软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2020,31(4):1113−1123 [doi: 10.13328/j.cnki.jos.005896]

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

d-正则(*k*,*s*)-SAT 问题的 NP 完全性*

符祖峰[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](安顺学院 电子与信息工程学院,贵州 安顺 561000)

通讯作者: 许道云, E-mail: dyxu@gzu.edu.cn

摘 要: 研究具有正则结构的SAT问题是否是NP完全问题,具有重要的理论价值.(k,s)-CNF公式类和正则(k,s)-CNF公式类已被证明存在一个临界函数$f(k)$,使得当$s\leqslant f(k)$时,所有实例都可满足;当$s\geqslant f(k)+1$时,对应的SAT问题是NP完全问题.研究具有更强正则约束的d-正则(k,s)-SAT问题,其要求实例中每个变元的正负出现次数之差不超过给定的自然数d.通过设计一种多项式时间的归约方法,证明d-正则(k,s)-SAT问题存在一个临界函数$f(k,d)$,使得当$s\leqslant f(k,d)$时,所有实例都可满足;当$s\geqslant f(k,d)+1$时,d-正则(k,s)-SAT问题是NP完全问题.这种多项式时间的归约变换方法通过添加新的变元和新的子句,可以更改公式的子句约束密度,并约束每个变元正负出现次数的差值.这进一步说明,只用子句约束密度不足以刻画CNF公式结构的特点,对临界函数$f(k,d)$的研究有助于在更强正则约束条件下构造难解实例.

关键词: d-正则(k,s)-CNF公式;SAT问题;NP完全性

中图法分类号: TP301

中文引用格式: 符祖峰,许道云.d-正则(k,s)-SAT 问题的 NP 完全性.软件学报,2020,31(4):1113−1123. http://www.jos.org.cn/1000-9825/5896.htm

英文引用格式: Fu ZF, Xu DY. NP-completeness of d-regular (k,s)-SAT problem. Ruan Jian Xue Bao/Journal of Software, 2020, 31(4):1113−1123 (in Chinese). http://www.jos.org.cn/1000-9825/5896.htm

NP-completeness of *d*-regular (*k*,*s*)-SAT Problem

FU Zu-Feng[1,2], XU Dao-Yun[1]

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

[2](School of Electronic and Information Engineering, Anshun University, Anshun 561000, China)

Abstract: The study on the NP-completeness of regular SAT problem is a subject which has important theoretical value. It is proved that there is a critical function $f(k)$ such that all formulas in ($k,f(k)$)-CNF are satisfiable, but ($k,f(k)+1$)-SAT is NP-complete, and there is such a critical function about regular (k,s)-SAT problem too. This work studies the regular SAT problem with stronger regular constraints. The regular (k,s)-CNF is the subclass of CNF in which each clause of formula has exactly k distinct literals and each variable occurs exactly s times. The d-regular (k,s)-CNF is a regular (k,s)-CNF formula that the absolute value of the difference between positive and negative occurrence number of each variable in the formula is at most d. A polynomial reduction from a k-CNF formula is presented to a d-regular (k,s)-CNF formula and it is proved that there is a critical function $f(k,d)$ such that all formulas in d-regular ($k,f(k,d)$)-CNF are satisfiable, but d-regular ($k,f(k,d)+1$)-SAT is already NP-complete. By adding new variables and new clauses, the reduction method not only can alter the ration of clauses to variables of any one CNF formula, but also can restrict the difference between positive and negative occurrences number of each variable. This shows that only using constrained density (the ration of clauses to variables) to character structural features of a CNF formula is not enough. The study of the critical function $f(k,d)$ is helpful to construct some hard instance under stronger regular constraints.

* 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: 2019-04-24; 修改时间: 2019-11-03; 采用时间: 2019-12-03

Key words: d-regular (k,s)-CNF; SAT-problem; NP-completeness

可满足性问题(satisfiability problem,简称 SAT 问题)研究给定的一个合取范式(conjunction normal form,简称 CNF)公式是否存在一组真值指派使得公式为真.每一个子句长度恰好为 k 的 SAT 问题被称为 k-SAT 问题.已经证明,当 $k\geqslant 3$ 时 k-SAT 问题是 NP 完全问题[1],说明在最坏情况下没有多项式时间的算法来判定这样的公式是否有可满足的真值指派.借助于多项式时间的归约转换技术,已发现某些结构规范的 3-CNF 公式类的可满足性问题也是 NP 完全问题,如(3,4)-CNF 公式类和正则(3,4)-CNF 公式类[2].

随机 k-SAT 问题的实验验证和理论分析已经表明[3,4],CNF 公式的子句约束密度α(公式中子句个数与变元个数的比值)是影响公式可满足性和求解难度的一个重要参数.存在一个与 k 相关的临界值点$\alpha(k)$,当$\alpha<\alpha(k)$时,随机 k-SAT 问题实例是高概率可满足,而当$\alpha>\alpha(k)$时,随机 k-SAT 问题实例是高概率不可满足.这种从满足到不可满足的突变现象,被称为随机 k-SAT 问题的相变现象.目前已发现很多 NP 完全问题都存在这种相变现象.其中关于约束满足问题(CSP),文献[5,6]提出了 RB 模型和 RD 模型,证明这些模型存在可满足性相变并给出了确切的相变点.文献[7]通过研究稀疏超图的铰链宽度,为随机约束满足问题的相变现象提供了一些理论上的证据.文献[8]研究了近似平衡(大多数变元出现在相同数目的约束中)的随机约束满足问题的可满足性相变,结果表明,这种平衡模型比不平衡模型更难求解.

正则(3,4)-CNF 公式类的子句约束密度都是 4/3,远小于随机 3-SAT 问题的相变点$\alpha(3)\approx 4.267$[9],说明正则(3,4)-CNF 公式应该是高概率可满足并且是易于求解的.但使用归约转换方法却证明正则(3,4)-SAT 问题是 NP 完全问题[2],说明在最坏情况下没有多项式时间的算法来求解正则(3,4)-CNF 公式.这是因为,这种归约方法通过添加新变元和新子句,稀释了原公式的子句约束密度,使得构造的新公式具有较低的子句约束密度,但其可满足问题却属于 NP 完全问题.显然,只用子句约束密度α来刻画 CNF 公式的结构特点是不够的.

改变每个变元的正负出现次数之差,就有可能改变 SAT 问题求解的难易程度.当每一个变元的正负出现次数都趋于相等时,这样的(k,s)-CNF 公式被称为平衡(k,s)-CNF 公式.Boufkhad 和 Dubois[10]首次提出了每个变元出现次数期望为s的平衡(k,s)-SAT 问题,并指出随机的平衡$(3,s)$-SAT 问题的可满足相变点为$2.46\leqslant\alpha_s(3)\leqslant 3.78$,低于随机 3-SAT 问题的相变点.平衡$(k,s)$-SAT 问题因为它的实例比一般的 k-SAT 问题实例更加难以计算而被广泛关注[11−15].

为了更深入地研究带正则结构的 CNF 公式,引入了d-正则(k,s)-CNF 公式类.d-正则(k,s)-CNF 公式是指每个变元的正负出现次数之差不超过自然数 d 的正则(k,s)-CNF 公式.当 $d=1$ 时,d-正则(k,s)-CNF 公式就是平衡(k,s)-CNF 公式.文献[13]分别使用一阶矩和二阶矩的方法,给出了 1-正则(k,s)-SAT 问题相变点$\alpha(k)$的上界和下界:存在一个常量 k_0,当 $k\geqslant k_0$ 时,$2k\ln 2-(k+1)\ln 2/2-1\leqslant\alpha(k)\leqslant 2k\ln 2$.文献[14]通过构造一个特殊的独立随机实验,使用一阶矩方法,给出了$\alpha(k)$更紧的上界$\alpha(k)\leqslant 2k\ln 2-(k-2)\ln 2/2$.文献[15]基于一阶复本对称破缺理论,获得了$\alpha(k)$一个更好的上界$\alpha(k)\leqslant 2k\ln 2-(k+1)\ln 2/2$,使得上下界之间只相差一个常数 1.

Kratochvíl[16]证明,当(k,s)-CNF 公式类中存在不可满足实例时,(k,s)-SAT 问题是 NP 完全问题.同时,Kratochvíl 提出,给定 $k\geqslant 3$,(k,s)-CNF 公式类存在一个临界函数$f(k)$,使得:

1) 在 $s\leqslant f(k)$时,任意(k,s)-CNF 公式一定是可满足的公式;

2) 当 $s\geqslant f(k)+1$ 时,(k,s)-CNF 公式的可满足问题是 NP 完全问题.

意味着在临界函数 $f(k)$的位置,CNF 公式的可满足情况出现了另外一种相变.关于(k,s)-SAT 问题的临界函数$f(k)$已有大量的研究成果[16−20],除了$f(3)=3,f(4)=4$,目前关于$f(k)$最好的结果是:

$f(5)\leqslant 7, 7\leqslant f(6)\leqslant 11, 13\leqslant f(7)\leqslant 17, 24\leqslant f(8)\leqslant 29, 41\leqslant f(9)\leqslant 51.$

本文给出了一种从 k-CNF 公式到 d-正则(k,s)-CNF 公式的多项式时间归约转换方法,证明了在 $k\geqslant 3$ 时,如果 d-正则(k,s)-CNF 公式类存在不可满足实例,则 d-正则(k,s)-SAT 问题是 NP 完全问题.进而证明了 d-正则(k,s)-CNF 公式类也存在临界函数$f(k,d)$,使得:

1) 在 $s\leqslant f(k,d)$时,每一个 d-正则(k,s)-CNF 公式一定是可满足的;

2) 当 $s \geqslant f(k,d)+1$ 时,d-正则(k,s)-CNF 公式的可满足问题是 NP 完全问题.

1 基础知识

本节介绍一些基本定义和符号约定.

一个文字 L 是一个布尔变元 x 或其否定$\neg x$,x 被称为正文字,$\neg x$ 被称为负文字.若干文字的析取构成了一个子句 C,简记为 $C=\{L_1,L_2,...,L_k\}$或 $C=L_1\vee L_2\vee...\vee L_k$,其长度为 k.若干子句的合取构成了一个 CNF 公式 F,简记为 $F=[C_1,C_2,...,C_m]$或 $F=C_1\wedge C_2\wedge...\wedge C_m$,其子句数为 m.公式 F 的子句数表示为$\#cl(F)$,变元数表示为$\#var(F)$,变元集合表示为 $var(F)$.一个变元在公式中正出现的次数表示为 $pos(F,x)$,负出现的次数表示为 $neg(F,x)$.一个公式 F 的不交拷贝公式 F'是指这两个公式的结构是完全一致的,但是其变元集的交集为空.如果 CNF 公式Φ和Ψ同时可满足或者同时不可满足,则称这两个公式是 SAT-等价.

正则(k,s)-CNF 公式是指每一个子句恰好含有 k 个变元,每个变元恰好出现在 s 个子句的 CNF 公式.d-正则(k,s)-CNF 公式还要求每个变元的正负出现次数不超过自然数 d.这里规定,当 s 为偶数时,d 取偶数;当 s 为奇数时,d 取奇数.因为如果一个变元在公式中出现的次数恰好为偶数,那么它的正负出现次数的差值只能是偶数,反之亦然.

定义 1. 设公式 $F=[C_1,C_2,...,C_m]$,称 F 是极小不可满足公式(minimal unsatisfiable formula,简记为 MU 公式),当且仅当公式 F 不可满足但删除任意一个子句 C_i 后可满足.$MU(k)$公式是指子句数减去变元数等于 k 的极小不可满足公式.

给定一个不可满足公式,总是能够通过删除子句而获得一个 MU 公式.

定义 2. 设公式 F 是一个 k-CNF 公式,称 F 是 xy 凝固的 d-正则(k,s)-CNF 公式,当且仅当 F 包含变元 x 和 y,并满足以下条件.

1) 变元 x 恰好出现在 1 个子句中,变元 y 恰好出现在 $s-1$ 个子句中,并且$|pos(F,y)-neg(F,y)-1|\leqslant d$;

2) 除变元 x 和 y 之外,其余每个变元刚好出现在 s 个子句中,且正负出现次数的差不超过 d;

3) 公式 F 可满足且所有的可满足指派在 x 和 y 上的值都是 true.

例如用下面矩阵表示的公式Ψ:

$$\begin{array}{c} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ y \\ x \end{array}\begin{pmatrix} + & - & + & + & & & - & - & & & & \\ + & - & - & - & + & + & & & & & & \\ + & - & & & - & - & + & + & & & & \\ & & - & + & & & & & - & + & - & + \\ & & & & - & + & & & - & + & + & - \\ & & & & & & + & & + & - & - & + \\ & & & & & & & + & & & & \end{pmatrix}.$$

很明显,公式Ψ是可满足的,且任意的可满足指派都要求变元 x 和 y 的值取 true.因此公式Ψ是一个 xy 凝固的 0-正则(3,6)-CNF 公式.

引理 1. 如果 $s\leqslant k$,则每一个(k,s)-CNF 公式都是可满足的公式[21].

引理 2. 设公式 F 有 n 个变元,其表示矩阵为

$$\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ \vdots \\ x_{n-1} \\ x_n \end{array}\begin{pmatrix} + & & & & - \\ - & + & & & \\ & - & & & \\ & & \ddots & & \\ & & & + & \\ & & & - & + \end{pmatrix},$$

则公式 F 是可满足的,且任意的满足指派中每个变元的取值都相等.

引理 2 的结论是显然的,因为公式 F 表示了一个蕴含圈:

$$x_1 \to x_2 \to \ldots \to x_n \to x_1 \to \ldots$$

2 *k*-CNF 公式到 *d*-正则(*k*,*s*)-CNF 公式的归约方法

本节假定 xy 凝固的 d-正则(k,s)-CNF 公式是存在的,设计一种从 k-CNF 公式到 d-正则(k,s)-CNF 公式的多项式时间归约方法,将 k-SAT 问题归约为 d-正则(k,s)-SAT 问题.

令Φ是一个 xy 凝固的 d-正则(k,s)-CNF 公式,Ψ为任意的一个 k-SAT 实例.不失一般性,假设公式 Ψ 中恰好有 m 个子句,每个子句有 k 个文字,则公式 Ψ 必然有 mk 个文字 $L_1,L_2,\ldots,L_{mk}$.

k-CNF 公式到 d-正则(k,s)-CNF 公式的归约方法分为以下几步.

1) 引入新的变元集 $Z=\{z_{ij}\}(1\leqslant i\leqslant m,1\leqslant j\leqslant k)$,依次替换公式$\Psi$中的 mk 个文字,从而构造公式Ψ_1.

$$\Psi_1 = \bigwedge\nolimits_{1\leqslant i\leqslant m}\bigvee\nolimits_{1\leqslant j\leqslant k} L'_{i,j},$$

$$L'_{i,j} = \begin{cases} z_{i,j}, & \text{if } L_{i,j} = z \\ \neg z_{i,j}, & \text{if } L_{i,j} = \neg z \end{cases}, z \in \mathrm{var}(\Psi).$$

2) 引入公式Φ的 $mk(k-2)$个不交拷贝$\Phi_{i,j}$,$1\leqslant i\leqslant mk$,$1\leqslant j\leqslant k-2$.公式$\Phi$中变元 x 在公式$\Phi_{i,j}$中被重命名为 $x_{i,j}$,变元 y 在公式$\Phi_{i,j}$中被重命名为 $y_{i,j}$.每一个$\Phi_{i,j}$都是 xy 凝固的 d-正则(k,s)-CNF 公式,这些$\Phi_{i,j}$和Φ满足任何两个之间都没有共享变元.定义变元集 $X=\{x_{i,j}\}$,$Y=\{y_{i,j}\}$,$1\leqslant i\leqslant mk$,$1\leqslant j\leqslant k-2$.构造公式 $\Psi_2 = \bigwedge\nolimits_{1\leqslant i\leqslant mk}\bigwedge\nolimits_{1\leqslant j\leqslant k-2}\Phi_{i,j}$.

3) 构造公式 $\Psi_3 = \bigwedge\nolimits_{1\leqslant i\leqslant mk} d_i$, 其中, $d_i = z_i \vee \neg z_j \vee \neg x_{i,1} \vee \neg x_{i,2} \vee \ldots \vee \neg x_{i,k-2}$, z_j 是公式Ψ_1 中与 z_i 替换了公式Ψ中同一个变元的下一个新变元,且如果 z_i 是这个变元的最后一次替换,则 z_j 是这个变元的第 1 次替换.

4) 引入新的变元集 $U=\{u_1,u_2,\ldots,u_{mks}\}$.用变元集 X,Y,Z,U 随机构造一个 k-CNF 公式 Ψ_4,并满足以下条件.

i) 变元集 Z 中每个变元 $z_{i,j}$ 恰好出现在 $s-3$ 个子句中,并且若 $z_{i,j}$ 在 Ψ_1 中是正出现,则

$$0 \leqslant \left| neg(\Psi_4, z_{i,j}) - pos(\Psi_4, z_{i,j}) - 1 \right| \leqslant d.$$

反之,则 $0 \leqslant \left| pos(\Psi_4, z_{i,j}) - neg(\Psi_4, z_{i,j}) - 1 \right| \leqslant d$.

ii) 变元集 X 中每个变元 $x_{i,j}$ 恰好出现在 $s-2$ 个子句中,并且 $0 \leqslant pos(\Psi_4, x_{i,j}) - neg(\Psi_4, x_{i,j}) \leqslant d$.变元集 Y 中每个变元 $y_{i,j}$ 恰好负出现在 1 个子句中.

iii) 变元集 U 中每个变元 u 恰好出现在 s 个子句中,并且 $pos(\Psi_4,u) - neg(\Psi_4,u) = \min(d,1)$.

iv) Ψ_4 的每一个子句中至少包含变元集 U 中任意变元的一次正出现.

5) 构造公式 $\Psi'=\{\Psi_1,\Psi_2,\Psi_3,\Psi_4\}$.

显然,公式 Ψ'中每一个变元的出现次数都是 s,而且每一个变元正负出现次数差都不超过 d.公式 Ψ'就是归约的结果.

归约方法首先构造 4 个子公式$\Psi_1,\Psi_2,\Psi_3,\Psi_4$,合取后获得 d-正则(k,s)-CNF 公式 Ψ'.从构造过程易于看出Ψ_1,Ψ_2,Ψ_3是一定可以在多项式时间内构造完成的.

Ψ_4是用变元集 X,Y,Z,U 随机组成的 k-CNF 公式,但需要满足 4 个条件.前 3 个条件都易于满足,第 4 个条件需要保证变元集 U 中变元的所有正出现次数超过所有的子句数.

Ψ_4中的文字总数为$(s-3)\times mk+(s-1)\times mk\times(k-2)+mks^2$,可知公式 Ψ_4 中的子句数为

$$\begin{aligned}\#cl(\Psi_4) &= (s-3)\times m+(s-1)\times m\times(k-2)+ms^2 \\ &= sm-3m+smk-mk+2m-2sm+ms^2 \\ &= ms^2+smk-sm-m-mk.\end{aligned}$$

变元集 U 中变元的所有正出现次数为 $pos(\Psi_4,u) = \dfrac{mks^2}{2}$. 因 $k\geqslant 3$ 得 $k^2-5\geqslant 0$.可知

$$-\frac{(k-1)^2}{(k-2)^2} + 2\frac{k+1}{k-2} \geqslant 0,$$

$$\left(s-\frac{k-1}{k-2}\right)^2-\frac{(k-1)^2}{(k-2)^2}+2\frac{k+1}{k-2}\geqslant 0,$$
$$ks^2\geqslant 2s^2+2ks-2s-2k-2,$$
$$\frac{mks^2}{2}\geqslant ms^2+mks-ms-mk-m,$$
$$pos(\Psi_4,u)\geqslant \#cl(\Psi_4).$$

可见变元集 U 中变元的所有正出现次数超过 Ψ_4 的子句数.公式 Ψ_4 是一定可以在多项式时间内构造完成的.即公式 Ψ' 的构造可以在$|\Psi|$的多项式时间内完成.

假定公式 Ψ 可满足,且指派 τ 满足了公式 Ψ,指派 $\tau_{i,j}$ 满足了公式 $\Phi_{i,j}$(因为公式 Φ 是可满足的,公式 $\Phi_{i,j}$ 是公式 Φ 的不交拷贝).定义一个指派 τ' 如下:

$$\tau'(v)=\begin{cases}\tau(z), & v\in var(\Psi_1),z\in var(\Psi)(v\text{替换}z)\\ \tau_{i,j}(v), & v\in var(\Phi_{i,j})\\ \text{true}, & v\in U\end{cases}.$$

可知指派 τ' 必然满足公式 Ψ_1,Ψ_2 和 Ψ_3.因为 Ψ_4 中每个子句都含有变元集 U 中变元的一次正出现,显然,指派 τ' 必然满足公式 Ψ_4.由公式 Ψ' 的结构可知指派 τ' 必然满足公式 Ψ'.可见公式 Ψ' 是可满足的.

假定 Ψ' 公式是可满足的,且指派 τ 满足了公式 Ψ'.显然,指派 τ 必然满足公式 Ψ_1,Ψ_2,Ψ_3 和 Ψ_4.因

$$\Psi_2=\bigwedge\nolimits_{1\leqslant i\leqslant mk}\bigwedge\nolimits_{1\leqslant j\leqslant k-2}\Phi_{i,j},$$

可知指派 τ 满足每一个 $\Phi_{i,j}$.又因 $\Phi_{i,j}$ 是 xy 凝固的 d-正则(k,s)-CNF 公式,得到 $\tau(x_{i,j})=\tau(y_{i,j})=$true.

显然, $\tau(\neg x_{i,j})=\text{false},1\leqslant i\leqslant mk,1\leqslant j\leqslant k-2$. 化简公式 Ψ_3 并根据引理 2 可得,如果变元 z_i 和 z_j 替换了公式 Ψ 中的同一个变元,则有 $\tau(z_i)=\tau(z_j)$.

定义一个指派 τ' 如下:

$\tau'(v)=\tau(z)$,如果 Ψ_1 中变元 z 替换了公式 Ψ 中的变元 v.

显然,指派 τ' 满足公式 Ψ.公式 Ψ 是可满足的.

由此可见,公式 Ψ 和公式 Ψ' 是 SAT-等价.

这种归约转换方法通过添加新的变元和新的子句,将任意的 k-CNF 公式的子句约束密度变成 $\alpha=s/k$,并且让每个变元的正负出现次数差不超过自然数 d.

3 *d*-正则(*k*,*s*)-SAT 问题的 NP 完全性和临界函数

本节将使用第 2 节介绍的归约方法来讨论 d-正则(k,s)-SAT 问题的 NP 完全性.

引理 3. 给定正整数 $k\geqslant 3$,如果存在 xy 凝固的 d-正则(k,s)-CNF 公式,则 d-正则(k,s)-SAT 问题是 NP 完全问题.

证明:如果存在 xy 凝固的 d-正则(k,s)-CNF 公式,则第 2 节中设计的归约方法可以在多项式时间内将任意的 k-CNF 公式 Ψ 归约到 d-正则(k,s)-CNF 公式 Ψ'.根据 k-SAT 问题的 NP 完全性,可知如果存在 xy 凝固的 d-正则(k,s)-CNF 公式,d-正则(k,s)-SAT 问题是 NP 完全问题. □

引理 4. 给定正整数 $k\geqslant 3$ 和 s,如果存在不可满足的 d-正则(k,s)-CNF 公式,则一定存在 xy 凝固的 d-正则(k,s)-CNF 公式.

证明:如果存在不可满足的 d-正则(k,s)-CNF 公式,则通过消减子句一定可以变成一个极小不可满足的(k,s)-CNF 公式,并且每个变元的正负出现次数不会超过$(s+d)/2$.假设公式 Φ 是一个不可满足的 d-正则(k,s)-CNF 公式,$\Phi=\Phi_1\wedge\Phi_2$,其中,Φ_1 是一个极小不可满足的(k,s)-CNF 公式.C_1 是公式 Φ_1 的子句集,C_2 是公式 Φ_2 的子句集.在 $var(\Phi_1)$ 中任选一个变元 y 和包含变元 y 的负出现的子句 c.定义 $\tilde{C}=(C_1\setminus\{c\})\cup\{\tilde{c}\}$,其中,$\tilde{c}=(c\setminus\{\neg y\})\cup\{x\}$, x 是一个没有出现在公式 Φ 的新变元.可知子句集 $\tilde{C}$ 对应的公式 $\tilde{\Phi}$ 是一个可满足公式,且任何一个可满足指派中变元 x 和变元 y 的取值都是 true.

假定公式Φ_2中的子句数为 m,文字个数为 mk.引入新的变元集 $X=\{x_1,x_2,...,x_{tk}\}$,其中,$t>2m/s$.用原有的 mk 个文字和新变元集 X 中变元,随机构造 k-CNF 公式Φ_3,并让公式Φ_3满足以下条件.(1) 变元集 X 中每个变元恰好出现$\lceil s/2\rceil$次正出现和$\lfloor s/2\rfloor$次负出现.(2) 每个子句恰好包含 k 个变元,并且每个子句至少包含变元集 X 中任一变元的一次正出现.构造公式 $\Phi'=\tilde{\Phi}\wedge\Phi_3$.

因为公式Φ_3中每个子句都包含 X 中任一变元的一次正出现,X 中变元取 true 的任意指派都可满足公式Φ_3,并且Φ_3的可满足性与公式 $\tilde{\Phi}$ 中变元取值无关.由此可知,Φ'一定是可满足的,且每一个可满足指派中变元 x 和变元 y 的取值都是 true.即 Φ'是一个 xy 凝固的 d-正则(k,s)-CNF 公式.

接下来证明构造公式Φ_3的可行性.

公式Φ_3中子句数目为 $m+ts$,变元集 X 中所有变元正出现的次数为 $tks/2$.

因 $t>2m/s$,则 $m<ts/2$,得 $m+ts<3ts/2$.又因 $k\geqslant 3$,则 $m+ts<kts/2$,即变元集 X 中所有变元正出现的次数大于公式Φ_3中子句数.因此公式Φ_3是可构造的.

显然,公式 Φ'是可以在多项式时间内构造,并且它是一个 xy 凝固的 d-正则(k,s)-CNF 公式. □

定理 1. 给定正整数 $k\geqslant 3$,若存在不可满足的 d-正则(k,s)-CNF 公式,则 d-正则(k,s)-SAT 问题是 NP 完全问题.

证明:由引理 3 和引理 4 可以直接得到. □

推论 1. 对于每一个 $k\geqslant 3$,存在某个正整数 $f(k,d)$,使得:(1) 在 $s\leqslant f(k,d)$时,d-正则(k,s)-SAT 问题的每一个实例一定是可满足的;(2) 在 $s\geqslant f(k,d)+1$ 时,d-正则(k,s)-SAT 问题是 NP 完全问题.称 $f(k,d)$为 d-正则(k,s)-SAT 问题的临界函数.

证明:由定理 1 可直接得到结论. □

$f(k,d)$还可以表示为

$$f(k,d)=\max\{s:d-\text{正则}(k,s)-CNF\cap UNSAT=\varnothing\}.$$

推论 2. $3\leqslant f(3,0)\leqslant 4$,$3<f(3,1)<5$,$f(3,2)=3$.

证明:根据引理 1 可知,所有的正则(3,3)-CNF 公式一定是可满足的,意味着所有的 d-正则(3,3)-CNF 公式一定是可满足的.可见 $f(3,d)\geqslant 3$.又因正则(3,4)-CNF 公式类中存在不可满足的实例[2],易知 $f(3,2)=3$.

构造一个公式 F_1,用矩阵表示如下:

$$\begin{array}{c} x_1\\ x_2\\ x_3\\ x_4\\ x_5\\ x_6 \end{array}\begin{pmatrix}
+ & + & + & - & - & - & & & & & & \\
+ & + & - & & & & - & & - & & + & \\
+ & - & + & & & & - & & - & & & + \\
& & & + & + & - & & - & & - & & + \\
& & & + & - & + & & - & & - & + & \\
& & & & & & + & + & - & - & - & +
\end{pmatrix}.$$

显然,公式 F_1 是一个 0-正则(3,6)-CNF 公式,但公式 F_1 是不可满足的.可知 0-正则(3,6)-SAT 问题是 NP 完全问题,且 $3\leqslant f(3,0)\leqslant 4$.

构造一个公式 F_2,用矩阵表示如下:

$$\begin{array}{c} x_1\\ x_2\\ x_3\\ x_4\\ x_5\\ x_6\\ x_7\\ x_8\\ x_9\\ x_{10}\\ x_{11}\\ x_{12} \end{array}\begin{pmatrix}
+ & + & + & - & - & & & & & & & & & & & & & & & \\
+ & + & - & & & - & - & & & & & & & & & & & & & \\
+ & - & + & & & - & - & & & & & & & & & & & & & \\
& & & + & - & & & + & & + & & - & & & & & & & & \\
& & & - & + & & & & + & & - & & - & & & & & & & \\
& & & & & & & + & + & - & & & & - & - & & & & & \\
& & & & & & & + & - & + & & & & - & - & & & & & \\
& & & & & & & & & & + & + & - & & & - & - & & & \\
& & & & & & & & & & + & - & + & & & - & - & & & \\
& & & & & & & & & & & & & & & + & - & + & - & + \\
& & & & & + & - & & & & & & & & & & & - & + & + \\
& & & & & & & & & & & & & - & + & & & + & + & -
\end{pmatrix}.$$

显然,公式 F_2 是一个 1-正则(3,5)-CNF 公式,但是公式 F_2 是不可满足的.可知 $3<f(3,1)<5$,1-正则(3,5)-SAT 问

题是 NP 完全问题. □

4 临界函数 $f(k,d)$的上下界

虽然$f(k)$不知道是否可计算,但对$f(k)$的上下界已有很多研究成果,并被证明其是严格递增的[16].本节主要研究每个变元正负出现次数均衡程度对正则(k,s)-CNF 公式可满足问题的影响,即研究临界函数 $f(k,d)$的上下界随k,d 的变化情况.

因为 d-正则(k,s)-CNF 公式属于(k,s)-CNF 公式,如果(k,s)-CNF 公式类中每一个实例公式均可满足,则 d-正则(k,s)-CNF 公式类中所有实例公式可满足.即可得$f(k,d)\geqslant f(k)$.又因为$f(k)$是严格递增的,可知$f(k,d)\geqslant f(k-1)$.

因为 d-正则(k,s)-CNF 公式一定是$(d+1)$-正则(k,s)-CNF 公式,显然可得 $f(k,d)\geqslant f(k,d+1)$.因为正则(k,s)-CNF 公式中变元正负出现的差最大为 $s-2$,显然可知,如果 $f(k)<s$,则一定有 $f(k,s-2)<s$.

定理 2. 当 $d=f(k)-1$ 时,$f(k,d)=f(k)$.

证明:当 $s\leqslant f(k)$时,任何(k,s)-CNF 公式都是可满足的.这些公式中每一个变元正负出现次数差最大为 $s-2\leqslant f(k)-1$.也就是说,当 $s\leqslant f(k)$时,任何$(f(k)-1)$-正则(k,s)-CNF 公式都是可满足的.

当 $s=f(k)+1$ 时,一定存在一个不可满足的(k,s)-CNF 公式.通过在这个公式中添加新的子句,使每个变元恰好出现 s 次,从而获得一个不可满足的正则(k,s)-CNF 公式.另外,这个公式中每一个变元正负出现次数差最大为$s-2=f(k)-1$.可见这是一个不可满足的$(f(k)-1)$-正则(k,s)-CNF 公式.

因此,当 $d=f(k)-1$ 时,$f(k,d)=f(k)$. □

由此看出,当 d 逐步增大时,松弛了对变元正负出现次数差值的约束,使得 $f(k,d)$趋近于 $f(k)$.

定理 3. $f(k+1,d)>f(k,d+2)$.

证明:假定存在一个不可满足的 d-正则$(k+1,f(k+1,d)+1)$-CNF 公式Φ,包含的变元为 $x_1,x_2,\ldots,x_n$.由定理 1 可得$f(k+1,d)\geqslant k$,可知 $\#cl(\Phi)>\#var(\Phi)$. 根据 Hall 婚配定理,公式Φ中必然存在一组不重复的子句 $c_i,i=1,2,\ldots,n$,使得每个子句 c_i 包含 x_i 或 $\neg x_i$.

从公式Φ构造公式 Φ',构造方法如下.

1) 在子句 c_i 中删除 x_i 的出现,$i=1,2,\ldots,n$;

2) 剩余的 $\#cl(\Phi)-n$ 个子句中,每个子句任意移除一个文字,并利用移除的文字和添加的新变元,组成新的子句,使得公式 Φ'成为一个$(d+2)$-正则$(k,f(k+1,d))$-CNF 公式.

不可满足公式的每个子句都删除一个文字(子句仍不为空)或在公式中增加新的子句,公式仍然不可满足.因此,如果公式Φ是不可满足的,易知公式 Φ'是不可满足的$(d+2)$-正则$(k,f(k+1,d))$-CNF 公式. □

定理 4. $f(k+1,2d)\leqslant 2f(k,d)+1$.

证明:令Φ是一个不可满足的 d-正则$(k,f(k,d)+1)$-CNF 公式,子句集为 C.定义公式

$$\Psi=\bigwedge_{c\in C}(c\vee x_c)\wedge(c\vee\neg x_c),$$

对于每一个 $c\in C,x_c$ 都是一个新变元.显然公式Ψ是一个$(k+1,2(f(k,d)+1))$-CNF 公式,并且每个变元正负出现次数的差值均不超过 $2d$.由 $(c\vee x_c)\wedge(c\vee\neg x_c)$ 和 c 是 SAT-等价可知,Ψ和Φ是 SAT-等价,得出公式Ψ是不可满足的.接着补充新子句,使Ψ变成 $2d$-正则$(k+1,2(f(k,d)+1))$-CNF 公式 Ψ',显然 Ψ'也是不可满足的.因此,$f(k+1,2d)\leqslant 2f(k,d)+1$. □

当 $d=0$ 时,显然有 $f(k+1,0)\leqslant 2f(k,0)+1$.

定理 5. $f(3k,4^{k-1}d)<3\times 4^{k-1}(f(k,d)+1)$.

证明:假定存在一个不可满足的 d-正则$(k,f(k,d)+1)$-CNF 公式Φ,包含变元 $x_1,x_2,\ldots,x_n$.

构造一个 0-正则(3,6)-CNF 公式 Φ',用矩阵表示如下:

$$\begin{array}{c} y_1 \\ y_2 \\ y_3 \\ y_4 \end{array}\begin{pmatrix} + & - & + & + & & & - & - \\ + & - & - & - & + & + & & \\ + & - & & & - & - & + & + \\ & & + & - & + & - & + & - \end{pmatrix}.$$

显然,公式 Φ' 是一个极小不可满足公式.将公式 Φ' 分裂为两个公式 Φ_1' 和 Φ_2',用矩阵分别表示如下:

$$\Phi_1':\begin{array}{c} y_1 \\ y_2 \\ y_3 \\ y_4 \end{array}\begin{pmatrix} + & - & + & \\ + & & - & + \\ + & + & & - \\ & - & + & + \end{pmatrix},\ \Phi_2':\begin{array}{c} y_1 \\ y_2 \\ y_3 \\ y_4 \end{array}\begin{pmatrix} - & - & + & \\ - & & - & + \\ - & + & & - \\ & + & - & - \end{pmatrix}.$$

由极小不可满足公式的性质可知,满足 Φ_1' 的真值指派一定不可能满足 Φ_2',满足 Φ_2' 的真值指派一定不可能满足 Φ_1'.构造 Φ_a' 的不交拷贝 $\Phi_{a,j}'$, $a\in\{1,2\}$,j=1,2,...,n.其中,Φ_a' 中的变元 y_i 在 $\Phi_{a,j}'$ 中被替换为新变元 $y_{i,j}$.

用公式 $\Phi_{1,j}'$ 替换公式 Φ 中 x_j 的所有正出现,用公式 $\Phi_{2,j}'$ 替换公式 Φ 中 x_j 的所有负出现,构造一个 CNF 公式 Ψ.可知公式 Ψ 中有 $4n$ 个变元,每个子句有 $3k$ 个文字,每个变元出现次数为 $3\times4^{k-1}(f(k,d)+1)$,每个变元正负出现次数之差不超过 $4^{k-1}d$.因为 $\Phi_{1,j}'$ 和 $\Phi_{2,j}'$ 不能同时满足,所以公式 Ψ 必不可满足. □

当 d=0 时,显然有 $f(3k,0)<3\times4^{k-1}(f(k,0)+1)$.

如果将公式 Φ' 更换为下面矩阵所表示的公式:

$$\begin{array}{c} y_1 \\ y_2 \\ y_3 \end{array}\begin{pmatrix} + & + & - & - & + & + & - & - \\ + & - & - & + & - & + & + & - \\ + & + & - & - & - & - & + & + \end{pmatrix},$$

那么,按照上面的推理方法易得结论:$f(3k,d)\leqslant4^k f(k,d)$.

因为 $f(k)$ 没有找到易于处理的计算方法,为了寻找 $f(k)$ 的上界,Hoory[17]定义了一个可计算的函数:

$$f_1(k):=\max\{s:(k,s)-CNF\cap MU(1)=\varnothing\},$$

并使用非递增整数序列生成方法来计算 $f_1(k)$ 的大小,进而获得 $f(k)$ 的上界.我们借鉴 Hoory 提出的方法来探寻 $f(k,d)$ 的上界.

给出如下定义:

$$f_1(k,d):=\max\{s:d-\text{正则}(k,s)-CNF\cap MU(1)=\varnothing\},$$

显然有 $f_1(k,d)\geqslant f(k,d)$.

定义 $\sigma=(a_1,a_2,...,a_n)$ 是一个有限的非递增的正整数序列(简称序列),如果 $a_1\geqslant a_2\geqslant...a_n\geqslant1$.称 a_i 是 σ 的一个项,σ 的长度为 n,ε 表示空序列.对于给定的有限的非负整数序列 σ,σ^{ord} 表示通过删除 0 项和非递增排序后生成非递增正整数序列的操作.

假定公式 $F=\{C_1,C_2,...,C_m\}$, $0\leqslant|C_1|\leqslant|C_2|\leqslant...\leqslant|C_m|\leqslant k$.令 $n=\max\{i:|C_i|<k\}$,定义一个序列:

$$\textstyle\sum_k(F)=(k-|C_1|,k-|C_2|,...,k-|C_n|).$$

如果公式 F 是一个 k-CNF 公式,则 $\sum_k(F)=\varepsilon$.

接下来定义一种序列的二元生成方法 $N(s,d)$.

序列的二元生成规则 $N(s,d)$.

输入:序列 σ_1 和 σ_2.

输出:序列 σ.

规则:从序列 σ_i 中缩减 s_i 个项,获得新序列 σ_i', i=1,2.要求 $s_1+s_2\leqslant s, s_1\leqslant(s+d)/2, s_2\leqslant(s+d)/2$.输出 $\sigma=(\sigma_1'\sigma_2')^{ord}$.

定义 $N(s,d)$-推演树 T 是一棵根树,其结点都标记为序列,并且如果 σ_1 和 σ_2 是 σ 的孩子结点,则 σ_1 和 σ_2 能够通过二元生成规则 $N(s,d)$ 推出序列 σ.给定一个序列集 Γ 和序列 σ,$\Gamma\vdash_{N(s,d)}\sigma$ 表示存在一棵 $N(s,d)$-推演树 T,其树根标记为 σ,叶子被标记的序列组成了 Γ.图 1 显示了 $(3)\vdash_{N(3,1)}(1,1,1,1,1)$.

$$\dfrac{\dfrac{\dfrac{(3)\quad(3)}{(2,2)}\quad(3)}{(2,1,1)}\qquad\dfrac{(3)\quad(3)}{(2,2)}}{(1,1,1,1,1)}$$

Fig.1 A $N(3,1)$-derivation tree

图 1 一棵 $N(3,1)$-推演树

定理 6. $(k)\vdash_{N(s,d)}\sigma$,当且仅当存在一个公式 $F\in MU(1)$,使得 $\sum_k(F)=\sigma$ 且每个变元出现次数不超过 s,正负出现次数均不超过$(s+d)/2$.

证明:(⇒)既然$(k)\vdash_{N(s,d)}\sigma$,说明存在一棵以σ为根、叶子为(k)的 $N(s,d)$-推演树 T.假定树 T 的推理步为 n(推理步是一棵 $N(s,d)$-推演树的非叶子结点个数).基于推理步进行归纳证明.

当 n=1 时,σ显然就是(k).当公式 $F=\{\phi\}$ 时,$\sum_k(F)=(k)$.

假定 n>1 时,树 T 必然显示从σ_1 和σ_2 能够依据二元生成规则 $N(s,d)$推理出σ.假定

$$\sigma_1=(a_1,\ldots,a_j),\ \sigma_2=(a_{j+1},\ldots,a_m),\ \sigma=(c_1,\ldots,c_m),$$

令 $I\subseteq\{1,\ldots,m\}$ 是从σ_1 和σ_2 推理出σ所用到的下标集.

根据归纳假设,可知存在公式$F_1,F_2\in MU(1)$,使得$\sum_k(F_1)=\sigma_1,\sum_k(F_2)=\sigma_2$,且$F_1$和$F_2$中每个变元次数不超过 s,正负出现次数均不超过$(s+d)/2$.假定 F_1 和 F_2 中不共享变元(可通过重命名变元保证),且其中恰好 k 个文字的子句组成的子公式分别是 F_1' 和 F_2'. 令 $F_1=\{C_1,\ldots,C_j\}\cup F_1',F_2=\{C_{j+1},\ldots,C_m\}\cup F_2'$, 满足 $a_i=k-|C_i|,i=1,\ldots,m$. 引入一个新的变元 x,构造公式 $F=\{D_1,\ldots,D_m\}\cup F_1'\cup F_2'$, 其中,

$$D_i=\begin{cases}C_i\vee x, & \text{if } i\in I \text{ and } i\leqslant j,\\ C_i\vee\neg x, & \text{if } i\in I \text{ and } i>j,\\ C_i, & \text{otherwise.}\end{cases}$$

显然,$F\in MU(1)$且$\sum_k(F)=\sigma$.根据二元生成规则 $N(s,d)$,可知公式 F 中 x 出现次数不超过 s,正负出现次数均不超过$(s+d)/2$.

(⇐)假设公式 $F\in MU(1)$,满足$\sum_k(F)=\sigma$, 每个变元出现次数不超过 s,且正负出现次数均不超过$(s+d)/2$.基于 F 中变元的数目 n 来归纳证明$(k)\vdash_{N(s,d)}\sigma$.

当 n=1 时,公式 $F=\{x,\neg x\},\sum_k(F)=(k-1,k-1)$. 显然,$(k)\vdash_{N(s,d)}(k-1)$.

当 n>1 时,任取公式 F 中的一个变元 x.因为 $MU(1)$公式一定可以分裂成两个没有共享子句的 $MU(1)$子公式[22],假定存在一个变元 x,使得

$$F=\{x\vee f_1,\ldots,x\vee f_j,\neg x\vee g_1,\ldots,\neg x\vee g_m,F_1,F_2\},$$

且满足 $F_1'=\{f_1,\ldots,f_j,F_1\}$ 和 $F_2'=\{g_1,\ldots,g_m,F_2\}$ 都是 $MU(1)$,其中, $j+m\leqslant s,j\leqslant(s+d)/2,m\leqslant(s+d)/2$.

因为 F_1' 和 F_2' 的变元数一定小于公式 F 的变元数,根据归纳假设,则一定有$(k)\vdash_{N(s,d)}\sigma_i$,且满足$\sum_k(F_i)=\sigma_i$, i=1,2.在推理过程中 $f_1,\ldots,f_j$ 与σ_1 中数值变小的项一一对应,$g_1,\ldots,g_m$ 与σ_2 中数值变小的项一一对应.可见,从σ_1 和σ_2 依据二元生成规则 $N(s,d)$,能够推理出σ.即$(k)\vdash_{N(s,d)}\sigma$. □

定理 7. $f_1(k,d)=\min\{s:(k)\vdash_{N(s,d)}\varepsilon\}-1$.

证明:因为$(k)\vdash_{N(s,d)}\varepsilon$,由定理 6 可知,必然存在一个 k-CNF 公式 F,满足每个变元出现次数不超过 s,正负出现次数均不超过$(s+d)/2$.显然公式 F 是一个不可满足的(k,s)-CNF 公式.又因每个变元正负出现次数不超过$(s+d)/2$,通过增加新变元和增加新子句的方法一定可以构造一个不可满足的 d-正则(k,s)-CNF 公式.可得 $f_1(k,d)=\min\{s:(k)\vdash_{N(s,d)}\varepsilon\}-1$. □

Hoory[17]给出了计算$f_1(k)$及其加速计算的方法,结合这些方法,我们给出了关于$f_1(k,d)$的一些结论,见表 1.

图 2 展示了从(5)到ε使用二元生成规则 $N(8,4)$的推理过程,其中,$\sigma=\sigma_1\oplus_{8,4}\sigma_2$表示从$\sigma_1$和$\sigma_2$依据二元生成规则 $N(8,4)$能够推理出σ.显然,4-正则(5,8)-CNF 类中存在不可满足实例.但是我们还没有发现 2-正则(5,8)-CNF 公式类,0-正则(5,8)-CNF 公式类中存在不可满足实例.

Table 1 Upper and lower bounds of $f_1(k,d)$

表 1 $f_1(k,d)$的上下界

$4\leqslant f_1(4,0)\leqslant 8$	$5\leqslant f_1(5,0)\leqslant 10$	$7\leqslant f_1(6,0)<16$
$4\leqslant f_1(4,1)\leqslant 7$	$5\leqslant f_1(5,1)\leqslant 9$	$7\leqslant f_1(6,1)<15$
$4\leqslant f_1(4,2)\leqslant 6$	$5\leqslant f_1(5,4)\leqslant 8$	$7\leqslant f_1(6,2)<14$
$f_1(4,3)=4$		$7\leqslant f_1(6,3)<13$
		$7\leqslant f_1(6,6)<12$

$$\sigma_1=(5)$$
$$\sigma_2=\sigma_1\oplus_{8,4}\sigma_1=(4,4)$$
$$\sigma_3=\sigma_1\oplus_{8,4}\sigma_2=(4,3,3)$$
$$\sigma_4=\sigma_1\oplus_{8,4}\sigma_3=(4,3,2,2)$$
$$\sigma_5=\sigma_1\oplus_{8,4}\sigma_4=(4,3,2,1,1)$$
$$\sigma_6=\sigma_1\oplus_{8,4}\sigma_5=(4,3,2,1)$$
$$\sigma_7=\sigma_6\oplus_{8,4}\sigma_6=(3,3,2,2,1,1)$$
$$\sigma_8=\sigma_1\oplus_{8,4}\sigma_7=(4,2,2,1,1)$$
$$\sigma_9=\sigma_1\oplus_{8,4}\sigma_8=(4,3,1,1)$$
$$\sigma_{10}=\sigma_8\oplus_{8,4}\sigma_9=(3,3,2,1,1,1)$$
$$\sigma_{11}=\sigma_9\oplus_{8,4}\sigma_9=(3,3,2,2)$$
$$\sigma_{12}=\sigma_9\oplus_{8,4}\sigma_{11}=(3,2,2,2,1,1)$$
$$\sigma_{13}=\sigma_1\oplus_{8,4}\sigma_{12}=(4,2,1,1,1)$$
$$\sigma_{14}=\sigma_1\oplus_{8,4}\sigma_{10}=(4,2,2,1)$$
$$\sigma_{15}=\sigma_1\oplus_{8,4}\sigma_{13}=(4,3,1)$$
$$\sigma_{16}=\sigma_{14}\oplus_{8,4}\sigma_{15}=(3,3,2,1,1)$$
$$\sigma_{17}=\sigma_{15}\oplus_{8,4}\sigma_{16}=(3,2,2,2,1)$$
$$\sigma_{18}=\sigma_{15}\oplus_{8,4}\sigma_{17}=(3,2,2,1,1,1)$$
$$\sigma_{19}=\sigma_1\oplus_{8,4}\sigma_{18}=(4,2,1,1)$$
$$\sigma_{20}=\sigma_{19}\oplus_{8,4}\sigma_{19}=(3,3,1,1)$$
$$\sigma_{21}=\sigma_{19}\oplus_{8,4}\sigma_{20}=(3,2,2,1)$$
$$\sigma_{22}=\sigma_{19}\oplus_{8,4}\sigma_{21}=(3,2,1,1,1)$$
$$\sigma_{23}=\sigma_1\oplus_{8,4}\sigma_{22}=(4,2,1)$$
$$\sigma_{24}=\sigma_{22}\oplus_{8,4}\sigma_{23}=(3,2,1,1)$$
$$\sigma_{25}=\sigma_{24}\oplus_{8,4}\sigma_{24}=(2,2,1,1)$$
$$\sigma_{26}=\sigma_{25}\oplus_{8,4}\sigma_{25}=(1,1,1,1)$$
$$\sigma_{27}=\sigma_{26}\oplus_{8,4}\sigma_{26}=\varepsilon$$

Fig.2 $N(8,4)$-derivation from (5) to ε

图 2 从(5)到ε的 $N(8,4)$-推演

5 结论和进一步工作

本文给出了一种从 k-CNF 公式到 d-正则(k,s)-CNF 公式的多项式时间归约方法,证明了如果 d-正则(k,s)-SAT 问题中存在不可满足的实例,则 d-正则(k,s)-SAT 问题就是 NP 完全问题.这说明,d-正则(k,s)-SAT 问题中存在着从可满足到 NP 完全问题的临界现象,对应的临界点称为临界函数 $f(k,d)$.这种多项式时间归约方法通过添加新变元和新子句,可以将任意的CNF公式的子句约束密度修改为s/k,并且让每个变元的正负出现次数差不超过 d.对临界函数$f(k,d)$的研究有助于在更多正则化约束条件下来寻找难解的 CNF 公式实例.

临界函数 $f(k,d)$不知是否可计算,但在文中 $f(k,d)$的上下界已被初步研究.下一步需要研究的问题是 0-正则(3,4)-SAT 问题中是否存在不可满足实例、更紧的$f(k,d)$的上下界以及 d 对正则(k,s)-SAT 问题的影响.

References:

[1] Cook SA. The complexity of theorem-proving procedures. In: Proc. of the 3rd Annual ACM Symp. on Theory of Computing. New York: ACM Press, 1971. 151−158. [doi: 10.1145/800157.805047]

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

[3] Crawford JM, Auton LD. Experimental results on the crossover point in satisfiability problems. In: Proc. of the 11th National Conf. on Artificial Intelligence. Washington: AAAI Press, 1993. 21–27. [doi: 10.1016/0004-3702(95)00046-1]

[4] Kirkpatrick S, Selman B. Critical behavior in the satisfiability of random boolean expressions. Science, 1994,264(5163):1297–1301. [doi: 10.1126/science.264.5163.1297]

[5] Xu K, Li W. Exact phase transitions in random constraint satisfaction problems. Journal of Artiial Intelligene Researh, 2000,12: 93–103. [doi: 10.1613/jair.696]

[6] Xu K, Li W. Many hard examples in exact phase transitions. Theoretical Computer Science, 2006,355(3):291–302. [doi: 10.1016/j.tcs.2006.01.001]

[7] Liu T, Lin X, Wang C, *et al*. Large hinge width on sparse random hypergraphs. In: Proc. of the 22nd Int'l Joint Conf. on Artificial Intelligence. Washington: AAAI Press, 2011. [doi: 10.5591/978-1-57735-516-8/IJCAI11-109]

[8] Liu T, Wang C, Xu W. Balanced random constraint satisfaction: Phase transition and hardness. In: Chen J, Lu P, eds. Frontiers in Algorithmics. Berlin, Heidelberg: Springer-Verlag, 2018. 238–250. [doi: 10.1007/978-3-319-78455-7_18]

[9] 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]

[10] 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]

[11] Sumedha, 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]

[12] Krishnamurthy S, Sumedha. Exact satisfiability threshold for k-satisfiability problems on a Bethe lattice. Physical Review E, 2015, 92(4):042144. [doi: 10.1103/PhysRevE.92.042144]

[13] Rathi V, Aurell E, Rasmussen LK, Skoglund M. Bounds on threshold of regular random *k*-SAT. In: Strichman O, Szeider S, eds. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer-Verlag, 2010,6175:264−277. [doi: 10.1007/978-3-642-14186-7_22]

[14] 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]

[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, 2017,45(12):7–13 (in Chinese with English abstract). [doi: 10.13245/j.hust.171202]

[16] Kratochvíl J, Savický P, Tuza Z. One more occurrence of variables makes satisfiability jump from trivial to NP-complete. SIAM Journal on Computing, 1993,22(1):203–210. [doi: 10.1137/0222015]

[17] 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]

[18] Savický P, Sgall J. DNF tautologies with a limited number of occurrences of every variable. Theoretical Computer Science, 2000,238(1):495–498. [doi: 10.1016/s0304-3975(00)00036-0]

[19] Gebauer H, Szabo T, Tardos G. The Local Lemma is asymptotically tight for SAT. Journal of the ACM, 2016,63(5):664–674. [doi: 10.1145/2975386]

[20] Hoory S, Szeider S. A note on unsatisfiable *k*-CNF formulas with few occurrences per variable. Society for Industrial and Applied Mathematics, 2006. 523–528. [doi: 10.1137/S0895480104445745]

[21] Tovey CA. A simplified NP-complete satisfiability problem. Discrete Applied Mathematics, 1984,8(1):85–89. [doi: 10.1016/0166-218x(84)90081-7]

[22] Davydov G, Davydova I, Büning HK. An efficient algorithm for the minimal unsatisfiability problem for a subclass of CNF. Annals of Mathematics & Artificial Intelligence, 1998,23(3–4):229–245. [doi: 10.1023/A:1018924526592]

附中文参考文献:

[2] 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]

[14] 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 1 RSB 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: 10.13328/j.cnki.jos.005129]

符祖峰(1978－),男,河南南阳人,副教授,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.