

警示传播算法收敛的充分条件^{*}

王晓峰^{1,2}, 许道云²

¹(北方民族大学 计算机科学系, 宁夏 银川 750021)

²(贵州大学 计算机科学系, 贵州 贵阳 550025)

通讯作者: 许道云, E-mail: dyxu@gzu.edu.cn



摘要: 信息传播算法求解可满足问题时会有惊人的效果, 难解区域变窄. 然而, 因子图带有环的实例, 信息传播算法不总有效, 常表现为不收敛. 对于这种现象, 至今缺少系统的理论解释. 警示传播(warning propagation, 简称 WP)算法是一种基础的信息传播算法, 对 WP 算法的收敛性研究是其他信息传播算法收敛性研究的重要基础. 在 WP 算法中, 将警示信息的取值从 $\{0, 1\}$ 松弛为 $[0, 1]$, 利用压缩函数的性质, 给出了 WP 算法收敛的一个充分条件. 选取了两组不同规模的随机 3-SAT 实例进行实验模拟, 结果表明: 当子句与变元的比值 $\alpha < 1.8$ 时, 该判定条件有效.

关键词: 警示传播算法; 收敛性; 可满足性问题; 因子图

中图法分类号: TP301

中文引用格式: 王晓峰, 许道云. 警示传播算法收敛的充分条件. 软件学报, 2016, 27(12): 3003-3013. <http://www.jos.org.cn/1000-9825/4940.htm>

英文引用格式: Wang XF, Xu DY. Sufficient conditions for convergence of the warning propagation algorithm. Ruan Jian Xue Bao/Journal of Software, 2016, 27(12): 3003-3013 (in Chinese). <http://www.jos.org.cn/1000-9825/4940.htm>

Sufficient Conditions for Convergence of the Warning Propagation Algorithm

WANG Xiao-Feng^{1,2}, XU Dao-Yun²

¹(Department of Computer Science, Beifang Minzu University, Yinchuan 750021, China)

²(Department of Computer Science, Guizhou University, Guiyang 550025, China)

Abstract: Message propagation algorithms are very effective in finding satisfying assignments for SAT instances, and hard region become narrower. However, message propagation algorithms do not always converge for graphs with cycles. Unfortunately, rigorous theory proof of this phenomenon is still lacking. Warning propagation (WP) algorithm is the most basic message propagation algorithm, we have derived the sufficient conditions for convergence of the warning propagation algorithm. Lastly, experimental results show that such conditions for convergence of WP are very effective in different group data from random satisfiable 3-SAT instances.

Key words: warning propagation algorithm; convergence; satisfiability problem; factor graph

约束可满足性问题(constraint satisfaction problem, 简称 CSP)是人工智能中的重要研究领域^[1-3]. CSP 问题是三元组 $\langle X, D, C \rangle$: X 表示变量的集合, 记为 $X = \{x_1, \dots, x_n\}$; D 是关于变量 x_1, \dots, x_n 的取值域的集合, 记为 $D = \{D_1, \dots, D_n\}$; C 是约束的集合, 每一个约束 $c \in C$ 表示为 $c = \langle \sigma, \rho \rangle$, 其中, 约束范围 σ 是变量的列表, 约束关系 ρ 是关

* 基金项目: 国家自然科学基金(61462001, 61262006, 61402017); 宁夏自然科学基金(NZ14108); 北方民族大学基金(2014XYZ03, 2014XBZ04); “计算机应用技术”自治区重点学科项目

Foundation item: National Natural Science Foundation of China (61462001, 61262006, 61402017); Natural Science Foundation of Ningxia Province of China (NZ14108); Natural Science Foundation of Beifang University of Nationalities (2014XYZ03, 2014XBZ04); Key Subject Foundation of Computer application technology of Ningxia

收稿时间: 2014-11-28; 修改时间: 2015-07-27; 采用时间: 2015-11-16; jos 在线出版时间: 2015-11-27

CNKI 网络优先出版: 2015-11-26 16:06:16, <http://www.cnki.net/kcms/detail/11.2560.TP.20151126.1606.005.html>

于 σ 中变量取值域的笛卡尔积的子集.可满足性判定问题(satisfiability,简称 SAT)是典型的 CSP 问题,即:给定一个合取范式公式(conjunctive normal form,简称 CNF) F ,SAT 判定问题指是否存在一组指派使得 F 为真.

SAT 问题是计算机科学中的核心问题,其理论和应用研究是计算机与数理逻辑界学者共同关注的一个重大问题.随着 SAT 问题的深入研究,各类具有特殊结构的 SAT 问题受到越来越多的学者重视.限制子句的长度或变量出现的次数,实质上是在特殊结构下研究 SAT 问题,目前研究较多的是 k -SAT 问题.当 $k \geq 3$ 时, k -SAT 问题是著名的 NP 完全问题.已经发现:在随机 k -SAT 实例中,子句个数与变元个数的比值 a 是一个重要参数,它不仅对公式的可满足性产生影响,还对公式可满足性判定的难易程度产生影响.一方面,在可满足性问题的相变现象的研究中,随机统计现象表明:满足与不可满足之间出现一种临界现象,对于随机 k -SAT 问题,存在可满足的相变点 a_d ,当 $a < a_d$ 时,实例高概率可满足;当 $a > a_d$,实例高概率不可满足^[4,5].在随机 3-SAT 问题中,尽管人们不知道 a_d 的确切值,但研究结果表明: a_d 至少为 3.52^[6],至多为 4.506^[7].另一方面,大多数优秀的可满足性判定算法能够解决的实例与 a 密切相关.如:基于纯文字规则的算法能够有效求解 $a < 1.6$ 区域的随机 3-SAT 实例(如 DPLL 算法);基于正文字或负文字出现次数最多而设计的算法可有效求解 $a < 3.52$ 区域的随机 3-SAT 实例(如 Zchaff 算法);WALKSAT 算法可有效求解 a 稍高于 3.95 区域的随机 3-SAT 实例^[8,9].

近年来,使用统计物理中的腔域方法^[10,11]给出了可满足相变点的近似值 $a_d \approx 4.267$,且存在阈值 $a_c \approx 3.921$.称 $a < a_c$ 为易解 SAT 区域,称 $a_c < a < a_d$ 为难解 SAT 区域.位于 $a < a_c$ 区域的可满足实例的解空间形成一个较大的解集簇,启发式局部搜索算法容易求解;位于 $a_c < a < a_d$ 区域的实例,解空间被分解成许多很小的解集簇,这些解集簇之间相互分开,且相距很远,不可能通过改变有限数目变量的赋值,使得可满足指派从一个解集簇转到另一个解集簇,使用局部搜索算法求解时收敛速度很慢,且求解难度较大^[12].SAT 问题的相变现象进一步催生了概率方法在 SAT 问题中的研究和应用.基于腔域方法设计的求解可满足性问题的信息传播算法,在难解区域的随机 3-SAT 实例上非常有效.如:置信传播(belief propagation,简称 BP)算法可有效求解 $a < 3.95$ 区域的随机 3-SAT 实例;调查传播(survey propagation,简称 SP)算法能够有效求解 $a < 4.26$ 区域的随机 3-SAT 实例.SP 算法是目前求解随机 3-SAT 问题最为有效的算法,能够在稍高于线性时间内求解位于难解 SAT 区域的具备 10^7 变量规模的实例,几乎能够有效求解接近相变点的可满足性实例.对 SP 算法而言,难解 SAT 区域变窄^[12-15].

信息传播算法不仅对固定定义域的随机可满足性问题非常有效(如随机 k -SAT 问题),还可用于求解具有增长定义域的随机可满足性问题.许可等人在 NP 完全问题相变现象的研究中,提出了具有精确相变和难解实例的 RB 模型.该模型是典型的具有增长定义域的随机可满足实例产生模型,被广泛应用于构造难解性问题^[16].随后,文献[17,18]分别从理论上和实验上说明:在相变的阈值附近,求解 RB 模型随机实例的难度随着问题规模的增加以指数形式增长.文献[19]表明,基于变量熵的 BP 算法在求解 RB 模型实例时非常有效.同样,信息传播算法可用于求解量化布尔公式(quantified boolean formulae,简称 QBF)的可满足性.殷明浩等人设计了一种基于 SP 的启发式算法 HSPQBF 用于求解 QBF 问题,实验结果表明,该算法有效^[20].

近几年,信息传播算法得到了计算机界学者的重视,已取得了一些令人瞩目的研究成果.例如,Ravanbakhsh 和 Greiner 通过引入参数,提出了求解 CSP 问题的 PBP(perturbed BP)算法和 PSP(perturbed SP)算法.实验结果表明:在随机 CSP 实例上,PBP 算法效果优于 BP 算法.同样,对于随机 SAT 实例,PSP 算法也优于 SP 算法^[21].特别是在组合优化问题方面,信息传播算法的应用已取得丰硕的成果,如 MAX-SAT 问题、Maximum weight matching 问题、Minimum cost network flow 问题、图的着色、最大独立集等^[22-27].

尽管信息传播算法求解组合优化问题非常有效,但仍有两个问题有待在理论上深入研究:

- 1) 信息传播算法是否收敛;
- 2) 如果信息传播算法收敛,得到的结果是否有效,即,近似度如何.如果信息传播不收敛,这种情况下算法失效.

已经知道:因子图为树型结构的实例,信息传播算法能够有效收敛;但对于因子图含有多个回路的图型结构的实例,信息传播算法不总是有效,常表现为不收敛^[13].

目前,对信息传播算法的收敛性研究已取得了一些成果.如,文献[14]中给出了信息传播算法的一般形式,其

本质是变量边缘概率分布的一种近似.文献[28]中分析了有限树型结构上信息传播算法的复杂性.文献[29]中分析了因子图含有一个环的图型结构的实例上信息传播算法的收敛性,表明信息传播算法能够有效收敛,且得到的结果是变量边缘分布的有效近似.文献[30]表明:对于任意结构的高斯图模型实例,信息传播算法能够正确收敛.Tatikonda 和 Jordan 分析了基于计算树的 Gibbs 测度序列的唯一性与 BP 算法收敛性之间的关系,导出了 BP 算法收敛的一个充分条件^[31].Heskes 利用 Bethe 自由能量最小值的唯一性,给出了 BP 算法有唯一固定点的有效条件,分析了信息传播算法与统计物理中自由能量函数之间的关系,指出信息传播算法收敛后得到的固定点是自由能量函数的稳定点,并给出了 BP 算法收敛到唯一固定点的充分条件,但未对信息传播算法是否收敛做出理论分析^[32].最近几年,一些学者根据信息更新函数的特征,并利用函数的压缩映射原理,给出了信息传播算法收敛的一些充分条件.如,Ihler 直接从 BP 算法的信息更新函数出发,获得了传递错误信息的界,并利用该错误界给出了信息传播算法收敛的一个充分条件^[33].Mooij 和 Kappen 利用向量空间中信息更新函数的压缩映射原理导出了 BP 算法收敛的充分条件,并指出:如果更新函数系数矩阵的谱半径严格小于 1,则 BP 算法收敛^[34].Shi 和 Schonfeld 等人得到了传递错误信息的一个更紧致的界,利用这个界,研究了信息传播算法的动态行为,并给出了算法收敛的一个充分条件^[35].文献[23]表明:BP 算法在立方级复杂度内能够有效收敛到最大权重匹配问题的最优解.文献[25]中分析了求解最小代价网络流(MCF)的 BP 算法的收敛性,研究结果表明:在亚多项式时间内,BP 算法收敛到 MCF 问题的最优解,并且这个最优解具有唯一性.与此同时,文献中还提供了简化了的 BP 算法,能够对 MCF 问题给出多项式时间随机近似方案(FPRAS).文献[36]中给出了信息传播算法收敛的一个概率条件,研究结果表明:BP 算法能够给出最大权重匹配问题和最小代价网络流问题的最优解,且算法迭代次数高概率的有多项式界.文献[37]中分析了 BP 算法的复杂性,提出 SBP(stochastic BP)算法,该算法的消息更新复杂性较 BP 算法的复杂性低了一个指数级别.

警示传播(warning propagation,简称 WP)算法是最为基础的信息传播算法,对 WP 算法的收敛性分析,有助于其他信息传播算法的收敛性分析.文献[38]中,Feige 和 Mossel 等人分析了 WP 算法的收敛性,但主要局限于植入指派的随机可满足实例产生模型 $\mathcal{P}_{n,p}^{plant}$,即:给定一个指派,以概率 p 随机选取满足指派的子句之集构成实例.具体地讲,在 $\mathcal{P}_{n,p}^{plant}$ 模型中取概率 p 充分大,其本质是高概率地最多只有一个赋值满足该实例.文献[39]证明了 WP 算法在 $G(n,k,p)$ 模型产生的 3CNF 公式上的收敛性,但只证明了 p 取值很小的一部分实例.目前,对警示传播算法的性能及收敛性的理论分析仍然不完善.

将 WP 算法的警示信息取值从 $\{0,1\}$ 松弛为 $[0,1]$,利用向量空间中压缩函数的性质,并借鉴文献[40]中的方法,给出了 WP 算法收敛的一个充分条件.具体地讲,对于 CNF 公式 $F=\{a,b,\dots\}$,含有 n 个变元 x_1,x_2,\dots,x_n .公式 F 的因子图中边的序列集合记为 $D=\{a \rightarrow i:(a,i) \in E\}$,构造 $|D| \times |D|$ 的非负矩阵 A , A 中的对角元素为 $A_{a \rightarrow i,a \rightarrow i} = \sqrt{2}\delta$,其他元素为 $A_{a \rightarrow i,b \rightarrow j} = 1_{V(j),a}(b)1_{V(a),i}(j)\theta(-J_j^a J_j^b)$, δ 为常数.如果非负矩阵 A 的谱半径 $\rho(A)$ 严格小于 1,则 WP 算法收敛,且与初始信息无关.其中, $1_X(x)$ 是标识函数,若 $x \in X$,则 $1_X(x)=1$;若 $x \notin X$,则 $1_X(x)=0$.最后,选取了两组规模不同的随机 3-SAT 实例进行实验模拟,结果表明:当子句与变元的比值 $\alpha < 1.8$ 时,该判定条件有效.

1 警示传播算法

设 $F=\{C_1,C_2,\dots,C_m\}$ 为 CNF 公式,含有 n 个变元 x_1,x_2,\dots,x_n .用 i 代表变元 x_i .公式 F 可以用二分图 $G=(C \cup X,E)$ 表示,称 G 为因子图.其中,变元结点集为 $X=\{1,2,\dots,n\}$,子句结点集为 $C=\{C_1,C_2,\dots,C_m\}$.图 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 含正文字 x_j ;

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

为了书写方法,不妨用字母 a,b,\dots 分别代表子句 C_1,C_2,\dots

- $V(a)$:表示出现在子句 a 中的变元集合, $V(a)=V^+(a) \cup V^-(a)$.其中,
 - $V^+(a)$:表示出现在子句 a 中的正文字对应的变元标识集合;

- $V(a)$:表示出现在子句 a 中的负文字对应的变元标识集合;
- $V(a)\setminus i:V^+(a)-\{i\}$.
- $V(j)$:表示含变元 x_j 的子句集合, $V(j)=V^+(j)\cup V^-(j)$.其中,
 - $V^+(j)$: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,若 x_j 出现在子句 a 中,则 $J_j^a = -1$;若 $\neg x_j$ 出现在子句 a 中,则 $J_j^a = 1$.在因子图的每条边 (a,i) 上,定义 WP 算法中的消息传递 $u_{a\rightarrow i}$ (常称为警示信息). $u_{a\rightarrow i}$ 表示子句 a 的可满足性对变元 x_i 的取值倾向.WP 算法的消息更新迭代方程如下:

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

t 表示迭代次数, $\theta(x)$ 是截尾函数.如果 $x \leq 0$,则 $\theta(x)=0$;否则, $\theta(x)=1$.若 a 中仅包含变元 x_i ,则置 $u_{a\rightarrow i}=1$.当 WP 算法收敛时,根据警示信息固定变元 x_i 的赋值:

$$H_i = - \sum_{b \in V^-(i)} J_i^b u_{b\rightarrow i}^* \quad (2)$$

如果 $H_i > 0$,则 $x_i=1$;如果 $H_i < 0$,则 $x_i=0$;否则, x_i 暂不赋值.一般地,将公式(1)改写为如下形式:

$$u_{a\rightarrow i}(t) = \prod_{j \in V^-(a)\setminus i} \theta \left(J_j^a \left(\sum_{b \in V^+(j)\setminus a} u_{b\rightarrow j}(t-1) - \sum_{b \in V^-(j)\setminus a} u_{b\rightarrow j}(t-1) \right) \right) \quad (3)$$

设

$$h_{j\rightarrow a} = \sum_{b \in V^+(j)\setminus a} u_{b\rightarrow j}(t-1) - \sum_{b \in V^-(j)\setminus a} u_{b\rightarrow j}(t-1) \quad (4)$$

称 $h_{j\rightarrow a}$ 为腔域.若变元 x_j 仅出现在 a 中,则置 $h_{j\rightarrow a}=0$.求解 CNF 公式 F 的 WP(warning propagation)算法如下:

Warning Propagation (CNF formula F).

1. 构造相应的因子图 $G(F)$;
2. 给因子图上的所有消息边 $u_{a\rightarrow i}(t=0)$ 随机赋值 0 或 1;
3. 重复如下过程,直到算法收敛(可设置最大迭代步 t_{\max} 强迫算法结束):
 - 3.1 对 $G(F)$ 中的边随机排列;
 - 3.2 根据随机边序列,利用公式(1)更新消息 $u_{a\rightarrow i}$;
4. 根据 H_i 计算部分指派 ψ ,对公式 F 进行简化;
5. 返回 ψ ;

WP 算法中的收敛,是指算法某 t 次迭代得到的信息 $u_{a\rightarrow i}(t)$ 与 $t-1$ 次的信息 $u_{a\rightarrow i}(t-1)$ 一致.当算法收敛时,得到警示信息的固定点,进而可固定部分变元的赋值,并返回部分指派 ψ ;否则,算法不收敛,返回失败.可以看出,算法的收敛性对算法的性能起关键作用.有如下结论:

定理 1^[13]. 如果公式对应的因子图为树型结构,那么 WP 算法收敛.

然而,在带有环的图型结构的因子图实例上,WP 算法不总收敛.目前,对算法收敛性的理论分析仍然不完善.因此,有必要对 WP 算法的收敛性给予理论解释.

2 主要结论及证明

警示传播算法是一种迭代算法,算法的收敛性影响算法的性能.如果警示传播算法能够有效收敛,则可以用来固定部分变元的赋值,从而对原问题进行简化.然而,对于因子图含有回路的实例,警示传播算法可能不收敛.下面我们给出警示传播算法收敛的充分条件.

定理 2. 如果非负方阵 M 的谱半径 $\rho(M)$ 严格小于 1,则 WP 算法收敛,且与初始信息无关.其中,矩阵 M 中的

元素为 $M_{a \rightarrow i, a \rightarrow i} = \sqrt{2}\delta, M_{a \rightarrow i, b \rightarrow j} = 1_{V(j) \setminus a}(b)1_{V(a) \setminus i}(j)\theta(-J_j^a J_j^b), 1_X(x)$ 是标识函数. 若 $x \in X$, 则 $1_X(x)=1$; 若 $x \notin X$, 则 $1_X(x)=0, \delta$ 为非负常数.

在 WP 算法中, 信息 $u_{a \rightarrow i}$ 取值为 0 或 1. 为了证明定理 2, 我们将 WP 算法的消息取值从 $\{0, 1\}$ 松弛为 $[0, 1]$. 具体地讲: 对于因子图的每条边 (a, i) , 定义消息传递 $\eta_{a \rightarrow i} \in [0, 1]$, 将警示传播算法的消息更新迭代方程(1)更改为

$$\eta_{a \rightarrow i}(t) = (\delta(\eta_{a \rightarrow i}(t-1))^2 - \delta(\eta_{a \rightarrow i}(t-1)) + 1) \prod_{j \in V(a) \setminus i} \theta \left(-J_j^a \left(\sum_{b \in V(j) \setminus a} J_j^b \eta_{b \rightarrow j}(t-1) \right) \right) \quad (5)$$

其中, δ 为非负常数.

相对于公式(1), 在公式(5)中, 我们构造了项 $H(\eta_{a \rightarrow i}(t-1)) = \delta(\eta_{a \rightarrow i}(t-1))^2 - \delta(\eta_{a \rightarrow i}(t-1)) + 1$. 注意到: 当 $\eta_{a \rightarrow i} \in \{0, 1\}$ 时, $H(\eta_{a \rightarrow i}(t-1)) = 1$. 进一步, 如果公式(5)中消息 $\eta_{a \rightarrow i}(t=0)$ 的初始取值仍然为 0 或 1, 则公式(5)退化为公式(1). 我们的想法是: 通过分析公式(5)的收敛性来解释公式(1)的收敛性, 进而给出 WP 算法收敛的有效条件.

设实例的因子图中边的序列集合为 $D = \{a \rightarrow i: (a, i) \in E\}$, 定义向量空间 $V = [0, 1]^{|D|}, (V, \|\cdot\|)$ 是一个赋范空间. 设向量 $x \in V, l_1$ -范数定义为

$$\|x\|_1 = \sum_{i=1}^{|D|} |x_i| \quad (6)$$

l_∞ -范数定义为

$$\|x\|_\infty = \max_{i \in \{1, \dots, |D|\}} |x_i| \quad (7)$$

向量空间 V 上的范数可诱导一个 V 上的度量 $d(v, w) = \|v - w\|, v, w \in V$. 显然, 度量空间 (V, d) 是完备的^[40]. 设函数 $f: V \rightarrow V$, 如果存在 $0 \leq K < 1$, 对于任意的 $v, w \in V$, 使得 $d(f(v) - f(w)) \leq Kd(v, w)$, 则称函数 f 是压缩函数. 如果 $f: V \rightarrow V$ 是完备度量空间 (V, d) 中的压缩函数, 那么对于任意的 $x \in V, f$ 收敛到唯一固定点 x_∞ . 也即, $x, f(x), f^2(x), \dots, x_\infty$.

引理 1. 设函数 $f: V \rightarrow V, d$ 是 V 上的度量. 对于某个 $N \in \mathbb{N}$, 假设 f^N 是关于 d 的压缩函数, 那么 f 收敛到唯一固定点 x_∞ , 且与初始信息无关.

证明: 对于任意的 $x \in V$, 考虑如下 $N-1$ 个序列:

$$\begin{aligned} &x, f^N(x), f^{2N}(x), \dots, \\ &f(x), f^{N+1}(x), f^{2N+1}(x), \dots, \\ &\vdots \\ &f^{N-1}(x), f^{2N-1}(x), f^{3N-1}(x), \dots \end{aligned}$$

因为 f^N 是关于 d 的压缩函数, 所以上述每一个序列都收敛到 x_∞ . 因此, 序列 $x, f(x), f^2(x), \dots$ 一定收敛到 x_∞ . 所以函数 f 收敛到唯一固定点 x_∞ , 且与初始信息无关. □

设映射 $A: V \rightarrow V, (V, \|\cdot\|)$ 是一个赋范空间, 诱导 A 的矩阵范数:

$$\|A\| = \sup_{v \in V} \|Av\| \quad (8)$$

$\|A\|_1$ -范数定义为

$$\|A\|_1 = \max_{j \in \{1, \dots, |D|\}} \sum_{i=1}^{|D|} |A_{ij}| \quad (9)$$

$\|A\|_\infty$ -范数定义为

$$\|A\|_\infty = \max_{i \in \{1, \dots, |D|\}} \sum_{j=1}^{|D|} |A_{ij}| \quad (10)$$

证毕. □

若函数 $f: V \rightarrow V$ 可微, 构造雅可比矩阵 A . 依据中值定理有如下引理.

引理 2^[40]. 设 $(V, \|\cdot\|)$ 是赋范空间, 映射 $f: V \rightarrow V$ 可微, 那么对于任意的 $x, y \in V$, 有公式(11)成立:

$$\|f(x) - f(y)\| \leq \|x - y\| \cdot \sup_{z \in V} \|f'(z)\| \quad (11)$$

根据引理 2,如果 $\sup_{z \in V} \|f'(z)\| < 1$,那么函数 f 收敛到固定点 x_∞ .假设算法中的消息更新并行执行,对于任意的向量 $\eta \in V$,函数 f 的分量 $f(\eta)_{a \rightarrow i}$ 定义为 $f(\eta)_{a \rightarrow i} = \eta_{a \rightarrow i}$.自然,消息更新方程(5)的并行执行过程可以用函数 f 的迭代过程来刻画.下面从引理 2 出发导出 WP 算法收敛的一个充分条件.

为了求解需要,我们对公式(5)进行参数化.通过观察,取 $\sinh \lambda_{a \rightarrow i} = \eta_{a \rightarrow i}$,则公式(5)变为如下形式:

$$\sinh \lambda_{a \rightarrow i}(t) = (\delta \sinh^2 \lambda_{a \rightarrow i}(t-1) - \delta(\sinh \lambda_{a \rightarrow i}(t-1) + 1) \prod_{j \in V(a) \setminus i} \theta \left(-J_j^a \sum_{b \in V(j) \setminus a} J_j^b \sinh \lambda_{b \rightarrow j}(t-1) \right)) \quad (12)$$

由于 $\eta_{a \rightarrow i} \in [0, 1]$,所以 $\lambda_{a \rightarrow i} \in [0, \log(1 + \sqrt{2})]$.定义向量空间 $V' = [0, \log(1 + \sqrt{2})]^{|D|}$,设函数 $h: V' \rightarrow V'$,对于任意的向量 $\lambda \in V'$,函数 h 的分量 $h(\lambda)_{a \rightarrow i}$ 定义为 $h(\lambda)_{a \rightarrow i} = \lambda_{a \rightarrow i}$.设 $i, j \in V(a), i \neq j, b \in V(j) \setminus a$.令:

$$G(\lambda_{a \rightarrow i}(t-1)) = \delta \sinh^2 \lambda_{a \rightarrow i}(t-1) - \delta(\sinh \lambda_{a \rightarrow i} F(\lambda_{c \rightarrow k}(t-1))) = \prod_{k \in V(a) \setminus i, j} \theta \left(-J_k^a \sum_{c \in V(k) \setminus a} J_k^c \sinh \lambda_{c \rightarrow k}(t-1) \right) (t-1) + 1.$$

对函数 h 求导,根据公式(12),我们有:

$$(h'(\lambda))_{b \rightarrow j}^{a \rightarrow i} = \frac{\partial \lambda_{a \rightarrow i}(t)}{\partial \lambda_{b \rightarrow j}(t-1)} = \frac{G(\lambda_{a \rightarrow i}(t-1)) \cdot F(\lambda_{c \rightarrow k}(t-1))}{\sqrt{1 + \sinh^2 \lambda_{a \rightarrow i}(t)}} \theta(-J_j^a J_j^b \cosh \lambda_{b \rightarrow j}(t-1)) \quad (13)$$

注意,上述是关于变量 $\lambda_{b \rightarrow j}$ 的求导,对于所有的子句 $c \in V(j) \setminus \{a, b\}, \lambda_{c \rightarrow j}$ 在公式(12)中均被看作常量.同样,对于所有的 $k \in V(a) \setminus \{i, j\}, d \in V(k) \setminus a, \lambda_{d \rightarrow k}$ 也被看做常量处理.令 $y(\lambda_{b \rightarrow j}) = -J_j^a \sum_{b \in V(j) \setminus a} J_j^b \sinh \lambda_{b \rightarrow j}$,考虑到截尾函数 $\theta(\cdot)$ 在公式(12)中的实际作用,先通过对函数 $y(\lambda_{b \rightarrow j})$ 进行求导,再进行截尾函数 $\theta(\cdot)$ 处理.因此,公式(12)通过求导后,截尾函数为 $\theta(-J_j^a J_j^b \cosh \lambda_{b \rightarrow j}(t-1))$.

又因为 $\cosh \lambda_{b \rightarrow j}(t-1) > 0$,所以 $\theta(-J_j^a J_j^b \cosh \lambda_{b \rightarrow j}(t-1)) = \theta(-J_j^a J_j^b)$.注意到:

$$\sup_{\lambda \in [0, \log(1 + \sqrt{2})]^{|D|}} \frac{G(\lambda_{a \rightarrow i}(t-1)) \cdot F(\lambda_{c \rightarrow k}(t-1))}{\sqrt{1 + \sinh^2 \lambda_{a \rightarrow i}(t)}} = 1 \quad (14)$$

因此,

$$|(h'(\lambda))_{b \rightarrow j}^{a \rightarrow i}| = \left| \frac{\partial \lambda_{a \rightarrow i}(t)}{\partial \lambda_{b \rightarrow j}(t-1)} \right| \leq \theta(-J_j^a J_j^b) \quad (15)$$

而

$$\begin{aligned} h'(\lambda)_{a \rightarrow i}^{a \rightarrow i} &= \frac{\partial \lambda_{a \rightarrow i}(t)}{\partial \lambda_{a \rightarrow i}(t-1)} \\ &= \frac{\delta \cosh \lambda_{a \rightarrow i}(t-1) (2 \sinh \lambda_{a \rightarrow i}(t-1) - 1) \prod_{j \in V(a) \setminus i} \theta \left(-J_j^a \sum_{b \in V(j) \setminus a} J_j^b \sinh \lambda_{b \rightarrow j}(t-1) \right)}{\sqrt{1 + \sinh^2 \lambda_{a \rightarrow i}(t)}} \\ &\leq \delta \cosh \lambda_{a \rightarrow i}(t-1) (2 \sinh \lambda_{a \rightarrow i}(t-1) - 1) \end{aligned} \quad (16)$$

由于 $\sinh \lambda_{a \rightarrow i} = \eta_{a \rightarrow i} \in [0, 1]$,因此 $h(\lambda)_{a \rightarrow i}^{a \rightarrow i} = \frac{\partial \lambda_{a \rightarrow i}(t)}{\partial \lambda_{a \rightarrow i}(t-1)} \leq \delta \cosh \lambda_{a \rightarrow i}(t-1)$.

又因为 $\cosh \lambda_{a \rightarrow i}(t-1) = \sqrt{1 + \sinh^2 \lambda_{a \rightarrow i}(t-1)} \leq \sqrt{2}$,

所以,

$$|(h'(\lambda))_{a \rightarrow i}^{a \rightarrow i}| = \left| \frac{\partial \lambda_{a \rightarrow i}(t)}{\partial \lambda_{a \rightarrow i}(t-1)} \right| \leq \sqrt{2} \delta \quad (17)$$

根据代数知识,利用函数 h 的导数构造雅可比矩阵 A .由压缩函数的性质及引理 2 可知:如果 A 的范数严格小于 1,则 h 为压缩函数,进而收敛到唯一固定点,且与初始信息无关.有如下结论:

定理 3. 对于 $\|A\|_1$ -范数,如果

$$\max_b \max_{j \in V(b)} \sum_{a \in V(j), b} \theta(-J_j^a J_j^b) < 1 - \sqrt{2}\delta \tag{18}$$

则公式(5)在 $\|A\|_1$ -范数下收敛,且与初始信息无关.

证明:根据公式(9)、公式(15)、公式(17),有:

$$\|A\|_1 = \|h'(\lambda)\|_1 = \max_{b \rightarrow j} \sum_{a \rightarrow i} \left| \frac{\partial \lambda_{a \rightarrow i}}{\partial \lambda_{b \rightarrow j}} \right| \leq \sqrt{2}\delta + \max_b \max_{j \in V(b)} \sum_{a \in V(j), b} \theta(-J_j^a J_j^b) \tag{19}$$

由压缩函数的性质和引理 2 可知:如果公式(20)成立,则函数 h 收敛到唯一固定点,且与初始信息无关:

$$\max_b \max_{j \in V(b)} \sum_{a \in V(j), b} \theta(-J_j^a J_j^b) < 1 - \sqrt{2}\delta \tag{20}$$

根据代数知识,函数 f 收敛到某一固定点,且与初始信息无关.所以,公式(5)在 $\|A\|_1$ -范数下收敛,且与初始信息无关. \square

同理可证如下结论.

定理 4. 对于 $\|A\|_\infty$ -范数,如果:

$$\max_a \max_{j \in V(a)} \sum_{b \in V(j), a} \theta(-J_j^a J_j^b) < 1 - \sqrt{2}\delta \tag{21}$$

则公式(5)在 $\|A\|_\infty$ -范数下收敛,且与初始信息无关.

证明:根据公式(10)、公式(15)、公式(17),有:

$$\|A\|_\infty = \|h'(\lambda)\|_\infty = \max_{a \rightarrow i} \sum_{b \rightarrow j} \left| \frac{\partial \lambda_{a \rightarrow i}}{\partial \lambda_{b \rightarrow j}} \right| \leq \sqrt{2}\delta + \max_a \max_{j \in V(a)} \sum_{b \in V(j), a} \theta(-J_j^a J_j^b) \tag{22}$$

由压缩函数的性质和引理 2 可知:如果公式(23)成立,则函数 h 收敛到唯一固定点,且与初始信息无关:

$$\max_a \max_{j \in V(a)} \sum_{b \in V(j), a} \theta(-J_j^a J_j^b) < 1 - \sqrt{2}\delta \tag{23}$$

根据代数知识,函数 f 收敛到某一固定点,且与初始信息无关.所以,则公式(5)在 $\|A\|_\infty$ -范数下收敛,且与初始信息无关. \square

设 M 是一个方阵, $\gamma_1, \dots, \gamma_n$ 是 M 的特征值,称 $\gamma(M) = \{\gamma_1, \dots, \gamma_n\}$ 为方阵 M 的谱, $\rho(M) = \sup\{|\gamma_i| : \gamma_i \in \gamma(M)\}$ 为方阵 M 的谱半径.

定理 5. 设函数 $h: V' \rightarrow V'$ 可微,进一步假设 $|h'(x)| \leq M$, M 是不依赖于 x 的非负矩阵.如果 $\rho(M) < 1$,那么对于任意 $x \in V'$, h 收敛到唯一固定点 x_∞ .也即, $x, h(x), h^2(x), \dots, x_\infty$ (注意: $|h'(x)| \leq M$ 表示由 $h'(x)$ 构成的雅可比矩阵中的每个元素的绝对值不大于 M 中对应元素的值).

证明:对于任意的 $x \in V'$, 假设 $n=1, 2, \dots$, 我们有:

$$\|(h^n)'(x)\| = \left\| \prod_{i=1}^n h'(h^{i-1}(x)) \right\| = \left\| \prod_{i=1}^n h'(h^{i-1}(x)) \right\| \leq \left\| \prod_{i=1}^n |h'(h^{i-1}(x))| \right\| \leq \left\| \prod_{i=1}^n M \right\| = \|M^n\| \tag{24}$$

由 Gelfand 定理^[40]: $\lim_{n \rightarrow \infty} \|M^n\|^{1/n} = \rho(M)$. 我们选取 $\varepsilon > 0$, 使得 $\rho(M) + \varepsilon < 1$. 对于某个 $N \in \mathbb{N}$, $\|M^N\| = (\rho(M) + \varepsilon)^N < 1$.

因此, $\|(h^N)'(x)\| < 1$. 所以, $\|(h^N)(x)\|$ 是压缩函数.

由引理 1 可知:函数 $h(x)$ 收敛到唯一固定点 x_∞ , 且与初始信息无关. \square

对于任意的 $i, j \in V(a), i \neq j, b \in V(j), a$, 定义 $|D| \times |D|$ 的非负方阵 M . M 中的元素:

$$M_{a \rightarrow i, a \rightarrow i} = \sqrt{2}\delta, M_{a \rightarrow i, b \rightarrow j} = 1_{V(j), a}(b) 1_{V(a), i}(j) \theta(-J_j^a J_j^b).$$

其中, $1_X(x)$ 是标识函数,若 $x \in X$, 则 $1_X(x) = 1$; 若 $x \notin X$, 则 $1_X(x) = 0$, δ 为非负常数. 根据公式(15)、公式(17)和定理 5, 我们有如下结论:

定理 6. 如果方阵 M 的谱半径 $\rho(M)$ 严格小于 1, 则迭代方程(5)式收敛, 且与初始信息无关.

证明:根据公式(15)、公式(17)和定理 5, 函数 h 收敛到唯一固定点, 且与初始信息无关. 进一步地, 函数 f 收敛到唯一固定点, 且与初始信息无关. 所以, 方程(5)收敛, 且与初始信息无关. \square

假设 M 的特征值 γ , 根据代数知识, 我们有: $|\gamma| \cdot \|x\| = \|\gamma x\| = \|Mx\| \leq \|M\| \cdot \|x\|$. 因此, 定理 6 的条件弱于定理 3 和定理 4.

设向量空间 $V^m = \{0, 1\}^D \subset V$, 定义函数 $g: V^m \rightarrow V^m$. 假设 WP 算法中的信息更新是并行执行, 对于任意的向量 $u \in V^m$, 函数 g 的分量 $g(u)_{a \rightarrow i}$ 定义为 $g(u)_{a \rightarrow i} = u_{a \rightarrow i}$. 自然, WP 算法的执行过程可用函数 g 的迭代过程来刻画. 根据函数 f 和 g 的构造, 显然, 对于任意的 $x \in V^m$, 有 $f(x) = g(x)$.

引理 3. 如果函数 f 是一个压缩映射, 那么函数 g 也为压缩映射.

证明: 由于 f 是一个压缩映射, 对于任意的 $x \in V$, 存在唯一固定点 x_∞ , 有 $x, f(x), f^2(x), \dots, x_\infty$, 也即: 函数 f 收敛到 x_∞ , 且与初始信息无关. 因此, 对于任意的 $y \in V^m$, 仍然有 $y, f(y), f^2(y), \dots, x_\infty$. 而当 $y \in V^m$ 时, 有 $f(y) = g(y)$. 进而有 $y, g(y), g^2(y), \dots, x_\infty$. 因此, 函数 g 为压缩映射, 收敛到固定点 x_∞ , 且与初始信息无关. \square

根据定理 6 和引理 3, 至此, 定理 2 得证.

注意到: 定理 2 中非负矩阵的元素由公式的结构决定. 因此, 参数 δ 的值取决于公式的结构. 显然, 对于定理 1 中的结论, 容易使用定理 2 来证明. 为了方便理解, 下面举例说明定理 2 的使用.

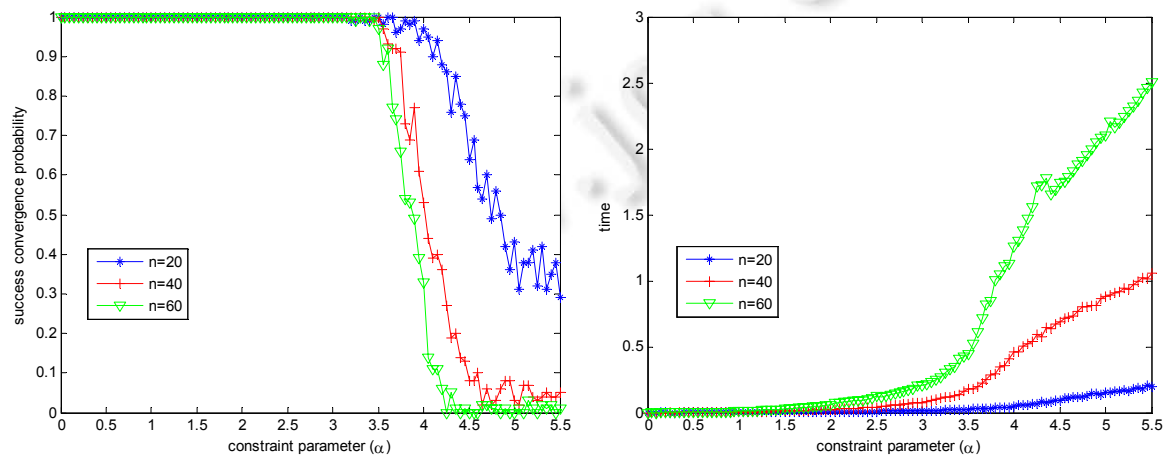
例 1: 设 CNF 公式 $F_1 = \{(x_1 \vee x_2), (x_2 \vee x_3), \dots, (x_{n-1} \vee x_n), (x_n \vee x_1)\}$, 其因子图含有一个环. 因此, 矩阵 M 的谱半径 $\rho(M) = \sqrt{2}\delta$. 若取 $0 < \delta < \frac{1}{\sqrt{2}}$, 则 $\rho(M) < 1$. 根据定理 2, 则 WP 算法在公式 F_1 上收敛.

同样, WP 算法在公式 $F_2 = \{(-x_1 \vee -x_2), (-x_2 \vee -x_3), \dots, (-x_{n-1} \vee -x_n), (-x_n \vee -x_1)\}$ 上也收敛.

3 数值实验及分析

在数值实验中, 使用模型 $G(n, 3, m)$ 产生随机 3-SAT 实例, 其中, n 表示变元个数, m 表示子句个数, 子句大小为 3. 模型 $G(n, 3, m)$ 按如下方式产生随机 3-SAT 实例: 随机均匀地在可能的子句 (这样的子句共有 $2^3 \cdot \binom{n}{3}$ 个) 中选择 $m = \alpha n$ 个子句, 被选中的子句之集构成实例. 在算法中, 取最大迭代次数 $t_{\max} = 10^3$.

我们选取 $G(n, 3, m)$ 随机实例进行模拟实验. 图 1 是 WP 算法分别在变元规模 $n=20, n=40, n=60$ 上的运行情况. 图 1(a) 是算法的收敛性随着约束参数 α 的变化情况, 图 1(b) 是算法的计算时间随着约束参数 α 的变化情况. 图中每个数据点由 $G(n, 3, m)$ 模型产生的 100 个随机实例构成. 可以看出: 当参数 $\alpha < 3.5$ 时, WP 算法高概率收敛; 当 $\alpha > 4.5$ 时, 算法高概率不收敛. 在 $\alpha \approx 3.5$ 点, WP 算法的收敛性发生了突变. 同样, 当参数 $\alpha < 3.5$ 时, WP 算法的运行时间较小; 当 $\alpha > 3.5$ 时, WP 算法的运行时间突然增大.



(a) 算法的收敛性 (b) 算法的计算时间

Fig.1 Change of convergence probability and time of WP with parameter α

图 1 WP 算法的收敛性和计算时间随约束参数 α 的变化情形

对于随机 3-SAT 实例, α 是重要的参数, 它影响实例的可满足性和判定难度. 存在可满足性的相变点 a_d , 当 $\alpha < a_d$ 时, 实例高概率可满足; 当 $\alpha > a_d$ 时, 实例高概率不可满足. 同时, 存在第 2 相变点 a_c , $a < a_c$ 为易解 SAT 区域, $a > a_c$ 为难解 SAT 区域. 大量的材料表明: a_d 至少为 3.52, 至多为 4.506, $a_c \approx 3.921$. 注意到, WP 算法在 $3.52 < \alpha < 4.506$ 区域的收敛情况不能从概率上给出确切的回答. 因此, 警示传播算法能够有效判定的随机 3-SAT 实例最多为 $\alpha < 3.52$ 区域. 从图 1(a) 中可以看出: 随着问题规模 n 的增大, WP 算法在 $G(n, 3, m)$ 实例上收敛的概率趋向于一个稳定值, 即, 收敛的概率曲线随着 n 的增大几乎趋于不变. 我们知道, 因子图的结构直接影响 WP 算法的收敛性. 事实上, 当参数 $\alpha > a_c$ 时, $G(n, 3, m)$ 模型产生的随机实例的因子图结构比较复杂, 求解难度较大. 从图 1(b) 中我们也可以看出: 大约在 $\alpha > 3.5$ 处, 算法的运行时间急剧上升.

为了验证定理 2 中的判定条件, 我们分别取了 $n=20$ 和 $n=60$ 的 $G(n, 3, m)$ 实例数据集进行模拟实验. 从图 2 中可以看出: 当 $\alpha < 1.8$ 时, 定理 2 能够判定 WP 算法的收敛性. 事实上, 文献[39]已经表明: 当约束参数 α 较小时, $G(n, 3, m)$ 实例的因子图的结构内部呈现树形状. 因此, 根据定理 2, 容易判定 WP 算法收敛. 随着 α 的增大, 因子图结构中开始出现圈, 并且圈的数量随着 α 的进一步增大而急剧增加. 由于定理 2 中的非负矩阵的元素构成与因子图的结构有关, 当因子图的结构变得复杂时, 非负矩阵的谱半径不再小于 1. 这时, 定理 2 中的判定条件失效. 如图 2 所示: 当 $\alpha > 1.8$ 时, 定理 2 中的判定条件逐渐失效.

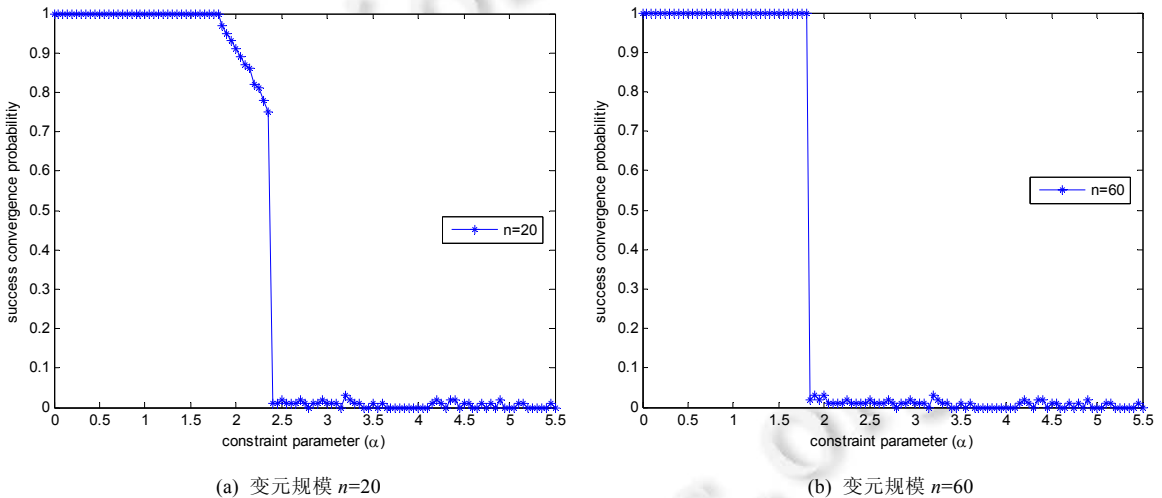


Fig.2 Results of Th.2

图 2 定理 2 的实验结果

如果一个因子图包含环, 且只含正文字或只含负文字的实例 F , 根据定理 2 容易判断, 则 WP 算法在 F 上收敛. 我们分别选取 3 组不同变元规模的由 $G(n, 3, m)$ 模型产生的随机 3SAT 可满足实例进行实验模拟, 这里取 $\alpha=2.0$. 这些实例只含正文字或只含负文字, 每组数据有 200 个实例(其中, 有 100 个只包含正文字的实例和 100 个只包含负文字的实例). 实验结果表明: 在这种随机实例上, WP 算法收敛. 表 1 是 WP 算法的收敛情况统计表. 用符号“+”表示只含正文字的实例; 用符号“-”表示只含负文字的实例. Var_num 表示实例的变元规模, Instance 表示实例集, Convergence 表示 WP 算法收敛的概率.

Table 1 The convergence of WP (Theory 2)

表 1 WP 算法的收敛情况(定理 2)

| Var_num | Instance | Convergence |
|---------|----------|-------------|
| 20 | + - | 1 |
| 40 | + - | 1 |
| 60 | + - | 1 |

4 结束语

本文分析了 WP 算法的收敛性,通过引入参数,将 WP 算法的信息取值从 $\{0,1\}$ 松弛为 $[0,1]$,利用向量空间中压缩函数的性质,给出了 WP 算法收敛的一个充分条件.大量的实验结果表明,该方法有效.同时,该方法提供了理解 WP 算法性能的理论依据,并为分析其他信息传播算法的收敛性奠定了理论基础.进一步工作是:给出 WP 算法更弱的收敛条件.

References:

- [1] Martin D, Alan F, Michael M. A probabilistic analysis of randomly generated binary constraint satisfaction. *Theory Computer Science*, 2003,290:1815–1828. [doi: 10.1016/S0304-3975(02)00317-1]
- [2] Creignou N, Eaude H. The SAT-UNSAT transition for random constraint satisfaction problems. *Discrete Mathematics*, 2009,309:2085–2099. [doi: 10.1016/j.disc.2008.04.025]
- [3] Martin OC, Monasson R, Zecchina R. Statistical mechanics methods and phase transitions in optimization. *Theory Computer Science*, 2001,265:3–67. [doi: 10.1016/S0304-3975(01)00149-9]
- [4] Kirkpatrick S, Selman B. Critical behavior in the satisfiability of random Boolean formulae. *Science*, 1994,264(5163):1297–1301. [doi: 10.1126/science.264.5163.1297]
- [5] Mertens S, Mezard M, Zecchina R. Threshold values of random k -SAT from the cavity method. *Random Structure Algorithms*, 2006,28:340–373. [doi: 10.1002/rsa.v28:3]
- [6] Kaporis A, Kirousis L, Lalas E. The probabilistic analysis of a greedy satisfiability algorithm. *Random Structures & Algorithms*, 2006,28(4):444–480. [doi: 10.1002/rsa.v28:4]
- [7] Dubois O, Boufkhad Y, Mandler J. Typical random 3-sat formulae and the satisfiability threshold. *Electronic Colloquium on Computational Complexity*, 2003,10(007):1–2.
- [8] Moskewicz MW, Madigan CF, Zhao Y. Chaff: Engineering an efficient SAT solver. In: *Proc. of the 38th Annual Design Automation Conf. New York: ACM Press*, 2001. 530–535. [doi: 10.1145/378239.379017]
- [9] Aurell E, Gordon U, Kirkpatrick S. Comparing beliefs, surveys, and random walks. *Advances in Neural Information Processing Systems*, 2004,17(1):1–8.
- [10] Mezard M, Parisi G. The cavity method at zero temperature. *Journal of Statistical Physics*, 2003,111(1-2):1–34.
- [11] Mezard M, Zecchina R. Random k -satisfiability problem: From an analytic solution to an efficient algorithm. *Physical Review E*, 2002,66(5):056126.
- [12] Maneva E, Mossel E, Wainwright M. A new look at survey propagation and its generalizations. *Journal of the ACM*, 2007,54(4):1089–1098.
- [13] Braunstein A, Mezard M, Zecchina R. Survey propagation: An algorithm for satisfiability. *Random Structures and Algorithms*, 2005,27(2):201–226.
- [14] Yedidia JS, Freeman WT, Weiss Y. Understanding belief propagation and its generalizations. *Artificial Intelligence*, 2003,8(1):239–269.
- [15] Braunstein A, Zecchina R. Survey and belief propagation on random k -SAT. *Lecture Notes in Computer Science*, 2004,2919(1):519–528.
- [16] Xu K, Li W. Exact phase transitions in random constraint satisfaction problems. *Journal of Artificial Intelligence Research*, 2000,12(1):93–103.
- [17] Xu K, Li W. Many hard examples in exact phase transitions. *Theory Computer Science*, 2006,355(1):291–302. [doi: 10.1016/j.tcs.2006.01.001]
- [18] Xu K, Boussemart F, Hemery F, Lecoutre C. Random constraint satisfaction: Easy generation of hard instances. *Artificial Intelligence*, 2007,171(1):514–534. [doi: 10.1016/j.artint.2007.04.001]
- [19] 赵春艳,郑志明.一种基于变量熵求解约束满足问题的置信传播算法. *中国科学(F:信息科学)*,2012,42(9):1170–1180. [doi: 10.1360/112011-693]
- [20] Yin MH, Zhou JP, Sun JG, Gu W. Heuristic survey propagation algorithm for solving QBF problem. *Ruan Jian Xue Bao/Journal of Software*, 2011,22(7):1538–1550 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/3859.htm> [doi: 10.3724/SP.J.1001.2011.03859]
- [21] Ravanbakhsh S, Greiner R. Perturbed message passing for constraint satisfaction problems. *Artificial Intelligence*, 2014. arXiv:1401.6686
- [22] Chieu HL, Lee WS. Relaxed survey propagation for the weighted maximum satisfiability problem. *Journal of Artificial Intelligence Research*, 2009,36:229–266.

- [23] Bayati M, Shah D, Sharma M. Max product for maximum weight matching: Convergence, correctness, and LP duality. *IEEE Trans. on Information Theory*, 2007,54(3):1241–1251.
- [24] Coja-Oghlan A, Mossel E, Vilenchik D. A spectral approach to analyzing belief propagation for 3-colouring. *Combinatorics, Probability and Computing*, 2009,18(6):881–912.
- [25] Gamarnik D, Shah D, Wei Y. Belief propagation for min-cost network flow: Convergence and correctness. *Operations Research*, 2012,60(2):410–428.
- [26] Sanghavi S, Malioutov DM, Willsky AS. Belief propagation and LP relaxation for weighted matching in general graphs. *IEEE Trans. on Information Theory*, 2011,57(4):2203–2212.
- [27] Sanghavi S, Shah D, Willsky AS. Message passing for maximum weight independent set. *IEEE Trans. on Information Theory*, 2009, 55(11):4822–4834.
- [28] Kschischang F, Frey B, Loeliger H. Factor graph and the sum-product algorithm. *Information Theory*, 2001,47(1):498–519.
- [29] Weiss Y. Correctness of local probability propagation in graphical models with loops. *Neural Computation*, 2000,12(1):1–41.
- [30] Weiss Y, Freeman WT. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. *Neural Computation*, 2001,13(10):2173–2200. [doi: 10.1162/089976601750541769]
- [31] Tatikonda S, Jordan MI. Loopy belief propagation and Gibbs measure. In: *Proc. of the 18th Annual Conf. on Uncertainty in Artificial Intelligence*. 2002. 493–500.
- [32] Heskes T. On the uniqueness of loopy belief propagation fixed points. *Neural Computation*, 2004,16:2379–2413. [doi: 10.1162/0899766041941943]
- [33] Ihler AT. Loopy belief propagation: Convergence and effects of message errors. *Machine Learning Research*, 2005,6:905–936.
- [34] Mooij JM, Kappen HJ. Sufficient conditions for convergence of the sum-product algorithm. *IEEE Trans. on Information Theory*, 2007,53:4422–4437.
- [35] Shi XQ, Schonfeld D, Tuninetti D. Message error analysis of loopy belief propagation for the sum-product algorithm. *Computer and Information Science*, 2010,1009:1–30.
- [36] Brunsch T, Cornelissen K, Manthey B, Roglin H. Smoothed analysis of BP for minimum cost flow and matching. *Journal of Graph Algorithms and Applications*, 2013,17(6):647–670.
- [37] Norshams N, Wainwright MJ. Stochastic belief propagation: A low complexity alternative to the sum product algorithm. *IEEE Trans. on Information Theory*, 2013,59(4):1981–2000.
- [38] Feige U, Mossel E, Vilenchik D. Complete convergence of message passing algorithms for some satisfiability problems. *Theory of Computing*, 2013,9(19):617–651. [doi: 10.4086/toc.2013.v009a019]
- [39] Wang XF, Xu DY, Wei L. Convergence of warning propagation algorithms for random satisfiable instances. *Ruan Jian Xue Bao/ Journal of Software*, 2013,24(1):1–11 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/4213.htm> [doi: 10.3724/SP.J.1001.2013.04213]
- [40] Dieudonné J. *Foundations of Modern Analysis*. New York: Academic Press, 1969.

附中中文参考文献:

- [20] 殷明浩,周俊萍,孙吉贵,谷文祥.求解 QBF 问题的启发式调查传播算法. *软件学报*,2011,22(7):1538–1550. <http://www.jos.org.cn/1000-9825/3859.htm> [doi: 10.3724/SP.J.1001.2011.03859]
- [39] 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. *软件学报*,2013,24(1):1–11. <http://www.jos.org.cn/1000-9825/4213.htm> [doi: 10.3724/SP.J.1001.2013.04213]



王晓峰(1980—),男,甘肃会宁人,博士,讲师,CCF 会员,主要研究领域为算法设计与分析,计算复杂性,人工智能.



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.