

# 用于求解正则(3,4)-SAT 实例集的修正警示传播算法

余光伟 许道云

(贵州大学计算机科学与技术学院 贵阳 550025)

**摘要** 利用极小不可满足公式的临界特性,可以将任意的一个 3-CNF 公式多项式时间归约转换为一个正则(3,4)-CNF 公式,从而得到一个保留 NP 完全性的正则(3,4)-SAT 问题。警示传播算法(Warning Propagation, WP)在归约转换后的正则(3,4)-SAT 实例集上高概率收敛,但在任意一个实例上都无法判断公式的可满足性,因此算法求解失效。对于一个归约转换后的正则(3,4)-CNF 公式,每一变元出现的正负次数之差具有趋于稳定的结构特征,基于该特征,提出基于变元正负出现次数规则的 WP 算法来求解归约转换后的正则(3,4)-SAT 实例。实验结果表明,修正的 WP 算法对正则公式的可满足性判定有效,从而可以利用公式的正则性特征进一步研究 WP 算法的收敛性特征条件。

**关键词** 极小不可满足公式,正则(3,4)-SAT 问题,警示传播算法

中图分类号 TP301 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.11.050

## Modified Warning Propagation Algorithm for Solving Regular (3,4)-SAT Instance Sets

SHE Guang-wei XU Dao-yun

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

**Abstract** Based on the critical characterization of the minimal unsatisfiable formula, a 3-CNF formula can be reduced to a regular (3,4)-CNF formula within polynomial time. Thus the regular (3,4)-SAT problem is NP-complete. The warning propagation algorithm(WP) converges in high probability on the regular (3,4)-SAT instance sets by reducing, but it can't determine the satisfiability of the formula on any instance, so the algorithm fails to solve the problem. For a reduced regular (3,4)-CNF formula, the difference between the positive and negative occurrences of each variable has been found to be stable. With this feature, a WP algorithm based on the rule of positive and negative occurrences was proposed to solve the reduced regular (3,4)-SAT instances. The experimental results show that the modified WP algorithm is effective for regular formulas. Therefore, the regularity of the formula can be used to study the convergence of the WP algorithm.

**Keywords** Minimal unsatisfiable formula, Regular (3,4)-SAT problem, Warning propagation algorithm

## 1 引言

可满足性问题(Satisfiability Problem, SAT 问题)是理论计算机科学领域的核心问题,它是指给定一个合取范式(Conjunction Normal Form, CNF)公式  $F$ , 判定是否存在一组布尔真值指派使得  $F$  为真。在 SAT 问题中,限制每个子句长度为  $k$  的 SAT 问题为  $k$ -SAT 问题。

随机  $k$ -SAT 问题是  $k$ -SAT 问题的一个子集合,它在  $k$ -SAT 问题的典型计算复杂性研究中有着重要的作用。在随机  $k$ -SAT 问题中,公式的规模(如变元数、子句数)以及公式的结构(如子句长度、子句约束密度)等都会影响到随机  $k$ -SAT 问题实例的可满足性和求解难度。例如,当  $k=2$  时,该问题属于 P 问题。2-SAT 问题易于判定且存在着线性时间的判定算法<sup>[1]</sup>。当  $k=3$  时,该问题属于 NP 完全问题<sup>[2]</sup>。在最坏情形下, NP 完全问题没有多项式时间的求解算法。特别地,一个重要的结构参数是子句约束密度  $\alpha$ , 它表示一个随

机  $k$ -CNF 公式中子句个数  $m$  与变元个数  $n$  的比值。实验验证和理论分析表明<sup>[3-5]</sup>: 当  $N \geq 1$  时,随着约束密度  $\alpha$  的逐渐增大,可满足指数目将逐渐减少,存在着某个与  $k$  有关的临界值点  $\alpha_c(k)$ , 当  $\alpha > \alpha_c(k)$  时,  $F$  以高概率不可满足; 当  $\alpha < \alpha_c(k)$  时,  $F$  以高概率可满足。这种从满足到不可满足的突变现象,称为随机  $k$ -SAT 问题的相变(Phase Transition)现象,临界值  $\alpha_c(k)$  称为随机  $k$ -SAT 问题可满足的相变点,且在相变位置点  $\alpha_c(k)$  前存在某个阈值  $\alpha_c(k)$ , 当  $\alpha < \alpha_c(k)$  时,公式的可满足性判定可能在多项式时间内完成,这称为易解 SAT 区域,当  $\alpha > \alpha_c(k)$  时,公式的判定难度很大,这称为难解 SAT 区域。目前,人们还无法找到  $\alpha_c(k)$  的精确值,围绕经典的随机 3-SAT 问题,相关研究者使用统计物理学中的腔域方法<sup>[6]</sup>给出了  $\alpha_c(3)$  的近似值,约为 4.267<sup>[7]</sup>。

警示传播算法(Warning Propagation, WP)<sup>[8]</sup>是一种基于腔域方法而设计的求解 SAT 问题的迭代算法,它能有效求解  $\alpha < 3.5$  区域的随机 3-SAT 实例。WP 算法有效的原因在于

到稿日期:2018-07-12 返修日期:2018-09-21 本文受国家自然科学基金项目(61762019,61462001)资助。

余光伟(1993-),男,硕士生,主要研究方向为算法设计与分析;许道云(1959-),男,教授,博士生导师,主要研究方向为可计算性与计算复杂性, E-mail: dyxu@gzu.edu.cn(通信作者)。



注意:子公式  $H_1^{[x]} + H_2^{[x]} + \dots + H_{2(s+t)}^{[x]} + T_1^{[x]} + T_2^{[x]} + \dots + T_{s+t}^{[x]}$  含有  $41(s+t)$  个 3-子句,并且每一个新引入的变元  $z_{l,k}$  ( $1 \leq l \leq 2(s+t), 0 \leq k \leq 9$ ) 在  $F^{[x]}$  中恰好出现 4 次。

基于上述方法,可以逐步构造一个公式  $F^{[\rho_1, \rho_2, \dots, \rho_n]}$ ,其每个子句的长度恰为 3,每个变元在公式中出现的次数恰为 4,且  $F^{[\rho_1, \rho_2, \dots, \rho_n]} \in \text{CNF}(93m, 124m)$ 。

与原公式相比,一个归约转换下的正则(3,4)-CNF 公式的结构发生了很大的变化,这种变化主要体现在两个方面:1)归约转换后的正则(3,4)-CNF 公式中变元出现的正负次数之差所占的比例趋于稳定;2)归约转换后的正则(3,4)-CNF 公式所对应的因子图会形成很多的圈。

为叙述方便,在一个 CNF 公式中,我们引入如下记号:1) $\text{Var}_{(s,t)}^i$ ,表示变元  $x_i$  在 CNF 公式中正出现  $s$  次,负出现  $t$  次;2) $\#\text{Var}_{(s,t)}$ ,表示 CNF 公式中正出现  $s$  次,负出现  $t$  次的变元集合;3) $\#\#\text{Var}_{(s,t)}$ ,表示  $\text{Var}_{(s,t)}^i$  的大小,即  $\text{Var}_{(s,t)}^i$  中包含变元的个数。

设  $F \in \text{CNF}(n, m)$  是一个 3-CNF 公式,在  $F$  中正文字出现的总数为  $S$ ,负文字出现的总数为  $T$ 。 $F'$  是  $F$  归约转换下的正则(3,4)-CNF 公式,则在  $F'$  中有:

$$\begin{cases} \#\text{Var}_{(0,4)} = 3m \\ \#\text{Var}_{(1,3)} = 6m \\ \#\text{Var}_{(2,2)} = 54m + T \\ \#\text{Var}_{(3,1)} = 36m + S \end{cases}$$

一个 CNF 公式可以表示为一个二分图,称这个二分图为因子图。因子图的一侧由变元集构成,另一侧由子句集构成。通常用圆形来表示变元节点,用矩形来表示子句节点,若变元  $x_i$  在子句  $C_j$  中正出现,则用实边来连接  $x_i$  与  $C_j$ ,若变元  $x_i$  在子句  $C_j$  中负出现,则用虚边来连接  $x_i$  与  $C_j$ 。例如,一个正则(3,4)-CNF 公式  $F' = (x_1 \vee x_2 \vee x_3) \wedge (x_1 \vee \neg x_2 \vee x_3) \wedge (x_1 \vee \neg x_2 \vee \neg x_3) \vee (\neg x_1 \vee \neg x_2 \vee \neg x_3)$  的因子图如图 1 所示。其中,  $C_1 = (x_1 \vee x_2 \vee x_3)$ ,  $C_2 = (x_1 \vee \neg x_2 \vee x_3)$ ,  $C_3 = (x_1 \vee \neg x_2 \vee \neg x_3)$ ,  $C_4 = (\neg x_1 \vee \neg x_2 \vee \neg x_3)$ 。

对于归约转换后的正则(3,4)-SAT 实例,有如下定理。

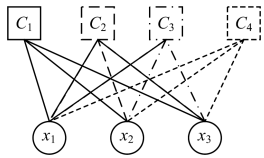


图 1 公式  $F'$  的因子图表示

Fig. 1 Factor graph representation of formula  $F'$

**定理 1** 假设公式  $F \in \text{CNF}(n, m)$  是一个 3-CNF 公式, $F$  中任意一个变元至少出现两次, $G$  是公式  $F$  所对应的因子图, $F'$  是  $F$  归约转换后的正则(3,4)-CNF 公式, $G'$  是  $F'$  所对应的因子图,则  $G'$  中的任意一个变元节点至少被包含在两个圈中。

证明:首先,证明  $G$  上的任意一个变元节点都至少被包含在一个圈中。可以断定图  $G$  上任意一个节点的度至少为 2。任意构造一条以变元  $x_1$  为起点的最长路径  $L: x_1 - C_2 - x_3 - C_4 \dots$ ,考察与  $x_1$  关联的边,这些边中任何一条边的另一端必然在  $L$  上,否则,将这个节点加入到  $L$  就可以得到一条

更长的路径。因为  $G$  中的每个节点的度数至少为 2,所以节点  $x_1$  要关联一条不在  $L$  上的边  $e$ ,这样,路径  $L$  上  $x_1$  到某子句节点  $C_j$  的子路就与边  $e$  构成了一个圈。

本文将  $G'$  上的变元节点分为两类:替换变元和引入变元。替换变元是指  $F$  中任意的变元  $x \in \#\text{Var}_{(s,t)}$  正则归约转换为  $F'$  后的变元集  $X^{[x]} = \{x_1, x_2, \dots, x_{s+t}\}$ ,引入变元是指  $F'$  中的其他变元。考察替换变元集  $X^{[x]}$  的局部结构,由正则归约转换技术, $X^{[x]}$  中的变元会形成一个包含  $2(s+t)$  个节点的圈(其中变元节点和子句节点各  $s+t$  个),称这类圈为 replace-cycle。则对于任意一个替换变元,它一定被包含在一个 replace-cycle 中。正则归约转换技术并未破坏  $G$  的整体结构,如果  $G$  中的任意两个节点存在一条连通的路,那么在  $G'$  中的替换变元之间也一定存在一条连通的路。同理, $G$  中的任意一个变元节点都至少被包含在一个圈中,那么  $G'$  中的替换变元也一定至少被包含在另一个圈中,称这类圈为 original-cycle。因此, $G'$  中的任一替换变元至少被包含在两个圈中。

最后考察  $G'$  中的引入变元。每一个引入变元的局部结构类似,以任意一个引入变元为起点,构造图  $G'$  上的一个游走,容易证明不同的游走至少会形成两个圈。

因此,因子图  $G'$  上的任一变元节点至少被包含在两个圈中。证毕。

### 3 警示传播算法在正则(3,4)-SAT 问题上的收敛性和有效性

前文已经提到,一个 CNF 公式  $F$  可以由一个二分图  $G = (X \cup Y, E)$  来表示,称为因子图。这里,变元节点集表示为  $X = \{1, 2, \dots, n\}$ ,子句节点集表示为  $Y = \{C_1, C_2, \dots, C_m\}$ 。 $G$  中的边分为两类:实边和虚边。

实边:  $(C_i, j) \in E \Leftrightarrow$  子句  $C_i$  包含正文字  $x_j$ 。

虚边:  $(C_i, j) \in E \Leftrightarrow$  子句  $C_i$  包含负文字  $\neg x_j$ 。

为叙述方便,本文用  $a, b, c, \dots$  来表示  $C_1, C_2, C_3, \dots$ ,并给出如下符号的表示意义。

$V(a)$  表示出现在子句  $a$  中的变元集,  $V(a) = V^+(a) \cup V^-(a)$ 。 $V^+(a)$  表示子句  $a$  中变元正出现的集合,  $V^-(a)$  表示子句  $a$  中变元负出现的集合。  $V(a) \setminus i = V(a) - i$ 。

$V(j)$  表示包含变元  $x_j$  的子句集合,  $V(j) = V^+(j) \cup V^-(j)$ 。 $V^+(j)$  表示变元  $x_j$  正出现的子句集合,  $V^-(j)$  表示变元  $x_j$  负出现的子句集合。  $V(j) \setminus a = V(j) - a$ 。

$J_j^a$  是一个标识参数,如果  $x_j \in a$ ,那么  $J_j^a = -1$ ;如果  $\neg x_j \in a$ ,那么  $J_j^a = 1$ 。

WP 算法是一种基于因子图的迭代算法。对于因子图中的每一条边  $(a, i)$ ,定义一个警示信息  $u_{a \rightarrow i}$ ,来表示变元  $x_i$  满足子句  $a$  的概率。警示信息更新函数的定义如下:

$$u_{a \rightarrow i}(t) = \prod_{j \in V(a) \setminus i} \theta(-J_j^a (\sum_{b \in V(j) \setminus a} J_j^b u_{b \rightarrow j}(t-1))) \quad (1)$$

其中,  $t$  是迭代步。  $\theta(x)$  定义为:如果  $x \leq 0$ ,那么  $\theta(x) = 0$ ;否则  $\theta(x) = 1$ 。如果  $V(a) \setminus i = \emptyset$ ,那么  $u_{a \rightarrow i} = 1$ 。如果  $V(j) \setminus a = \emptyset$ ,那么  $u_{a \rightarrow i} = 0$ 。通常式(1)被写为:

$$u_{a \rightarrow i}(t) = \prod_{j \in V(a) \setminus i} \theta(J_j^a (\sum_{b \in V^+(j) \setminus a} u_{b \rightarrow j}(t-1) - \sum_{b \in V^-(j) \setminus a} u_{b \rightarrow j}(t-1))) \quad (2)$$

如果所有的  $u_{a \rightarrow i}(t) = u_{a \rightarrow i}(t-1)$ , 那么就可以得到稳定的警示信息  $u_{a \rightarrow i}^* = u_{a \rightarrow i}(t)$ , 通过计算

$$H_i = - \sum_{b \in V^+(i)} J_{i,b}^b u_{b \rightarrow i}^* = \sum_{b \in V^+(i)} u_{b \rightarrow i}^* - \sum_{b \in V^-(i)} u_{b \rightarrow i}^* \quad (3)$$

就可以固定变元  $x_i$  的值: 如果  $H_i > 0$ , 则  $x_i = 1$ ; 如果  $H_i < 0$ , 则  $x_i = 0$ 。否则,  $x_i$  随机赋值为 0 或 1。

同时计算一个冲突标识  $c_i$ :

$$c_i = \begin{cases} 1, & \text{如果 } (\sum_{b \in V^+(i)} u_{b \rightarrow i}^*)(\sum_{b \in V^-(i)} u_{b \rightarrow i}^*) > 0 \\ 0, & \text{其他} \end{cases} \quad (4)$$

求解 SAT 问题的 WP 算法如算法 1 所示。

**算法 1** WP(for SAT Problem Instance  $F$ )

- Step1 构造公式  $F$  的因子图  $G$ , 变元集合  $X, V = \emptyset$ 。
- Step2 对  $G$  中的每一条边的警示信息  $u_{a \rightarrow i}$  随机赋值 0 或 1。
- Step3 for  $t=1$  to  $t=t_{\max}$  ( $t_{\max}$  是设置的最大迭代步, 用来强制算法终止)
  - Step3.1 以某种顺序, 使用式(2)更新警示信息  $u_{a \rightarrow i}(t)$ 。
  - Step3.2 对于每一条边  $a \rightarrow i$ , 如果  $u_{a \rightarrow i}(t) = u_{a \rightarrow i}(t-1)$ , 那么转向 Step4; 否则, 转向 Step5。
- Step4 如果  $t < t_{\max}$ , 得到一组稳定的警示信息  $u_{a \rightarrow i}^* = u_{a \rightarrow i}(t)$ 。
  - Step4.1 使用式(3)计算所有的  $H_i$ , 使用式(4)计算所有的  $c_i$ 。
  - Step4.2 如果存在  $c_i = 1$ , 返回 UNSAT, 否则, 转向 Step4.3。
  - Step4.3 如果存在  $H_i \neq 0$ , 用  $H_i$  固定  $x_i, S = \{i | \text{由 } H_i \text{ 的变元 } x_i\}$ , 使用变元  $x_i (i \in S)$  清洗图  $G$  为  $G'$  (对公式  $F$  进行化简), 构造图  $G'$  的变元集  $X'$ , 转向 Step4.5; 否则, 转向 Step4.4。
  - Step4.4 选择  $X$  中的一个变元  $x_i$ , 随机赋值 0 或 1,  $S' = \{i | \text{随机赋值的变元 } x_i\}$ , 使用变元  $x_i (i \in S')$  清洗图  $G$  为  $G'$ , 构造图  $G'$  的变元集  $X'$ 。
  - Step4.5 计算  $V = V \cup S \cup S', X = X', G = G'$ 。
  - Step4.6 如果  $X' \neq \emptyset$ , 转向 Step2。
  - Step4.7 返回  $V$  中变元的可满足指派。
- Step5 如果  $t = t_{\max}$ , 返回 UN-CONVERGENCE。

文献[14]表明, WP 算法在正则 (3,4)-SAT 实例集上高概率收敛。具体地, 在模型  $G(n, k, m)$  (其中,  $n$  表示变元个数,  $k$  表示子句长度,  $m$  表示子句个数) 产生的随机 3-SAT 实例集上, 当子句约束密度  $\alpha < 3.91$  时, WP 算法高概率收敛; 当子句约束密度  $\alpha > 3.91$  时, WP 算法高概率不收敛; 但在归约转换后的正则 (3,4)-SAT 实例集上, WP 算法的收敛性有了较大提高。本文取变元规模  $n = 40$ , 比较随着参数  $\alpha$  的变化, 随机 3-SAT 实例集和归约转换后的正则 (3,4)-SAT 实例集上 WP 算法的收敛性, 结果如图 2 所示。

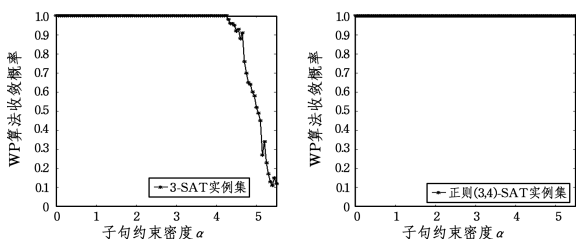


图 2 WP 算法在随机 3-SAT 实例集和归约转换下正则 (3,4)-SAT 实例集上的收敛性对比

Fig. 2 Convergence comparison between random 3-SAT instance sets and reduced regular (3,4)-SAT instance sets

虽然 WP 算法在归约转换后的正则 (3,4)-SAT 实例集

上高概率收敛, 但算法是否有效, 即 WP 算法得到的警示信息是否可以有效固定部分变元的赋值, 从而判断正则 (3,4)-SAT 问题的可满足性, 还需进一步验证。我们发现, WP 算法在归约转换后的正则 (3,4)-SAT 问题实例集上无法给出有效的警示信息, 算法高概率判断失效。具体地, 在随机产生模型  $G(n, k, m)$  上, 取变元规模  $n = 40$ , 比较在参数  $\alpha \leq 3.5$  的范围内, 随机 3-SAT 实例集和其归约转换下正则 (3,4)-SAT 实例集上 WP 算法能返回变元可满足指派的概率, 结果如图 3 所示。

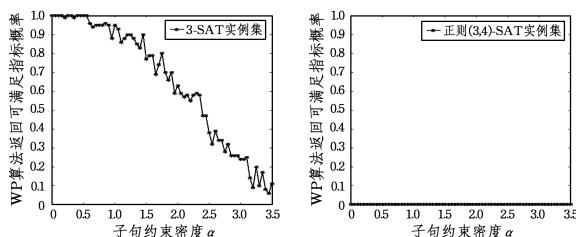


图 3 WP 算法在随机 3-SAT 实例集和归约转换下正则 (3,4)-SAT 实例集上求解有效性的对比

Fig. 3 Effectiveness comparison between random 3-SAT instance sets and reduced regular (3,4)-SAT instance sets

从图 3 可以看出, WP 算法在归约转换后的正则 (3,4)-SAT 实例集上无法有效给出变元的可满足指派。这是因为 WP 算法本质上是一种随机算法, 算法收敛得到的稳定警示信息实际上只能固定其中的部分变元, 其他的变元只能随机赋值。随着子句约束密度  $\alpha$  的增大, 变元可满足指派的个数将逐渐减少。若随机赋予变元一个不可满足的值, 则会导致算法判断失效, 这也解释了为什么尽管 WP 算法收敛但不是总能返回一个可满足赋值指派。在以极小不可满足公式为构造工具的正则 (3,4)-SAT 实例集上, 由于极小不可满足公式的临界特性, 有大量的变元赋值是确定的。WP 算法对这些确定变元的随机赋值就导致了算法的判定失效, 这正是本文对算法修正的方向。

**4 修正的警示传播算法求解正则 (3,4)-SAT 问题**

针对 WP 算法在正则 (3,4)-SAT 实例集上返回可满足指派的失效的问题, 本文提出一种基于变元正负出现次数的修正策略, 其算法如算法 2 所示。

**算法 2** WP(for Regular (3,4)-SAT Instance  $F'$ , Reduced by 3-SAT Instance  $F$ )

- Step1 构造公式  $F$  的因子图  $G$ , 变元集合  $X, V = \emptyset$ 。
- Step2 标记  $G$  中属于由  $F$  中同一变元归约转换下的变元集  $O_k (k = 1, 2, \dots, n)$ 。
- Step3 对  $G$  中的每一条边的警示信息  $u_{a \rightarrow i}$  随机赋值 0 或 1。
- Step4 for  $t=1$  to  $t=t_{\max}$ 
  - Step4.1 以某种顺序, 使用式(2)更新警示信息  $u_{a \rightarrow i}(t)$ 。
  - Step4.2 对于每一条边  $a \rightarrow i$ , 如果  $u_{a \rightarrow i}(t) = u_{a \rightarrow i}(t-1)$ , 那么转向 Step5; 否则, 转向 Step6。
- Step5 如果  $t < t_{\max}$ , 得到一组稳定的警示信息  $u_{a \rightarrow i}^* = u_{a \rightarrow i}(t)$ 。
  - Step5.1 使用式(3)计算所有的  $H_i$ , 使用式(4)计算所有的  $c_i$ 。
  - Step5.2 如果存在  $c_i = 1$ , 那么返回 UNSAT, 否则, 转向 Step5.3。
  - Step5.3 计算  $X$  中每一变元的  $\text{Var}_i^{\pm}(s, t)$ 。

Step5.4 如果存在  $x_i \in \# \text{Var}_{(s,0)}$ , 则  $x_i = 1$ ; 如果存在  $x_i \in \# \text{Var}_{(0,1)}$  或  $x_i \in \# \text{Var}_{(1,3)}$ , 则  $x_i = 0$ . 令  $P = \{i | x_i \in \# \text{Var}_{(s,0)} \cup \# \text{Var}_{(0,1)} \cup \# \text{Var}_{(1,3)}\}$ , 使用变元  $x_i (i \in P)$  清洗图  $G$  为  $G'$ , 构造图  $G'$  的变元集  $X'$ , 计算  $X = X', G = G', V = V \cup P$ , 转向 Step3; 否则, 转向 Step5.5.

Step5.5 遍历  $X$  中的变元, 当某一变元  $x_i \in \# \text{Var}_{(3,1)}$ , 则  $x_i = 0$ , 使用变元  $x_i$  清洗图  $G$  为  $G'$ , 构造图  $G'$  的变元集  $X'$ , 计算  $X = X', G = G', V = V \cup \{x_i\}$ , 转向 Step3; 否则, 转向 Step5.6.

Step5.6 如果存在  $H_i \neq 0$ , 用  $H_i$  固定  $x_i$ ,  $S = \{i | \text{由 } H_i \text{ 固定的变元 } x_i\}$ , 使用变元  $x_i (i \in S)$  清洗图  $G$  为  $G'$ , 构造图  $G'$  的变元集  $X'$ , 转向 Step5.8; 否则, 转向 Step5.7.

Step5.7 选择  $X$  中的一个变元  $x_i$ , 如果  $x_i \in O_k (k=1, \dots, n)$ , 那么随机赋值  $O_k$  中所有变元都为 0 或 1; 否则  $x_i = 0, S' = \{i | \text{计算 Step5.6 固定的变元}\}$ , 使用变元  $x_i (i \in S')$  清洗图  $G$  为  $G'$ , 构造图  $G'$  的变元集  $X'$ .

Step5.8 计算  $V = V \cup S \cup S', X = X', G = G'$ .

Step5.9 如果  $X' \neq \emptyset$ , 转向 Step3.

Step5.10 返回  $V$  中变元的可满足指派.

Step 6 如果  $t = t_{\max}$ , 返回 UN-CONVERGENCE.

该算法主要的修正是在 Step5.4 和 Step5.7 中. 在 Step5.4 中, 当变元  $x_i \in \# \text{Var}_{(s,0)}$  或  $x_i \in \# \text{Var}_{(0,1)}$  时,  $x_i$  是公式中的纯文字, 删除  $x_i$  所在的子句将不影响公式的可满足性判定; 当  $x_i \in \# \text{Var}_{(1,3)}$  时, 正则归约转换的规则迫使  $x_i = 0$  才能使公式可满足. 在 Step5.7 中, 当  $x_i \in O_k (k=1, \dots, n)$  时, 由正则归约转换的规则,  $O_k$  中所有变元取值都一样才能使公式可满足.

## 5 实验结果与分析

基于上文提出的求解算法, 取  $G(n, k, m)$  模型上正则归约转换下的两组实例集进行测试, 算法返回的可满足指派概率曲线如图 4 所示.

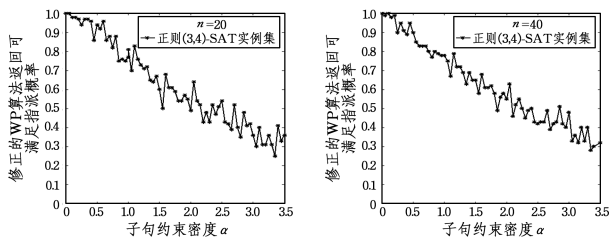


图 4 修正的 WP 算法在两组正则 (3,4)-SAT 实例集上返回可满足指派的概率

Fig. 4 Assignment probability of modified WP algorithm on two regular (3,4)-SAT instance sets

图 4 中的每个数据点是由  $G(n, k, m)$  模型产生的 100 个随机实例构成, 实验环境为: CPU 3.20 GHz, 4.00 GB 内存, Win 10 操作系统, C 语言编程, CodeBlock 编译. 变元规模分别取  $n=20$  和  $n=40$ .

从实验结果可以看出: 修正的 WP 算法在正则 (3,4)-SAT 实例集上有效.

修正 WP 算法的有效性表明了 WP 算法在求解 SAT 问题时与公式的结构有着很大的联系. 在不改变公式可满足性的前提下, 正则归约转换技术加长了 CNF 公式对应的因子图

上变元到变元、变元到子句、子句到变元的路径, 从而使 WP 算法的收敛性有了较大的提高; 但同时也会使 WP 算法求解失效, 这是因为在归约转换的正则 (3,4)-SAT 实例集上有着大量的变元在公式任何的可满足赋值中是固定的, WP 算法无法给出这些所有固定变元确定的赋值, 所以导致了算法的失效. 本文提出的修正算法是对公式的一种分解, 其分别作用于分解后独立的子公式, 从而使得算法再次有效.

**结束语** 本文针对 WP 算法在求解归约转换下正则 (3,4)-SAT 问题上的失效, 提出了一种修正策略, 得到了一种 WP 修正算法. 实验证明了修正的 WP 算法有效. 本文的结论有助于进一步理解 WP 算法, 结合其他的方法设计出求解 SAT 问题更有效的算法.

本文提出该算法的主要目的是对 WP 算法的收敛性进行理论分析或改进 WP 算法. 前文提到, WP 算法之所以有效, 关键就在于算法能够收敛. 基于 WP 算法在归约转换后的正则 (3,4)-SAT 实例集上近乎完全收敛的实验现象, 我们相信正则 (3,4)-CNF 公式的结构性质会对算法收敛性的理论分析有着潜在的帮助. 同时, 算法 1 对本文中正则公式可满足性判定的失效也表明 WP 算法仍有改进的余地. 修正策略本质上是对公式的一种分解, 可以理解为把公式对应的因子图分离为若个不同“区域”, 修正的 WP 算法分别作用于这些独立的“区域”, 从而使算法 1 对公式可满足性的判定再次有效.

受此启发, 在使用 WP 算法对任意一个 CNF 公式进行求解时, 我们能否将公式分离为不同的“区域”, 然后更改 WP 算法的信息传递方式为“区域”内的传播, 使之具有更好的求解效果? 基于 WP 修正算法, 我们将进一步利用 CNF 公式的正则性深入研究 WP 算法的收敛性特性与收敛条件的本征性质.

## 参考文献

- [1] ASPVALL B, PLASS M F, TARJAN R E. A linear-time algorithm for testing the truth of certain quantified boolean formulas [J]. Information Processing Letters, 1979, 8(3): 121-123.
- [2] COOK S A. The complexity of theorem-proving procedures [C] // Proceedings of the 3rd Annual ACM Symposium on the Theory of Computing (STOC'71). 1971: 151-158.
- [3] CRAWFORD J M, ANTON L D. Experimental results on the crossover point in satisfiability problems [C] // Proceedings of the 11th National Conference on Artificial Intelligence (AAAI-93). Washington, DC, USA, 1993: 21-27.
- [4] KIRKPATRICK S, SELMAN B. Critical behavior in the satisfiability of random boolean expressions [J]. Science, 1994, 264(5163): 1297-1301.
- [5] FRIEDGUT E, BOURGAIN J. Sharp thresholds of graph properties, and the k-sat problem [J]. Journal of the American Mathematical Society, 1999, 12(4): 1017-1055.
- [6] MÉZARD M, PARISI G. The Cavity Method at Zero Temperature [J]. Journal of Statistical Physics, 2003, 111(1-2): 1-34.
- [7] MÉZARD M, PARISI G, ZECCHINA R. Analytic and algorithmic solution of random satisfiability problems [J]. Science, 2002, 297(5582): 812-815.
- [8] BRAUNSTEIN A, MÉZARD M, ZECCHINA R. Survey propa-

gation: An algorithm for satisfiability[J]. *Random Structures and Algorithms*, 2005, 27(2): 201-226.

- [9] FEIGE U, MOSSEL E, VILENCHIK D. Complete convergence of message passing algorithms for some satisfiability problems [C]// *Proceedings of Random06*. 2008: 339-350.

- [10] WANG X F, XU D Y, WEI L. Convergence of Warning Propagation Algorithms for Random Satisfiable Instances[J]. *Journal of Software*, 2013, 24(1): 1-11. (in Chinese)

王晓峰, 许道云, 韦立. 随机可满足实例集上警示传播算法的收敛性[J]. *软件学报*, 2013, 24(1): 1-11.

- [11] WANG X F, XU D Y. Sufficient Conditions for Convergence of the Warning Propagation Algorithm[J]. *Journal of Software*, 2016, 27(12): 3003-3013. (in Chinese)

王晓峰, 许道云. 警示传播算法收敛的充分条件[J]. *软件学报*, 2016, 27(12): 3003-3013.

- [12] XU D Y. Applications of minimal unsatisfiable formulas to polynomially reduction for formulas[J]. *Journal of Software*, 2006, 17(5): 1204-1212. (in Chinese)

许道云. 极小不可满足公式在多项式归约中的应用[J]. *软件学报*, 2006, 17(5): 1204-1212.

- [13] XU D Y, WANG X F. A Regular NP-Complete Problem and Its Inapproximability[J]. *Journal of Frontiers of Computer Science & Technology*, 2013, 7(8): 691-697. (in Chinese)

许道云, 王晓峰. 一个正则 NP-完全问题及其不可近似性[J]. *计算机科学与探索*, 2013, 7(8): 691-697.

- [14] WANG X F, QIANG L I, DING H S. Convergence of Warning Propagation Algorithm for Regular Structure Instances[J]. *Computer Science*, 2015, 42(1): 279-284. (in Chinese)

王晓峰, 李强, 丁红胜. 规则实例集上警示传播算法的收敛性[J]. *计算机科学*, 2015, 42(1): 279-284.

(上接第 303 页)

零元素的比例仅约为  $10^{-6}$ 。综合两个数据集上的实验结果得出, 在面对不同稀疏程度的真实张量数据时, 本文给出的 ParSCP-ALS 算法相较于现有并行和串行算法获得更好的性能。

**结束语** 本文提出的结合 GPU 技术的并行 CP 张量分解 ParSCP-ALS 算法是对 CP-ALS 张量分解算法进行改进后的并行实现。首先, 对 CP-ALS 中的 Tensor Khatri-Rao 乘积进行简化, 并给出 SCP-ALS 算法; 然后, 结合张量数据的稀疏特性与 SCP-ALS 算法, 实现原 CP-ALS 算法的并行化, 也就是 ParSCP-ALS 算法。基于模拟数据集和真实数据集的算法性能对比实验, 都验证了 ParSCP-ALS 算法能有效提升 CP 张量分解的计算效率。模拟数据集和真实数据集上的实验结果都表明, 相比于现有的并行张量算法, ParSCP-ALS 算法有着更优的性能, 能够高效地处理大规模稀疏张量的分解问题。

## 参 考 文 献

- [1] LI X T. Research on Key Technologies of Multi-Relational Graph Mining Based on Tensors[D]. Harbin: Harbin Institute of Technology, 2013. (in Chinese)

李旭涛. 基于张量的多关系图挖掘关键技术研究[D]. 哈尔滨: 哈尔滨工业大学, 2013.

- [2] YANG B, LIU D Y, LIU J M, et al. Complex Network Clustering Algorithms[J]. *Journal of Software*, 2009, 20(1): 54-66. (in Chinese)

杨博, 刘大有, LIU J M, 等. 复杂网络聚类方法[J]. *软件学报*, 2009, 20(1): 54-66.

- [3] MØRUP M. Applications of tensor (multiway array) factorizations and decompositions in data mining[J]. *Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery*, 2011, 1(1): 24-40.

- [4] HITCHCOCK F L. The expression of a tensor or a polyadic as a sum of products[J]. *Journal of Mathematics and Physics*, 1927, 6(1-4): 164-189.

- [5] KOLDA T G, BADER B W. Tensor Decompositions and Appli-

cations[J]. *Siam Review*, 2009, 51(3): 455-500.

- [6] HARSHMAN R A. Foundations of the PARAFAC procedure: Model and conditions for an ‘explanatory’ multi-mode factor analysis [J]. *UCLA Working Papers in Phonetics*, 1970, 16: 1-84.

- [7] CARROLL J D, PRUZANSKY S, KRUSKAL J B. CANDIDINC: a general approach to multi dimension analysis of many-way arrays with linear constraints on parameters[J]. *Psychometrika*, 1980, 45(1): 3-24.

- [8] PAPALEXAKIS E E, FALOUTSOS C, SIDIROPOULOS N D. ParCube: Sparse Parallelizable Tensor Decompositions[J]. *Acm Transactions on Knowledge Discovery from Data*, 2012, 10(1): 521-536.

- [9] ANTIKAINEN J, HAVEL J, JOSTH R, et al. Nonnegative Tensor Factorization Accelerated Using GPGPU [J]. *IEEE Transactions on Parallel & Distributed Systems*, 2011, 22(7): 1135-1144.

- [10] ZOU B, LI C, TAN L, et al. GPURTENSOR: Efficient tensor factorization for context-aware recommendations[J]. *Information Sciences*, 2015, 299: 159-177.

- [11] KANG U, PAPALEXAKIS E, HARPALE A, et al. GigaTensor: scaling tensor analysis up by 100 times-algorithms and discoveries [C] // *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2012: 316-324.

- [12] COOK S. CUDA Programming-A Developer’s Guide to Parallel Computing with GPUs[M]. Beijing: China Mechine Press, 2014. (in Chinese)

COOK S. CUDA 并行程序设计[M]. 北京: 机械工业出版社, 2014.

- [13] BELL N, GARLAND M. Efficient Sparse Matrix-Vector Multiplication on CUDA [OL]. [http://www.users.csbsju.edu/~mheroux/fau2013\\_csci317/SPMV-Garland-Bell.pdf](http://www.users.csbsju.edu/~mheroux/fau2013_csci317/SPMV-Garland-Bell.pdf).

- [14] LEY M. DBLP: some lessons learned[J]. *Proceedings of the VLDB Endowment*, 2009, 2(2): 1493-1500.

- [15] HARPER F M, KONSTAN J A. The MovieLens Datasets: History and Context[J]. *ACM Transactions on Interactive Intelligent Systmes (TiIS)*, 2015, 5(4): 19.