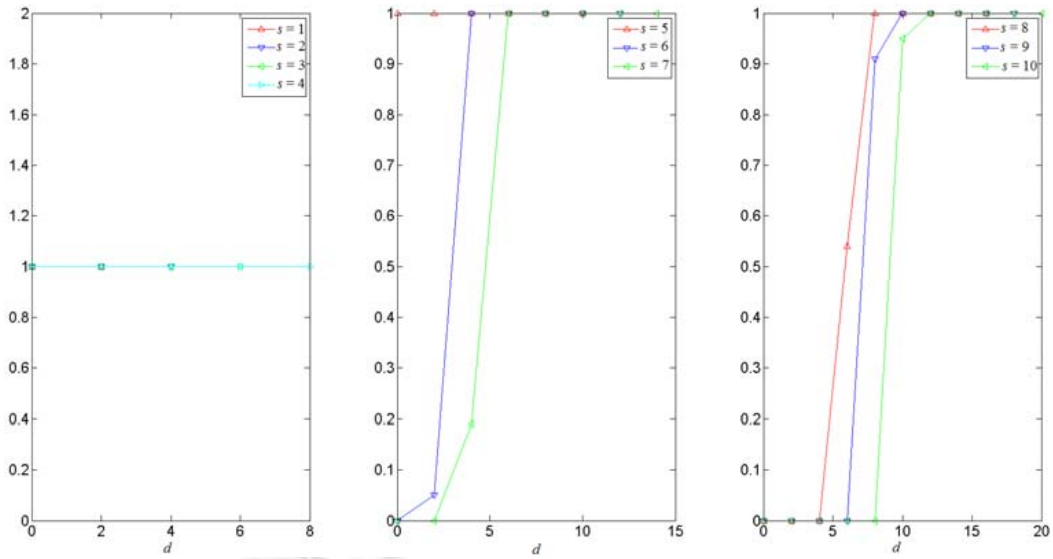
均匀 k-SAT 模型用于生成具有 M 个子句和 N 个变量的随机 k-CNF 公式,其中,M 和 N 是正整数.该模型先从 $2^k\binom{N}{k}$ 个可能的子句中均匀且相互独立地选取 M 个子句,而后将这些子句合取成一个 k-CNF 公式.设 F 是一个随机 3-CNF 公式,其子句数 M 与变量数 N 的比值为α.1999 年,Friedgut 等人证明了随机 3-SAT 问题存在 SAT-UNSAT 相变现象[2]:存在一个变量数 N 的函数α_3,使得当$\alpha<\alpha_3$时,随机 3-CNF 公式是高概率可满足的;而当$\alpha>\alpha_3$时,随机 3-CNF 公式是高概率不可满足的.虽然到目前为止还不知道α_3的具体值,但文献[3,4]得到了 $3.52\leqslant\alpha_3\leqslant 4.4898$,而文献[5]基于一个单调性假设得到了$\alpha_3\leqslant 4.262$.此外,模拟实验估计$\alpha_3$ 的值取得了较好的结果[6−9]:$\alpha_3\approx 4.267$,并且随机 3-CNF 公式的难解实例集中在$\alpha=4.267$ 附近.综上,比值α不仅与随机 3-CNF 公式是否可满足有关,还与求解该类公式的难度有关.

设 F 是一个 CNF 公式,x 是 F 的一个变量.称文字 $x(\neg x)$在 F 中的出现次数是 x 在 F 中的正(负)出现次数,并称 x 在 F 中的正出现次数与负出现次数之和是 x 在 F 中的出现次数.设整数 $s>0$,如果在一个 k-CNF 公式中每个变量均出现 s 次,则称该公式是正则(k,s)-CNF 公式.文献[10]利用多项式归约技术证明了正则(3,4)-SAT 问题是 NP-完全的.这说明,存在难解的正则(3,4)-CNF 公式实例.文献[11]利用 Zchaff 求解器[12](该求解器是目前求解 3-CNF 公式最有效的完备求解器)进行的模拟实验表明:当每个变量正负出现次数之差的绝对值至多是 1 时,正则$(3,s)$-CNF 公式的随机难解实例集中在 $s=11$ 附近,并且 $s=11$ 时求解该类实例的难度远大于比值$\alpha=4.267$ 时求解随机 3-CNF 公式实例的难度.注意到,正则$(3,s)$-CNF 公式的子句数与变量数的比值恒为 $s/3$,因此,文献[10,11]的上述结论说明:当 $s=4$ 或 11,即比值$\alpha=4/3$ 或 11/3 时,存在 3-CNF 公式的难解实例,只不过实例是正则的.受此启发,我们考虑了取定 s 时正则$(3,s)$-CNF 公式的随机难解实例生成问题.注意到:对于正则$(3,s)$-CNF 公式来说,s 取定意味着比值$\alpha=s/3$.因此,当 s 取定时,需要找到类似于比值α的新参数,使得该新参数不仅影响正则$(3,s)$-CNF 公式是否可满足,还影响求解该类公式的难度.文献[13−16]分别从 SAT-UNSAT 相变、解的个数等方面研究了严格正则(k,s)-CNF 公式,其中,s 是偶数,每个变量的正负出现次数均为 $s/2$.显然,在严格正则(k,s)-CNF 公式中,每个变量正负出现次数之差均为 0.结合文献[11,13−16],我们提出了每个变量正负出现次数之差的绝对值均为 d 的严格 d-正则(k,s)-CNF 公式.该类公式是正则(k,s)-CNF 公式的一个子类,而 d 是取定的非负整数.注意到:当 s 是偶数时,参数 d 的可能取值为 0,2,…,s;当 s 是奇数时,参数 d 的可能取值为 1,3,…,s.为方便讨论,在本文中,我们选择 s 是偶数的情形,即严格 d-正则$(k,2r)$-CNF 公式的情形,其中,$r>0$ 是整数.为保持符号一致,我们将严格 d-正则$(k,2r)$-CNF 公式仍表述为严格 d-正则$(k,2s)$-CNF 公式.

类似于使用均匀 k-SAT 模型生成随机 k-CNF 公式,借鉴已有模型[11,13,17,18],我们提出一个新的模型来生成严格 d-正则随机$(k,2s)$-CNF 公式.在这个模型生成公式时,先由 N 个变量得到 $2Ns$ 个文字,其中,每个变量均出现 $2s$ 次,并且均以相等的概率正出现 $s+\frac{d}{2}$ 或 $s-\frac{d}{2}$ 次;再从这 $2Ns$ 个文字的置换全体中均匀地选择一个置换,并根据该置换生成一个公式.为方便起见,称该模型是 SDRRK2S 模型.在下一节,我们将详细叙述这一模型.

本文立足于 3-CNF 公式的随机难解实例生成.为此,我们通过模拟实验观察了 s 取定时,参数 d 与严格 d-正则随机(3,2s)-CNF 公式是否可满足及其求解难度的关系.我们分别选取了 $s=1,2,\ldots,10$,并且对于每一个 s 值分别

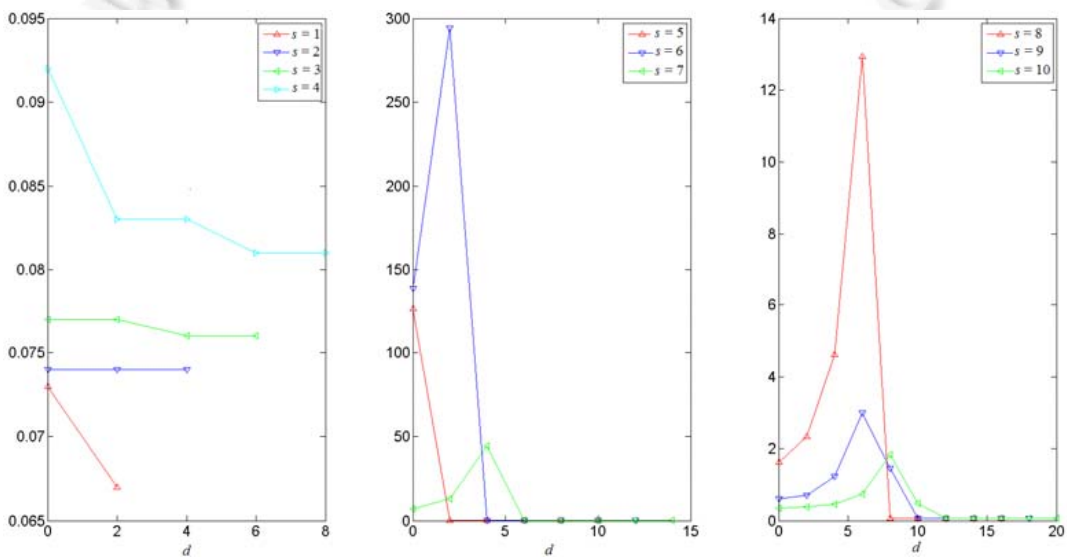
选取了 N=180 和 210.对于取定的 s,N,k=3 以及 $d\in\{0,2,\ldots,2s\}$,先使用 SDRRK2S 模型生成 100 个实例,再使用 Zchaff 求解器[12]逐一求解,最后统计可满足实例占比以及求解一个实例的平均时间.图 1、图 2 给出了 N=180 时的模拟实验结果.注意到,N=210 时也有类似的结果.为节约篇幅,我们没有展示 N=210 时的实验结果.



(a) s=1,2,3,4(d=0,2,…,2s) (b) s=5,6,7(d=0,2,…,2s) (c) s=8,9,10(d=0,2,…,2s)

Fig.1 Ratio of satisfiable instances when N=180 and $s\in\{1,2,\ldots,10\}$

图 1 当 N=180 而 s=1,2,…,10 时的可满足实例占比



(a) s=1,2,3,4(d=0,2,…,2s) (b) s=5,6,7(d=0,2,…,2s) (c) s=8,9,10(d=0,2,…,2s)

Fig.2 Average time in seconds for solving an instance when N=180 and $s\in\{1,2,\ldots,10\}$

图 2 当 N=180 而 s=1,2,…,10 时,求解一个实例的平均时间(s)

取定整数 5<s<11,由图 1 和图 2 可知,严格 d-正则随机(3,2s)-SAT 问题存在 SAT-UNSAT 相变现象和 HARD-EASY 相变现象.另外,由图 1(b)、图 2(b)、图 1(c)和图 2(c)可知:对于取定的整数 5<s<11,严格 d-正则随机(3,2s)-

SAT 问题的 HARD-EASY 相变现象与 SAT-UNSAT 相变现象有着一定联系.例如:当 s=6 时,SAT-UNSAT 相变点位于 d=2 与 d=4 之间(如图 1(b)所示),而 HARD-EASY 相变点位于 d=2 处(见图 2(b)).再如:当 s=10 时,SAT-UNSAT 相变点位于 d=8 与 d=10 之间(如图 1(c)所示),而 HARD-EASY 相变点位于 d=8 处(见图 2(c)).因此,在理论上研究严格 d-正则随机(3,2s)-SAT 问题在 s 取定时的 SAT-UNSAT 相变现象有助于 3-CNF 公式的随机难解实例生成.在本文中,我们借鉴已有做法[13],通过构造一个特殊随机实验和使用一阶矩方法给出了严格 d-正则随机(3,2s)- SAT 问题在 s 取定时可满足临界值的一个下界.模拟实验结果验证了理论证明所得下界的正确性.

1 SDRRK2S 模型

2005 年,Boufkhad 等人[17]提出了一种不同于均匀 k-SAT 模型的新模型来生成具有 N 个变量和αN 个子句的随机 k-CNF 公式.显然,该模型所生成公式的子句数与变量数的比值恰好是α.在该模型生成公式时,每个变量对应的两个文字均以概率 $p=\left\lfloor\frac{k\alpha}{2}\right\rfloor+1-\frac{k\alpha}{2}$ 出现 $\left\lfloor\frac{k\alpha}{2}\right\rfloor$ 次,以概率 $1-p$ 出现 $\left\lfloor\frac{k\alpha}{2}\right\rfloor+1$ 次.这说明,每个变量出现次数的期望值均为 $k\alpha$.因此,文献[17]所讨论的公式已经非常接近正则的 k-CNF 公式.记文字多重集$\{x_1,\ldots,x_1,\neg x_1,\ldots,\neg x_1,x_2,\ldots,x_2,\neg x_2,\ldots,\neg x_2,\ldots,x_N,\ldots,x_N,\neg x_N,\ldots,\neg x_N\}$为 L,其中,对于每一个 $i\in[N]=\{1,2,\ldots,N\}$,x_i 与$\neg x_i$ 的个数都是 $\left\lfloor\frac{k\alpha}{2}\right\rfloor$ 或 $\left\lfloor\frac{k\alpha}{2}\right\rfloor+1$.文献[17]的模型先均匀地将集合 L 划分成 $\frac{|L|}{k}$ 个部分,使得每个部分均包含 k 个文字(在本文中,符号|·|总表示某个集合的元素个数);然后,通过将 $\frac{|L|}{k}$ 个部分各自构成子句来得到一个随机公式.文献[17]的模拟实验(使用了文献[3]提出的贪心算法)表明:所生成的公式存在 SAT-UNSAT 相变现象,而且当 k=3 时,难解实例集中在α=3.5 附近.这说明,文献[17]的模型有助于 3-CNF 公式的随机难解实例生成.此外,文献[17]还分别使用一阶矩方法和文献[3]提出的贪心算法证明了:当 k=3 并且α>3.7822 时,所生成的公式是高概率不可满足的;而当 k=3 并且α<2.46 时,所生成的公式是高概率可满足的.随后,Rathi 等人[18]通过在文献[17]的模型上增加一个限制条件来生成随机 k-CNF 公式.该条件要求每个变量对应的两个文字具有相同的出现次数.在文献[18]中,Rathi 等人分别使用一阶矩方法和二阶矩方法证明了:当 k=3 并且α>3.7822 时,所生成的公式是高概率不可满足的;而当 k=3 并且α<2.667 时,所生成的公式是高概率可满足的.

如果在一个子句中,每个变量至多出现一次,则称该子句是合法子句;否则是非法子句.一个格局公式[17]可能包含非法子句.文献[17,18]的模型生成的是格局公式.当一个格局公式不包含非法子句时,称之为简单公式.周锦程等人[11]提出了一个模型来生成不包含重复子句的简单正则随机(3,s)-CNF 公式.在文献[11]的模型生成公式时:若 s 是偶数,则每个变量对应的两个文字均出现 s/2 次;若 s 是奇数,则每个变量均出现 s 次,但以相等的概率正出现 $\frac{s+1}{2}$ 或 $\frac{s-1}{2}$ 次.表面上,文献[11]的模型由文字多重集(见文献[17]的相应叙述)得到随机公式的方式与文献[17]的情形有很大的不同,但二者本质上是相同的,只不过文献[11]的模型除去了不是简单公式或包含重复子句的实例.文献[11]的模拟实验(使用 Zchaff 求解器[12])表明:所生成的公式存在 SAT-UNSAT 相变现象,而且难解实例集中在 s=11 附近(即比值α=3.667 附近).这说明文献[11]的模型虽然不同于文献[17]的模型,但仍然有助于 3-CNF 公式的随机难解实例生成.文献[11]还使用一阶矩方法证明了:当 s>11 时,所生成的公式是高概率不可满足的.此外,由于所生成的公式是简单公式而且不包含重复子句,文献[11]的模型不太容易生成公式实例.因此,在文献[13]中,周锦程等人放弃了简单公式以及不包含重复子句的限制,修改了文献[17]的模型以生成正则随机(k,2s)-CNF 公式.在文献[13]的模型生成公式时,每个变量对应的两个文字均出现 s 次,从而每个变量均出现 2s 次.本质上,文献[13]的模型是通过均匀地选择文字多重集(见文献[17]的相应叙述)的一个置换来得到一个随机公式的,从而得到的也是格局公式.在文献[13]中,周锦程等人通过构造一个特殊随机实验,并结合一阶矩方法证明了:当 k=3 并且比值α>3.7822 时,所生成的公式是高概率不可满足的.此外,当 k 较大时,周锦程等人[15]利用一阶复本对称破缺理论并结合一阶矩方法改进了文献[13]的结果.

SDRRK2S 模型用来生成严格 d-正则随机(k,2s)-CNF 公式.注意到:在一个严格 d-正则(k,2s)-CNF 公式中,

每个变量均出现 2s 次,并且正负出现次数之差的绝对值均为 d.这意味着,每个变量均只能正出现 $s+\frac{d}{2}$ 或 $s-\frac{d}{2}$ 次.因此,在 SDRRK2S 模型中,借鉴了文献[11]的模型当 s 是奇数时的情形来确定每个变量的正出现次数:每个变量均出现 2s 次,但以相等的概率正出现 $s+\frac{d}{2}$ 或 $s-\frac{d}{2}$ 次.另一方面,当 k=3 并且比值 α>3.7822 时,文献[13,17,18]的模型生成的公式均是高概率不可满足的;但文献[13]的模型由文字多重集得到随机公式的方式更便于计数.因此,在 SDRRK2S 模型中,采用了由文字多重集的一个置换得到一个随机公式的方式.

设整数 k>2,s>0,N>0 以及 $d\in\{0,2,\ldots,2s\}$,则 SDRRK2S 模型可叙述如下.

Input:子句长度 k,变量出现次数 2s,变量个数 N,变量正负出现次数之差的绝对值 d,其中,2Ns 是 k 的倍数.

Step 1. 对于每一个 $i\in[N]$,以随机方式生成多重集 $A_i=\{x_i,\ldots,x_i,\neg x_i,\ldots,\neg x_i\}$,其中,$A_i$ 中有 2s 个元素,而且 x_i 的个数以相等的概率取 $s+\frac{d}{2}$ 或 $s-\frac{d}{2}$.

Step 2. 记多重集 $A=\bigcup_{i=1}^{N}A_i$,并均匀地从 A 的置换全体中选择一个置换.不妨设该置换为

$$a_1,a_2,\ldots,a_k,a_{k+1},a_{k+2},\ldots,a_{2k},\ldots,a_{2Ns-k+1},a_{2Ns-k+2},\ldots,a_{2Ns}.$$

Step 3. 根据 Step 2 中的置换生成公式 F 如下:

$$F=(a_1\vee a_2\vee\ldots\vee a_k)\wedge(a_{k+1}\vee a_{k+2}\vee\ldots\vee a_{2k})\wedge\ldots\wedge(a_{2Ns-k+1}\vee a_{2Ns-k+2}\vee\ldots\vee a_{2Ns}).$$

Output:输出公式 F.

注意,SDRRK2S 模型生成的公式是含有 N 个布尔变量的格局公式.因此,该类公式可能包含非法子句.实际上,由文献[17]可知:当 k=3 时,SDRRK2S 模型以正概率输出简单公式.此外,由文献[17]还可知:当 k=3 并且 SDRRK2S 模型生成的格局公式存在 SAT-UNSAT 相变现象时,相应的简单公式也存在相同的相变.再加上本文关注的是严格 d-正则随机(3,2s)-SAT 问题在 s 取定时的可满足临界,因此我们没有在 SDRRK2S 模型中加入简单公式的限制.

2 可满足临界分析

本节主要使用文献[13]的方法研究严格 d-正则随机(3,2s)-SAT 问题在 s 取定时的可满足临界.这一过程主要分成两步:一是通过构造特殊随机实验得到一个特殊真值指派是一个严格 d-正则随机(k,2s)-CNF 公式解的概率的渐近表达式;二是使用一阶矩方法得到严格 d-正则随机(3,2s)-SAT 问题在 s 取定时可满足临界值的下界.

2.1 相关准备

设 F 是一个严格 d-正则随机(k,2s)-CNF 公式.如果 F 的文字全体组成的多重集是 L,则由 SDRRK2S 模型可知,L 可作为该模型 Step 2 中的多重集 A.由 SDRRK2S 模型还可知,该模型可由 L 生成(2Ns)!个公式.因此,该(2Ns)!个公式中可满足的公式占比即为 F 是可满足的概率.如果 F 的文字全体组成的多重集是 L',则由 SDRRK2S 模型可知,此时 F 是可满足的概率与 F 的文字全体组成的多重集是 L 时相应的概率相等.因此在本节中,约定 F 的文字全体组成的多重集是某个取定的集合.

设 F 是一个严格 d-正则随机(k,2s)-CNF 公式.注意到,F 是否可满足取决于是否存在作为 F 解的真值指派.因此,可考虑使用真值指派和公式共同描述该公式可满足的可能性.设 l 是 F 的一个文字,$\sigma\in\{0,1\}^N$ 是一个真值指派.如果 $\sigma(l)$=1,则称 l 是公式 F 的由 σ 决定的 1-文字;否则是公式 F 的由 σ 决定的 0-文字.在不致混淆的情况下,简称 l 是 1-文字或 0-文字.设 $A_{(F\&\sigma)}$ 是 1-文字全体组成的多重集,而 $B_{(F\&\sigma)}$ 是 0-文字全体组成的多重集.进一步,设 $S_{(F\&\sigma)}$ 是 SDRRK2S 模型由多重集 $A_{(F\&\sigma)}\cup B_{(F\&\sigma)}$ 生成的公式全体组成的多重集.于是,真值指派 σ 是公式 F 解的概率如下:

$$\frac{|\{H\,|\,H\in S_{(F\&\sigma)}\text{并且}H\text{的每个子句均包含1-文字}\}|}{|S_{(F\&\sigma)}|}.$$

为方便起见,记为 $P_{(\sigma\to F)}$.

设 F 是一个严格 d-正则随机(k,2s)-CNF 公式.对于每一个 $x\in\{x_1,x_2,\ldots,x_N\}$,若 x 在 F 中的正出现次数不小于负出现次数,令 $\sigma_F(x)=1$;否则,令 $\sigma_F(x)=0$.显然,σ_F 是为 F 定义的特殊真值指派.作为下一小节使用一阶矩方法的准备,我们给出了概率值 $P_{(\sigma_F\to F)}$ 的两个性质.

(1) 对于任意的真值指派 $\sigma\in\{0,1\}^N$,都有 $P_{(\sigma\to F)}\leqslant P_{(\sigma_F\to F)}$;

(2) 通过构造一个特殊随机实验,可得到 $P_{(\sigma_F\to F)}$ 的一个渐近表达式.

首先,给出得到 $P_{(\sigma_F\to F)}$ 最大性性质的一个引理.

引理 1. 设 F 是一个严格 d-正则随机(k,2s)-CNF 公式,$\sigma\in\{0,1\}^N$ 是一个真值指派.如果存在真值指派 $\tau\in\{0,1\}^N$,使得 $|A_{(F\&\sigma)}|\leqslant|A_{(F\&\tau)}|$,则 $P_{(\sigma\to F)}\leqslant P_{(\tau\to F)}$.

证明:不妨设 F 含 M 个子句,则 $|A_{(F\&\sigma)}\cup B_{(F\&\sigma)}|=kM=|A_{(F\&\tau)}\cup B_{(F\&\tau)}|$.于是可设:

$$A_{(F\&\sigma)}=\{t_1,t_2,\ldots,t_a\},B_{(F\&\sigma)}=\{t_{a+1},t_{a+2},\ldots,t_{kM}\},A_{(F\&\tau)}=\{T_1,T_2,\ldots,T_{a+b}\},B_{(F\&\tau)}=\{T_{a+b+1},T_{a+b+2},\ldots,T_{kM}\},$$

其中,a 和 b 是使得 $a+b\leqslant kM$ 的非负整数.设 H 是 $S_{(F\&\sigma)}$中任意一个公式.如果对于任意的 $i\in[M]$和 $j\in[k]$,l_{ij} 表示 H第 i 个子句中的第 j 个文字,则 $l_{11},l_{12},\ldots,l_{1k},l_{21},l_{22},\ldots,l_{2k},\ldots,l_{M1},l_{M2},\ldots,l_{Mk}$ 是多重集 $A_{(F\&\sigma)}\cup B_{(F\&\sigma)}$的一个置换.于是,由对 $A_{(F\&\sigma)}$和 $B_{(F\&\sigma)}$的假设可知,该置换还可表示如下:

$$t_{1'},t_{2'},\ldots,t_{k'},t_{[(2-1)k+1]'},t_{[(2-1)k+2]'},\ldots,t_{(2k)'},\ldots,t_{[(M-1)k+1]'},t_{[(M-1)k+2]'},\ldots,t_{(Mk)'},$$

其中,下标全体组成的集合恰好是[kM].如果对于任意的 $i\in[M]$和 $j\in[k]$,令 $L_{ij}=T_{[(i-1)k+j]'}$,则如下的格局公式:

$$(L_{11}\vee L_{12}\vee\ldots\vee L_{1k})\wedge(L_{21}\vee L_{22}\vee\ldots\vee L_{2k})\wedge\ldots\wedge(L_{M1}\vee L_{M2}\vee\ldots\vee L_{Mk})$$

必定属于 $S_{(F\&\tau)}$.进一步,如果令该格局公式为 $f(H)$,则容易证明,f 是 $S_{(F\&\sigma)}$和 $S_{(F\&\tau)}$之间的双射函数.设多重集 $S_1=\{G|G\in S_{(F\&\sigma)}$并且 G 的每个子句均包含 1-文字$\}$,$S_2=\{H|H\in S_{(F\&\tau)}$并且 H 的每个子句均包含 1-文字$\}$,显然,对于任意的 $G_0\in S_1$ 都有:G_0 的每个子句均包含 $A_{(F\&\sigma)}$中的文字.因此,由双射函数 f 的定义可知,$f(G_0)$的每个子句均包含 $A_{(F\&\tau)}$中的文字,即 $f(G_0)\in S_2$.进一步,由双射函数 f 是从 S_1 到 S_2 的单射可知,$|S_1|\leqslant|S_2|$.最后,由 $|S_{(F\&\sigma)}|=(kM)!=|S_{(F\&\tau)}|$ 可知,$P_{(\sigma\to F)}\leqslant P_{(\tau\to F)}$.引理 1 证毕. □

设 F 是一个严格 d-正则随机(k,2s)-CNF 公式,σ_F 是为 F 定义的特殊真值指派.由 σ_F 的定义可知:对于任意的真值指派 $\sigma\in\{0,1\}^N$,都有 $|A_{(F\&\sigma)}|\leqslant|A_{(F\&\sigma_F)}|$.于是,由引理 1 可得如下定理.

定理 1. 设 F 是一个严格 d-正则随机(k,2s)-CNF 公式,σ_F 是为 F 定义的特殊真值指派,则对于任意的真值指派 $\sigma\in\{0,1\}^N$,都有 $P_{(\sigma\to F)}\leqslant P_{(\sigma_F\to F)}$.

定理 1 给出了 $P_{(\sigma_F\to F)}$ 的最大性性质.因此,本小节余下的主要内容将是得到 $P_{(\sigma_F\to F)}$ 的一个渐近表达式.注意到,F 包含 2Ns 个文字和 $\frac{2Ns}{k}$ 个子句.又注意到,F 的由 σ_F 决定的 1-文字有 $\left(s+\frac{d}{2}\right)N$ 个,0-文字有 $\left(s-\frac{d}{2}\right)N$ 个.因此,可先构造一个特殊随机实验来得到 $P_{(\sigma_F\to F)}$ 的一个定性表达式.假设有编号分别为 1,2,…, $\frac{2Ns}{k}$ 的 $\frac{2Ns}{k}$ 个盒子,其中,每个盒子均包含编号分别为 1,2,…,k 的 k 个格子.如下进行随机实验:首先,按编号的自然顺序将盒子排序;其次,以概率 0<q<1 进行 2Ns 重伯努利实验,并记实验结果分别为

$$X_1,X_2,\ldots,X_k,X_{k+1},X_{k+2},\ldots,X_{2k},\ldots,X_{2Ns-k+1},X_{2Ns-k+2},\ldots,X_{2Ns};$$

最后,对于任意的 $i\in\left[\frac{2Ns}{k}\right]$ 和 $j\in[k]$,当 $X_{(i-1)k+j}=1$ 时,将第 i 个盒子编号为 j 的格子贴上标签 label-1,否则贴上标签 label-0.记在实验中每个盒子均包含被贴上标签 label-1 的格子为事件 A,并记标签 label-1 被贴 $\left(s+\frac{d}{2}\right)N$ 次为事件 B.由 0<q<1 可知,事件 B 发生的概率 $P(B)>0$.注意到:对于任意的 $i\in\left[\frac{2Ns}{k}\right]$ 和 $j\in[k]$,将第 i 个盒子编号为 j 的格子贴上标签 label-1 可看作在公式的第 i 个子句的第 j 个位置放置 1-文字.因此,由 0<q<1 可知:

$$P_{(\sigma_F\to F)}=\frac{\dfrac{|\{H\mid H\in S_{(F\&\sigma_F)}\text{并且 }H\text{ 的每个子句均包含 1-文字}\}|}{\left[\left(s+\frac{d}{2}\right)N\right]!\left[\left(s-\frac{d}{2}\right)N\right]!}q^{\left(s+\frac{d}{2}\right)N}(1-q)^{\left(s-\frac{d}{2}\right)N}}{\dfrac{|S_{(F\&\sigma_F)}|}{\left[\left(s+\frac{d}{2}\right)N\right]!\left[\left(s-\frac{d}{2}\right)N\right]!}q^{\left(s+\frac{d}{2}\right)N}(1-q)^{\left(s-\frac{d}{2}\right)N}}=\frac{P(A\cap B)}{P(B)}=\frac{P(B\mid A)P(A)}{P(B)}\quad(1)$$

由事件 A 和 B 的含义可知:

$$P(A)=[1-(1-q)^k]^{\frac{2Ns}{k}},P(B)=\binom{2Ns}{\left(s+\frac{d}{2}\right)N}q^{\left(s+\frac{d}{2}\right)N}(1-q)^{\left(s-\frac{d}{2}\right)N}\quad(2)$$

注意到,可以由 Stirling 近似得到 $P(B)$的一个渐近表达式.因此,由公式(1)可知,得到 $P(B|A)$的一个渐近表达式成为得到 $P_{(\sigma_F\to F)}$ 渐近表达式的关键.幸运的是,可以使用所谓的Local Limit Law[19]得到$P(B|A)$的一个渐近表达式.为了使用 Local Limit Law,我们从文献[19]中摘录了一个定理作为下面的引理.

引理 2(large power and Gaussian forms)[19]. 设 $A(z)=\sum_{j=0}^{+\infty}a_jz^j$ 和 $B(z)=\sum_{j=0}^{+\infty}b_jz^j$ 是满足以下条件的两个解析函数:(1) $A(z)$和 $B(z)$在 0 处解析并且系数非负;(2) $\gcd\{j|b_j>0\}=1$;(3) $B(0)\neq0$;(4) $A(z)$和 $B(z)$的收敛半径满足 $1<R_b\leqslant+\infty,R_b\leqslant R_a$.定义两个常数: $\mu=\dfrac{B'(1)}{B(1)},\sigma^2=\dfrac{B''(1)}{B(1)}+\dfrac{B'(1)}{B(1)}-\left[\dfrac{B'(1)}{B(1)}\right]^2>0$.设 $N=\mu n+x\sqrt{n}$,其中,x 属于实数域上某个有限区间.如果$[z^N](A(z)B(z)^n)$表示 $A(z)B(z)^n$ 的 N 次项系数,则:

$$\frac{1}{A(1)B(1)^n}[z^N](A(z)B(z)^n)=\frac{1}{\sigma\sqrt{2\pi n}}\mathrm{e}^{-\frac{x^2}{2\sigma^2}}\left(1+O\left(n^{-\frac{1}{2}}\right)\right),$$

其中, $O\left(n^{-\frac{1}{2}}\right)$ 表示存在常数 $C>0$,使得当 n 足够大时,有 $\left|O\left(n^{-\frac{1}{2}}\right)\right|\leqslant Cn^{-\frac{1}{2}}$.

设 $X_0=X_1+X_2+\ldots+X_k\geqslant1$,其中,$X_1,X_2,\ldots,X_k$ 是上述构造的随机实验中的前 k 重伯努利实验结果.对 X_0 独立重复观测 $\dfrac{2Ns}{k}$ 次.如果记各次实验结果之和为 X,则:

$$P\left[X=\left(s+\frac{d}{2}\right)N\right]=P(B\mid A)\quad(3)$$

其中,A 和 B 是公式(1)上方提及的两个事件.接下来,使用引理 2 得到 $P\left[X=\left(s+\dfrac{d}{2}\right)N\right]$ 的一个渐近表达式.为了能够使用引理 2,令 $Y_0=X_0-1$,此时,如果对 Y_0 独立重复观测 $\dfrac{2Ns}{k}$ 次并记各次实验结果之和为 Y,则:

$$P\left[Y=\left(s+\frac{d}{2}\right)N-\frac{2Ns}{k}\right]=P\left[X=\left(s+\frac{d}{2}\right)N\right]\quad(4)$$

因此,接下来的工作就是利用引理 2 得到 $P\left[Y=\left(s+\dfrac{d}{2}\right)N-\dfrac{2Ns}{k}\right]$ 的一个渐近表达式.

注意到,Y_0 的概率生成函数为 $f(x)=\dfrac{1}{1-(1-q)^k}\sum_{i=1}^{k}\binom{k}{i}q^i(1-q)^{k-i}x^{i-1}$.因此,可设 $A(z)\equiv1,B(z)\equiv f(z)$.容易看出,$A(z)$和 $B(z)$满足引理 2 条件.经过计算可得 $\mu=\dfrac{kq}{1-(1-q)^k}-1$.由 $f(x)$的系数及文献[20]可知:

$$\sigma^2>0\quad(5)$$

其中,σ是 k 和 q 的表达式.于是,由引理 2 可知,如果$\left(\frac{kq}{1-(1-q)^k}-1\right)\frac{2Ns}{k}$是正整数,则:

$$P\left[Y=\left(\frac{kq}{1-(1-q)^k}-1\right)\frac{2Ns}{k}\right]=P\left[Y=\left(\frac{kq}{1-(1-q)^k}-1\right)\frac{2Ns}{k}+0\sqrt{\frac{2Ns}{k}}\right]$$
$$=\frac{\left[z^{\left(\frac{kq}{1-(1-q)^k}-1\right)\frac{2Ns}{k}+0\sqrt{\frac{2Ns}{k}}}\right]\left(A(z)B(z)^{\frac{2Ns}{k}}\right)}{A(1)B(1)^{\frac{2Ns}{k}}} \tag{6}$$
$$=\frac{1}{2\sigma\sqrt{\frac{\pi Ns}{k}}}\left[1+O\left(\left(\frac{2Ns}{k}\right)^{-\frac{1}{2}}\right)\right]$$

进一步,如果$\left(s+\frac{d}{2}\right)N-\frac{2Ns}{k}=\left(\frac{kq}{1-(1-q)^k}-1\right)\frac{2Ns}{k}$,则由公式(3)、公式(4)以及公式(6),可得 $P(B|A)$的一个渐近表达式如下:

$$P(B\mid A)=\frac{1}{2\sigma\sqrt{\frac{\pi Ns}{k}}}\left[1+O\left(\left(\frac{2Ns}{k}\right)^{-\frac{1}{2}}\right)\right] \tag{7}$$

综上,由公式(1)、公式(2)以及公式(7)可得如下定理.

定理 2. 设 F 是一个严格 d-正则随机(k,2s)-CNF 公式,σ_F 是为 F 定义的特殊真值指派,而 q 是上述构造的随机实验中的概率值.如果$\frac{q}{1-(1-q)^k}=\frac{2s+d}{4s}$,则:

$$P_{(\sigma_F\to F)}=\frac{[1-(1-q)^k]^{\frac{2Ns}{k}}\left[1+O\left(\left(\frac{2Ns}{k}\right)^{-\frac{1}{2}}\right)\right]}{2\sigma\sqrt{\frac{\pi Ns}{k}}\binom{2Ns}{\left(s+\frac{d}{2}\right)N}q^{\left(s+\frac{d}{2}\right)N}(1-q)^{\left(s-\frac{d}{2}\right)N}},$$

其中,N 是 F 的变量个数;$O\left(\left(\frac{2Ns}{k}\right)^{-\frac{1}{2}}\right)$表示存在常数 C>0,使得当$\frac{2Ns}{k}$足够大时,有$\left|O\left(\left(\frac{2Ns}{k}\right)^{-\frac{1}{2}}\right)\right|\leqslant C\left(\frac{2Ns}{k}\right)^{-\frac{1}{2}}$;而$\sigma$如公式(5)所示.

进一步,由 Stirling 近似可得定理 2 的如下推论.

推论 1. 设 F 是一个严格 d-正则随机(k,2s)-CNF 公式,σ_F 是为 F 定义的特殊真值指派.如果 2s>d 并且方程$\frac{q}{1-(1-q)^k}=\frac{2s+d}{4s}$在(0,1)内存在根 q,则:

$$P_{(\sigma_F\to F)}=\frac{[1-(1-q)^k]^{\frac{2Ns}{k}}\left[1+O\left(\left(\frac{2Ns}{k}\right)^{-\frac{1}{2}}\right)\right]\sqrt{k}\left(1+\frac{d}{2s}\right)^{\frac{N(2s+d)+1}{2}}\left(1-\frac{d}{2s}\right)^{\frac{N(2s-d)+1}{2}}}{2^{2Ns+1}\sigma\, q^{\left(s+\frac{d}{2}\right)N}(1-q)^{\left(s-\frac{d}{2}\right)N}[1+o(1)]},$$

其中, $\lim\limits_{N\to+\infty}o(1)=0.$

在下一小节,我们将研究严格 d-正则随机(3,2s)-SAT 问题的可满足临界.这意味着 k=3.因此,作为本小节的

结束,经过简单计算,给出 k=3 和 2s>d 时方程 $\frac{q}{1-(1-q)^k}=\frac{2s+d}{4s}$ 在(0,1)内的根 q 如下.

引理 3. 设 F 是一个严格 d-正则随机(3,2s)-CNF 公式.如果 2s>d,则方程 $\frac{q}{1-(1-q)^3}=\frac{2s+d}{4s}$ 在(0,1)内存在唯一根: $q=\frac{3(2s+d)-\sqrt{(2s+d)(10s-3d)}}{2(2s+d)}$.

设 F 是一个严格 d-正则随机(k,2s)-CNF 公式,σ_F 是为 F 定义的特殊真值指派.显然,定理 1 和推论 1 分别给出了 $P_{(\sigma_F\to F)}$ 的最大性性质和一个渐近表达式.这两个结论是下一小节使用一阶矩方法得到严格 d-正则随机(3,2s)-SAT 问题在 s 取定时可满足临界值下界的基础.

2.2 可满足临界

设 F 是一个严格 d-正则随机(k,2s)-CNF 公式,Ω是 F 的解的个数.由 Markov 不等式可知,P(F 是可满足的)=$P(\Omega\geqslant 1)\leqslant E[\Omega]$.于是,当 $E[\Omega]$<<1 时,公式 F 是高概率不可满足的.注意到:如果 $\lim\limits_{N\to+\infty}\ln E[\Omega]=-\infty$,则当 N 足够大时,$E[\Omega]$<<1.因此,如果 $\lim\limits_{N\to+\infty}\ln E[\Omega]=-\infty$,则当 N 足够大时,公式 F 是高概率不可满足的.基于这一结论,我们给出了下面 3 个引理,并由此得到了严格 d-正则随机(3,2s)-SAT 问题在 s 取定时可满足临界值的一个下界.

注意到:如果 $\lim\limits_{N\to+\infty}\frac{\ln E[\Omega]}{N}<0$,则 $\lim\limits_{N\to+\infty}\ln E[\Omega]=-\infty$.因此,先给出下面的关于 $\lim\limits_{N\to+\infty}\frac{\ln E[\Omega]}{N}$ 的引理.

引理 4. 设 F 是一个严格 d-正则随机(3,2s)-CNF 公式,Ω是 F 解的个数.如果 2s>d,则 $\lim\limits_{N\to+\infty}\frac{\ln E[\Omega]}{N}\leqslant g(d)$,其中,

$$g(x)=\left(1-\frac{4s}{3}\right)\ln 2+\left(\frac{5s}{3}+\frac{x}{2}\right)\ln(2s+x)-\frac{4s}{3}\ln s+\left(s-\frac{x}{2}\right)\ln(2s-x)-\left(\frac{s}{3}+\frac{x}{2}\right)\ln[3(2s+x)-\sqrt{(2s+x)(10s-3x)}]-$$
$$\left(s-\frac{x}{2}\right)\ln[-(2s+x)+\sqrt{(2s+x)(10s-3x)}],x\in[0,2s-2].$$

证明:由定理 1 可知, $E[\Omega]\leqslant 2^N P_{(\sigma_F\to F)}$.于是,由 2$s$>$d$、引理 3 和推论 1 可知:

$$\lim_{N\to+\infty}\frac{\ln E[\Omega]}{N}\leqslant\lim_{N\to+\infty}\left(\ln 2+\frac{\ln P_{(\sigma_F\to F)}}{N}\right)=(1-2s)\ln 2+\frac{2s}{3}\ln[1-(1-q)^3]+\left(s+\frac{d}{2}\right)\ln\left(1+\frac{d}{2s}\right)+$$
$$\left(s-\frac{d}{2}\right)\ln\left(1-\frac{d}{2s}\right)-\left(s+\frac{d}{2}\right)\ln q-\left(s-\frac{d}{2}\right)\ln(1-q).$$

进一步,由引理 3 可知,该引理成立. □

设 $g(x)$和 d 如引理 4 所示.如果 $g(d)$<0,则由引理 4 可知, $\lim\limits_{N\to+\infty}\frac{\ln E[\Omega]}{N}<0$.因此,需要研究什么条件使得 $g(x)$<0.注意到,$x\in$[0,2s−2].为此,我们研究了 $g(x)$在 0 处和 2s−2 处的取值情况.

引理 5. 设 $g(x)$是引理 4 中定义的函数,则:(1) 当 $s\geqslant 6$ 时,$g(0)$<0;(2) $g(2s-2)$>0.

证明:先来证明结论(1).经过计算可得, $g(0)=\ln 2-\frac{s}{3}\ln(3-\sqrt{5})-s\ln(-1+\sqrt{5})$.为此,考虑函数:

$$h(x)=\ln 2-\frac{x}{3}\ln(3-\sqrt{5})-x\ln(-1+\sqrt{5}),x\in[1,+\infty).$$

注意到:

$$h'(x)=-\frac{1}{3}\ln[(3-\sqrt{5})(-1+\sqrt{5})^3]=-\frac{1}{3}\ln(-88+40\sqrt{5})<-\frac{1}{3}\ln(-88+40\times 2.23)=-\frac{1}{3}\ln 1.2<0,$$
$$h(6)=-\ln\frac{(3-\sqrt{5})^2(-1+\sqrt{5})^6}{2}=-\ln[32(-11+5\sqrt{5})^2]<-\ln[5.6^2(-11+5\times 2.236)^2]=-2\ln 1.008<0.$$

因此,由 $s\geqslant 6$ 可知,结论(1)成立.再来证明结论(2).经过计算可得:

$$g(2s-2)=\ln 2+\left(\frac{8s}{3}-1\right)\ln(2s-1)-\frac{4s}{3}\ln s-\left(\frac{4s}{3}-1\right)\ln[3(2s-1)-\sqrt{(2s-1)(2s+3)}]-\ln[\sqrt{(2s-1)(2s+3)}-(2s-1)].$$

为此,考虑定义在[1,+∞)上的函数:

$$H(x)=\ln 2+\left(\frac{8x}{3}-1\right)\ln(2x-1)-\frac{4x}{3}\ln x-\left(\frac{4x}{3}-1\right)\ln[3(2x-1)-\sqrt{(2x-1)(2x+3)}]-\ln[\sqrt{(2x-1)(2x+3)}-(2x-1)].$$

经过计算和整理可得, $H'(x)=\frac{4}{3}\ln\frac{(2x-1)^2}{3x(2x-1)-x\sqrt{(2x-1)(2x+3)}}$.

注意到, $(2x-1)^2-3x(2x-1)+x\sqrt{(2x-1)(2x+3)}=\sqrt{2x-1}(x\sqrt{2x+3}-(x+1)\sqrt{2x-1})>0$.

又注意到, $H(1)=\frac{1}{3}\ln\frac{8}{(3-\sqrt{5})(-1+\sqrt{5})^3}=\frac{1}{3}\ln\frac{11+5\sqrt{5}}{4}>0$.

于是,由 $s\geqslant 1$ 可知,$g(2s-2)=H(s)>0$,即结论(2)得证. □

设 $g(x)$是引理 4 中定义的函数.由引理 5 可知,研究什么条件使得 $g(x)<0$,还需考虑 $g(x)$在定义域内的单调性.为此,有下述引理.

引理 6. 引理 4 中定义的函数 $g(x)$在定义域内单调递增.

证明:经过计算可得:

$$g'(x)=\frac{1}{2}\ln\frac{(2s+x)(-(2s+x)+\sqrt{(2s+x)(10s-3x)})}{(2s-x)(3(2s+x)-\sqrt{(2s+x)(10s-3x)})}=\frac{1}{2}\ln\frac{2(2s+x)}{3x-2s+\sqrt{(2s+x)(10s-3x)}}.$$

注意到:

$$\ln\frac{2(2s+x)}{3x-2s+\sqrt{(2s+x)(10s-3x)}}=\ln\left[1+\frac{6s-x-\sqrt{(2s+x)(10s-3x)}}{3x-2s+\sqrt{(2s+x)(10s-3x)}}\right]>\frac{6s-x-\sqrt{(2s+x)(10s-3x)}}{2(2s+x)}>0.$$

因此,$g'(x)>0$.于是,结论得证.引理 6 证毕. □

现在,根据引理 4~引理 6 并结合一阶矩方法,可得到严格 d-正则随机(3,2s)-SAT 问题在 s 取定时可满足临界值的一个下界如下.

定理 3. 设 F 是一个严格 d-正则随机(3,2s)-CNF 公式.如果 $s\geqslant 6$ 并且 $2s>d$,则当 $d<d_0$ 并且 N 足够大时,公式 F 是高概率不可满足的,其中,d_0 是引理 4 中定义的函数 $g(x)$在区间(0,2s−2)内的唯一零点.

注意到,F 是一个严格 d-正则随机(3,2s)-CNF 公式.因此,可以为 F 定义特殊真值指派 σ_F.由 σ_F 的定义可知:当 $2s=d$ 时,σ_F 是 F 的解,即 F 是可满足的.因此,在定理 3 中加入了条件 $2s>d$.

证明:由引理 5 和引理 6 可知:当 $s\geqslant 6$ 时,引理 4 中定义的函数 $g(x)$在区间(0,2s−2)内存在唯一零点 d_0.设 Ω 是 F 解的个数,则由 $2s>d$、引理 4、引理 6 以及 $d<d_0$ 可知, $\lim\limits_{N\to+\infty}\frac{\ln E[\Omega]}{N}\leqslant g(d)<g(d_0)=0$,从而 $\lim\limits_{N\to+\infty}\ln E[\Omega]=-\infty$.于是,由本小节开始时的讨论可知:当 $d<d_0$ 并且 N 足够大时,公式 F 是高概率不可满足的.定理 3 证毕. □

定理 3 的证明给出了唯一零点 d_0 的存在性.由函数 $g(x)$表达式的复杂性可知,不太容易给出 d_0 的表达式.作为替代,表 1 给出了当 s=10,20,…,100 时 d_0 的数值解.注意:对于任意取定的 s,求出 d_0 的数值解是容易的.

Table 1 Numerial solutions of the null point d_0 when $s\in\{10,20,\ldots,100\}$

表 1 当 s=10,20,…,100 时,唯一零点 d_0 的数值解

s	d_0	s	d_0
10	2.803 7	60	59.682 4
20	11.831 0	70	72.972 7
30	22.603 8	80	86.574 7
40	34.358 5	90	100.437 7
50	46.775 6	100	114.523 4

3 模拟实验

根据定理 3 的条件 $d<d_0$,结合表 1,我们分别选择了 s=20,30,40,50,60;同时,根据条件 N 足够大,我们分别选

择了 N=165,180,195,210.现在,以 s=30 为例,说明模拟实验如何进行.当 s=30 时,由表 1 可知,d_0 的数值解是 22.6038.为此,需分别取 d=0,2,…,22.当 s=30 时,对于任意的 $N\in${165,180,195,210}和 $d\in${0,2,…,22},我们按照如下步骤进行一次实验.

Step 1. 设 k=3,用 SDRRK2S 模型生成 100 个实例.

Step 2. 使用 Zchaff 求解器[12]分别求解这 100 个实例.记求解器成功求解的实例总数是 n,并记不可满足实例总数是 n_u.

Step 3. 计算 n_u/n 并记之为 r_s.显然,r_s 表示在成功求解实例中不可满足实例所占比例.

注意到:当 s=30 时,总共需要进行 4×12=48 次实验,而且在各次实验中,三元对(s,N,d)互不相同.为方便讨论,记此 48 次实验中的任意一个是 E.设 F 是一个严格 d-正则随机(3,2s)-CNF 公式,其中,参数 s,N 以及 d 恰好是实验 E 中三元对的各相应值,则实验 E 的 100 个实例可看作对公式 F 的 100 次模拟.进一步,实验 E 的 r_s 可看作对 F 是不可满足的概率的模拟.这说明,我们的实验是合理的.

表 2~表 4 分别给出了 s=20,30,40 时的模拟实验结果,其中,11.8310,22.6038,34.3585 分别是 s=20,30,40 时 d_0 的数值解.由表 2 可知:当 s=20 时,对于任取的 d<11.8310 和 $N\in${165,180,195,210},都有 r_s=1.注意到,r_s 可看作对公式是不可满足的概率的模拟.因此,表 2 支持定理 3.同理,表 3 和表 4 也均支持定理 3.

Table 2 Values of r_s when s=20 and d<11.8310

表 2 当 s=20 并且 d<11.8310 时的 r_s 值

d	0	2	4	6	8	10
N=165	1	1	1	1	1	1
N=180	1	1	1	1	1	1
N=195	1	1	1	1	1	1
N=210	1	1	1	1	1	1

Table 3 Values of r_s when s=30 and d<22.6038

表 3 当 s=30 并且 d<22.6038 时的 r_s 值

d	0	2	4	6	8	10	12	14	16	18	20	22
N=165	1	1	1	1	1	1	1	1	1	1	1	1
N=180	1	1	1	1	1	1	1	1	1	1	1	1
N=195	1	1	1	1	1	1	1	1	1	1	1	1
N=210	1	1	1	1	1	1	1	1	1	1	1	1

Table 4 Values of r_s when s=40 and d<34.3585

表 4 当 s=40 并且 d<34.3585 时的 r_s 值

d	0	2	4	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
N=165	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
N=180	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
N=195	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
N=210	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

此外,同表 2~表 4 中各个 r_s 均等于 1 一样,当 s=50,60 时,模拟实验结果也均为 r_s=1(为节约篇幅,我们没有展示这些结果).因此,s=50,60 时的模拟实验结果也均支持定理 3.综上,模拟实验结果验证了理论证明所得下界的正确性.

注意到:在上述模拟实验中,各个 r_s 均等于 1.这说明定理 3 所得下界仍然比较粗糙,需要进一步改进.

4 结 论

基于严格正则(k,2s)-CNF 公式,我们提出了每个变量正负出现次数之差的绝对值均为 d 的严格 d-正则(k,2s)-CNF 公式.我们使用了新提出的 SDRRK2S 模型生成严格 d-正则随机(k,2s)-CNF 公式.初步进行的模拟实验表明:当整数 5<s<11 取定时,参数 d 不仅与严格 d-正则随机(3,2s)-CNF 公式是否可满足有关,还与求解该类公式的难度有关.因此,立足于 3-CNF 公式的随机难解实例生成,我们研究了严格 d-正则随机(3,2s)-SAT 问题在 s

取定时的可满足临界.通过构造一个特殊随机实验以及使用一阶矩方法,我们得到了该可满足临界值的一个下界.随后进行的模拟实验结果,验证了理论证明所得下界的正确性.

需要说明的是:从模拟实验结果看,本文所得下界仍然比较粗糙,需要进一步改进.此外,研究讨论该可满足临界值的上界以确定 SAT-UNSAT 相变点也十分必要.这些工作将为研究参数 d 如何影响公式求解难度以及设计随机难解实例生成算法奠定基础.

References:

[1] Levin LA. Universal sequential search problems. Problemy Peredachi Informatsii, 1973,9(3):115−116.

[2] 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]

[3] Kaporis AC, Kirousis LM, Lalas EG. The probabilistic analysis of a greedy satisfiability algorithm. Random Structures & Algorithms, 2006,28(4):444−480. [doi: 10.1002/rsa.20104]

[4] Díaz J, Kirousis L, Mitsche D, Pérez-Giménez X. On the satisfiability threshold of formulas with three literals per clause. Theoretical Computer Science, 2009,410(30-32):2920−2934. [doi: 10.1016/j.tcs.2009.02.020]

[5] Lundow PH, Markström K. Revisiting the cavity-method threshold for random 3-SAT. Physical Review E, 2019,99(2):022106(5).

[6] Crawford JM, Auton LD. Experimental results on the crossover point in random 3-SAT. Artificial Intelligence, 1996,81(1-2):31−57. [doi: 10.1016/0004-3702(95)00046-1]

[7] Selman B, Kirkpatrick S. Critical behavior in the computational cost of satisfiability testing. Artificial Intelligence, 1996,81(1-2): 273−295. [doi: 10.1016/0004-3702(95)00056-9]

[8] Monasson R, Zecchina R, Kirkpatrick S, Selman B, Troyansky L. Determining computational complexity from characteristic ‘phase transitions’. Nature, 1999,400(6740):133−137. [doi: 10.1038/22055]

[9] Braunstein A, Mézard M, Zecchina R. Survey propagation: An algorithm for satisfiability. Random Structures & Algorithms, 2002, 27(2):201−226. [doi: 10.1002/rsa.20057]

[10] Xu DY, Wang XF. A regular NP-complete problem and its inapproximability. Journal of Frontiers of Computer Science and Technology, 2013,7(8):691−697 (in Chinese with English abstract). [doi: 10.3778/j.issn.1673-9418.1305025]

[11] 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, 2016,42(12):2563−2571 (in Chinese with English abstract). [doi: 10.13700/j.bh.1001-5965.2015.0845]

[12] Mahajan YS, Fu ZH, Malik S. Zchaff2004: An efficient SAT solver. In: Hoos HH, Mitchell DG, eds. Theory & Applications of Satisfiability Testing. Berlin: Springer-Verlag, 2004. 360−375. [doi: 10.1007/11527695_27]

[13] 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 with English abstract). http://www.jos.org.cn/1000-9825/5129.htm [doi: 10.13328/j.cnki.jos.005129]

[14] Sumedha, Krishnamurthy S, Sahoo S. Balanced *k*-satisfiability and biased random *k*-satisfiability on trees. Physical Review E, 2013, 87(4):042130(9). [doi: 10.1103/PhysRevE.87.042130]

[15] Zhou JC, Xu DY, Lu YJ. Satisfiability threshold of regular (*k*,*r*)-SAT problem via 1RSB theory. Journal of Huazhong University of Science and Technology (Natural Science Edition), 2017,45(12):7−13 (in Chinese with English abstract). [doi: 10.13245/j.hust.171202]

[16] Coja-Oghlan A, Wormald N. The number of satisfying assignments of random regular *k*-SAT formulas. Combinatorics, Probability and Computing, 2018,27:496−530. [doi: 10.1017/S0963548318000263]

[17] Boufkhad Y, Dubois O, Interian Y, Selman B. Regular random *k*-SAT: Properties of balanced formulas. Journal of Automated Reasoning, 2005,35:181−200. [doi: 10.1007/978-1-4020-5571-3_9]

[18] Rathi V, Aurell E, Rasmussen L, Skoglund M. Bounds on threshold of regular random *k*-SAT. In: Strichman O, Szeider S, eds. Theory & Applications of Satisfiability Testing-SAT 2010. Berlin: Springer-Verlag, 2010. 264−277. [doi: 10.1007/978-3-642-14186-7_22]

[19] Flajolet P, Sedgewick R. Analytic Combinatorics. Cambridge: Cambridge University Press, 2009. 592−593. [doi: 10.1017/CBO9780511801655]

[20] Richardson T, Urbanke R. Modern Coding Theory. Cambridge: Cambridge University Press, 2008. 508−510. [doi: 10.1017/CBO9780511791338]

附中文参考文献:

[10] 许道云,王晓峰.一个正则 NP-完全问题及其不可近似性.计算机科学与探索,2013,7(8):691−697. [doi: 10.3778/j.issn.1673-9418.1305025]

[11] 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,2016,42(12):2563−2571. [doi: 10.13700/j.bh.1001-5965.2015.0845]

[13] 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 [doi: 10.13328/j.cnki.jos.005129]

[15] 周锦程,许道云,卢友军.基于 1RSB 的正则(k,r)-SAT 问题可满足临界.华中科技大学学报(自然科学版),2017,45(12):7−13. [doi: 10.13245/j.hust.171202]

王永平(1980－),男,讲师,主要研究领域为计算复杂性,可计算性分析.

许道云(1959－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为计算复杂性,可计算性分析.