

RB 模型实例集上置信传播算法的收敛性*

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: 置信传播算法求解 $RB(k, n, \alpha, r_c, p)$ 模型实例时非常有效, 几乎能够有效求解接近可满足性相变点的难解实例. 然而, 因子图带有回路的实例, 置信传播算法不总是有效, 常表现为不收敛. 对于这种现象, 至今缺少系统的理论解释. 置信传播算法是最为基础的信息传播算法, 对置信传播算法的收敛性分析是其他信息传播算法收敛性分析的重要基础. 在 $RB(k, n, \alpha, r_c, p)$ 模型中, 取 $k=2$, $\alpha > \frac{1}{k}$, $r_c > 0$ 均为常数, 且满足 $kr_c^{\frac{\alpha}{1-\alpha}} \geq 1$. 证明了如果 $p \in (0, n^{-2\alpha})$, 则置信传播算法在 $RB(k, n, \alpha, r_c, p)$ 模型产生的随机实例集上高概率收敛. 最后, 在 $RB(k, n, \alpha, r_c, p)$ 模型上选取了几组不同的数据进行数值模拟, 实验结果表明该结论有效. 当问题规模 n 增大时, 在 $RB(k, n, \alpha, r_c, p)$ 模型的可满足区域, 实验收敛区间趋于一个固定范围, 而理论收敛区间逐渐变窄. 原因在于, $RB(k, n, \alpha, r_c, p)$ 模型是一个具有增长定义域的随机 CSP 实例产生模型, 不协调赋值的数目与参数 p 及问题规模 n 有关.

关键词: 置信传播算法; 收敛性; 约束可满足性问题; RB 模型

中图法分类号: TP181

中文引用格式: 王晓峰, 许道云. RB 模型实例集上置信传播算法的收敛性. 软件学报, 2016, 27(11): 2712-2724. <http://www.jos.org.cn/1000-9825/4877.htm>

英文引用格式: Wang XF, Xu DY. Convergence of the belief propagation algorithm for RB model instances. Ruan Jian Xue Bao/ Journal of Software, 2016, 27(11): 2712-2724 (in Chinese). <http://www.jos.org.cn/1000-9825/4877.htm>

Convergence of the Belief Propagation Algorithm for RB Model Instances

WANG Xiao-Feng^{1,2}, XU Dao-Yun²

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

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

Abstract: Belief propagation algorithm is very effective in finding satisfying assignments for $RB(k, n, \alpha, r_c, p)$ model instances where hard region becomes narrower. However, belief propagation algorithm does not always converge for factor graphs with cycles. Unfortunately, rigorous theoretical proof of this phenomenon is still lacking. Belief propagation algorithm is the most basic message passing algorithms. This study derives the convergence conditions of the belief propagation algorithm for solving $RB(k, n, \alpha, r_c, p)$ model instances. In the $RB(k, n, \alpha, r_c, p)$ model with $k=2$, $\alpha > \frac{1}{k}$, $r_c > 0$ and $kr_c^{\frac{\alpha}{1-\alpha}} \geq 1$, which proves that BP will be converged with high probability if $p \in (0, n^{-2\alpha})$. Experimental results show that such convergence conditions of belief propagation algorithm are very effective in two different group data from the random $RB(k, n, \alpha, r_c, p)$ model instances. In the $RB(k, n, \alpha, r_c, p)$ model, when n increases, the experimental convergence interval is fixed range, while the theory convergence interval become narrower. It is because the number of incompatible assignment are determined by n and p in the $RB(k, n, \alpha, r_c, p)$.

* 基金项目: 国家自然科学基金(61462001, 61262006)

Foundation item: National Natural Science Foundation of China (61462001, 61262006)

收稿时间: 2014-03-09; 修改时间: 2015-03-10; 采用时间: 2015-08-07; jos 在线出版时间: 2016-03-30

CNKI 网络优先出版: 2016-03-30 11:43:38, <http://www.cnki.net/kcms/detail/11.2560.TP.20160330.1143.001.html>

Key words: belief propagation algorithm; convergence; constraint satisfaction problem; RB model

约束可满足问题(constraint satisfaction problem,简称 CSP)是人工智能中的一个重要研究领域^[1-3].一个 CSP 问题是由 n 个变量和 m 个约束组成.每个变量从其定义域 D 中取值.从变量集中随机选取 k 个不同的变量构成一个约束,并且有一个相应的不协调集合 Q 来限制这 k 个变量的取值为不可满足的赋值^[4].从可能的 $\binom{n}{k}$ 个约束中随机选取 m 个构成 CSP 的一个随机实例.求解这个随机实例就是至少找到一个解,也就是这 n 个变量的一组赋值,使得这 m 个约束同时被满足;或者证明无解.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.一般情况下,随机 CSP 是 NP 完全的.在随机 CSP 问题中,约束个数 m 与变量个数 n 之比 $a=m/n$ 是一个重要参数,它不仅对实例的可满足性产生影响,还对求解实例的难易程度产生影响.

- 一方面,大量的理论分析和实验结果表明,NP 完全问题存在可满足性相变现象,即存在一个阈值 a_d ,几乎所有的随机实例在 $a < a_d$ 时是高概率可满足的(satisfiable,简称 SAT);相反地,几乎所有的随机实例在 $a > a_d$ 时是高概率不可满足的(unsatisfiable,简称 UNSAT)^[5-8].称阈值 a_d 为相变点.
- 另一方面,大多数优秀的算法能够求解的实例与 a 密切相关.在相变点 a_d 附近,实例的求解难度最大.如比较典型的随机 3SAT 问题,基于纯文字规则的算法只能有效求解 $a < 1.6$ 区域的随机 3SAT 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 3SAT 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 3SAT 实例^[9,10].

最近几年,统计物理中的方法得到了广泛应用,基于该方法设计的求解可满足性问题的信息传播算法,在接近相变点时仍然有非常出色的表现.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 3SAT 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 3SAT 实例^[11-17],几乎能够有效求解接近相变点的实例.这弥补了以往其他随机算法在接近相变点时就会失效的缺陷.

信息传播算法不仅对固定定义域的随机 CSP 实例非常有效(如随机 k SAT),还可用于求解具有增长定义域的随机 CSP 实例.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 $RB(k,n,\alpha,r_c,p)$ 模型,该模型是一个典型的具有增长定义域的随机 CSP 实例产生模型,被广泛应用于构造难解性问题^[18].随后,文献[19-21]分别从理论和实验上说明在相变点的阈值附近,求解 $RB(k,n,\alpha,r_c,p)$ 模型实例的难度随着问题规模的增加以指数形式增长.文献[22]表明,基于变量熵的 BP 算法在求解 $RB(k,n,\alpha,r_c,p)$ 模型实例时仍然非常有效,几乎能够求解接近相变点的实例.

尽管信息传播算法求解 CSP 实例时非常有效,但仍有两个问题有待理论上深入研究:

- 1) 信息传播算法是否收敛.
- 2) 如果信息传播算法收敛,得到的结果是否是变量边缘概率分布的有效近似;如果信息传播算法不收敛,这种情况下算法失效.

BP 算法之所以能够有效求解 CSP 实例,缘于该算法能够在 CSP 实例上正确收敛,进而得到变量的有效边缘概率分布,用以固定部分变量的赋值.然而在某些实例上,信息传播算法不是总有效,常表现为不收敛.我们已经知道:对于因子图为树结构和带有一个环的图结构的 CSP 实例,BP 算法能够有效收敛^[23,24].但对于因子图带有多个环的 CSP 实例,BP 算法不是总收敛.最近几年,一些学者初步研究了 BP 算法的收敛性,取得了一些结果.例如,文献[25]表明,对于任意结构的高斯图模型实例,BP 算法能够正确收敛.Tatikonda 等人分析了基于计算树的 Gibbs 测度序列的唯一性与 BP 算法的收敛性之间的关系,给出了 BP 算法收敛的有效条件^[26].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^[27].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,并利用该界给出了 BP 算法收敛的有效条件^[28].Mooij 等人利用商空间中信息更新函数的压缩映射原理导出了 BP 算法收敛的有效条件,指出,如果更新函数的系数矩阵的谱半径严格小于 1,则 BP 算法收敛^[29].Shi 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 BP 算法的动态行为,

并给出了算法收敛的有效条件^[30].文献[31]中分析了求解最小代价网络流(minimum cost network flow,简称MCF)的BP算法的收敛性,研究表明,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,文献中还提供了一个简化了的BP算法,能够对MCF问题给出一个多项式时间随机近似方案.文献[32]中给出了BP算法收敛的一个概率条件,研究表明,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地有多项式界.文献[33]中分析了BP算法的复杂性,提出一个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.

上述有关置信传播算法的收敛性证明方法大多都忽略了产生问题的模型,而是把注意力集中在具体问题本身,增加了这一问题的研究难度.事实上,基于实例的产生模型来研究置信传播算法的收敛性是非常有意义的.研究这一问题的优点至少包含以下3点:

- (1) 为了研究一般性复杂问题,往往需要构造该问题的实例产生模型,而该模型一般为一个随机模型(如,CSP问题的实例产生模型B,D,RB,随机3-SAT实例产生模型等),利用随机模型的特征来分析问题所对应的因子图的某些结构性质的概率界.我们知道,置信传播算法是基于实例因子图的信息迭代算法,因子图的结构影响算法的收敛性.因此,利用因子图中某些结构性质的概率界来估算置信传播算法迭代步数的上界,以提高算法运行的效率,这是非常有意义的.
- (2) 实例产生模型一般具有规则结构,规则图中的许多良好的性质对于研究实例因子图的结构性质具有潜在的作用.
- (3) 基于这种规则结构的实例,置信传播算法的收敛性研究则更为具体和直观,有助于分析置信传播算法中的消息传递方式和消息迭代的稳定性.

RB模型是一个具有规则因子图结构的随机实例产生模型,目前得到了国内外学者的重视,是国际上应用最为广泛的难解实例产生模型,该模型的难解实例被应用于各种国际算法竞赛(如,ACM竞赛等)^[34].基于此,我们研究基于RB模型的置信传播算法的收敛性.文献[22]中提出了一种基于变量熵的置信传播算法来求解RB模型实例,实验结果表明,该算法能够有效求解接近相变点的RB可满足实例.然而,未能从理论上解释这种现象.

RB模型是一个具有增长定义域的实例产生模型,已有的关于算法的收敛性分析都是基于具有固定定义域的情形.文献[29]中利用向量空间中函数的压缩映射原理给出了置信传播算法收敛的一般条件,但该判定条件的计算过程比较复杂,且适合于因子图结构较为简单的CSP实例.本文借鉴文献[29]中的证明方法,充分利用RB实例产生模型中不相容赋值的概率分布,分析基于变量熵的置信传播算法的收敛性,给出了该算法收敛的一个概率条件.具体来讲,在 $RB(k,n,\alpha,r_c,p)$ 模型中,取 $k=2$, $\alpha > \frac{1}{k}$, $r_c > 0$ 均为常数,且满足 $ke^{-\frac{\alpha}{r_c}} \geq 1$,概率 p 控制每个约束的不协调赋值的个数.我们证明,如果 $p \in (0, n^{-2\alpha})$,则基于变量熵的BP算法在 $RB(k,n,\alpha,r_c,p)$ 模型产生的随机二元可满足实例集上高概率收敛.最后,在 $RB(k,n,\alpha,r_c,p)$ 模型上选取了几组不同的数据进行数值模拟,实验表明该结论有效.更进一步地,当问题规模 n 增大时,在 $RB(k,n,\alpha,r_c,p)$ 模型的可满足区域,实验收敛区间趋于一个固定范围,而理论收敛区间逐渐变窄.原因在于, $RB(k,n,\alpha,r_c,p)$ 模型是一个具有增长定义域的随机CSP实例产生模型,不协调赋值的数目与参数 p 及问题规模 n 有关.

1 问题描述

设有一个变量集合 $X = \{x_1, \dots, x_n\}$,每个变量从其定义域 $D = \{1, \dots, d_n\}$ 中取值.从变量集中随机选取 $k(k \geq 2)$ 个不同的变量 $\{x_i^1, \dots, x_i^k\}$ 构成约束 C_i .对于每个约束 C_i ,都有一个相应的不协调赋值集合 $Q_{C_i} \subset D^k$ 来限制这 k 个变量的取值.若这 k 个变量的取值属于集合 Q_{C_i} ,则称此约束是不可满足的;反之,则称此约束是可满足的.由以下两步生成一个 $RB(k,n,\alpha,r_c,p)$ 模型实例:

- 1) 从可能的 $\binom{n}{k}$ 个约束中随机选取 $m = r_c n \ln n$ 个约束构成集合 $\{C_i\}_{i=1}^m$,其中, $r_c > 0$ 是常数;
- 2) 对于每个约束 C_i ,从可能的 d_n^k 个赋值中随机不重复地选取 $|Q_{C_i}| = pd_n^k$ 个赋值作为不协调赋值集合,其

中, $0 < p < 1$ 称为约束紧度, $d_n = n^\alpha$, $\alpha > 0$ 为常数.

这 m 个约束就组成 $RB(k, n, \alpha, r_c, p)$ 模型的一个实例. 如果存在这 n 个变量的一组赋值, 使得所有的约束同时被满足, 则这组赋值就是该实例的一个解.

在文献[18]中, 用一阶矩和二阶矩方法证明了 $RB(k, n, \alpha, r_c, p)$ 模型存在精确的可满足性相变现象. 具体来讲, 用 $\Pr\{SAT\}$ 表示 $RB(k, n, \alpha, r_c, p)$ 模型实例是可满足的概率, 则有如下结论成立.

定理 1^[18]. 设 $p_c = 1 - e^{-\frac{\alpha}{k}}$. 如果 k, α, r_c 满足 $ke^{-\frac{\alpha}{k}} \geq 1$, 且 $\alpha > \frac{1}{k}, r_c > 0$ 均为常数, 则

$$\lim_{n \rightarrow \infty} \Pr\{SAT\} = \begin{cases} 0, & \text{if } p > p_c \\ 1, & \text{if } p < p_c \end{cases} \quad (1)$$

由此可见, 当变量规模 n 充分大以后, $RB(k, n, \alpha, r_c, p)$ 模型的随机实例在 p_c 处发生了可满足性的突变. 也就是说: 当 $p < p_c$ 时, $RB(k, n, \alpha, r_c, p)$ 模型产生的随机实例以接近概率 1 可满足; 当 $p > p_c$ 时, $RB(k, n, \alpha, r_c, p)$ 模型产生的随机实例以接近概率 1 不可满足. 当 $k \geq 2$ 时, $RB(k, n, \alpha, r_c, p)$ 模型实例的判定问题是 NP 完全的. 因此, 为了解方便, 本文只考虑 $k=2$ 的 $RB(k, n, \alpha, r_c, p)$ 模型实例 F . 该方法容易推广到 $k > 2$ 的多元实例的情形.

二元 ($k=2$) $RB(k, n, \alpha, r_c, p)$ 模型实例的因子图表示形式如图 1 所示. 因子图是一个二部图, 图中的结点有两类: 一类是变量结点(图中用圆环表示, 记作 i, j, k, \dots); 另一类是约束结点(图中用方框表示, 记作 a, b, c, d, \dots).

在因子图的每条边 (i, a) 上, 定义变量发送给约束的消息 $u_{i \rightarrow a}(s_i)$ 以及约束发送给变量的消息 $\eta_{a \rightarrow i}(s_i)$. 其中, $u_{i \rightarrow a}(s_i)$ 表示在没有约束 a 的情况下, 变量 i 取值为 $s_i \in D$ 的概率, $\eta_{a \rightarrow i}(s_i)$ 表示约束 a 要求变量 i 取值为 $s_i \in D$ 的概率. $V(i)$ 表示包含变量 i 的约束结点集合, $V(a)$ 表示约束 a 中的变量结点集合.

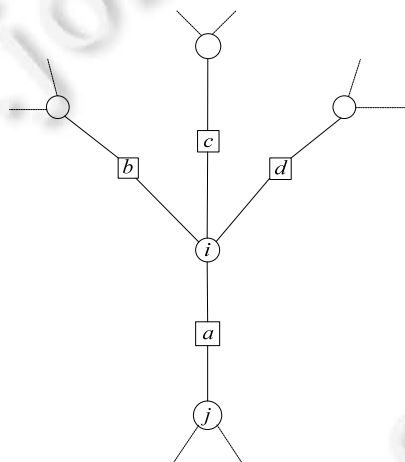


Fig.1 Factor graph of binary RB model instances
图 1 二元 RB 模型实例的因子图

求解二元 $RB(k, n, \alpha, r_c, p)$ 模型实例的 BP 算法的消息迭代方程为

$$u_{j \rightarrow a}^{(t)}(s_j) = \frac{1}{Z^{j \rightarrow a}} \prod_{b \in V(j) \setminus a} \eta_{b \rightarrow j}^{(t)}(s_j) \quad (2)$$

$$\eta_{a \rightarrow i}^{(t+1)}(s_i) = \frac{1}{Z^{a \rightarrow i}} \sum_{s_j \in D, j \in V(a) \setminus i} \delta_{(s_i, s_j) \in Q_a} u_{j \rightarrow a}^{(t)}(s_j) \quad (3)$$

其中, $Z^{j \rightarrow a}$ 和 $Z^{a \rightarrow i}$ 为归一化因子, $\delta_{(s_i, s_j) \in Q_a} = \begin{cases} 0, & \text{if } (s_i, s_j) \in Q_a \\ 1, & \text{o.w.} \end{cases}$ 为标识函数. 因此, 概率为

$$\Pr\{\delta_{(s_i, s_j) \in Q_a} = 0\} = p, \quad \Pr\{\delta_{(s_i, s_j) \in Q_a} = 1\} = 1 - p.$$

将公式(2)和公式(3)合并, 得到如下迭代方程:

$$\eta_{a \rightarrow i}^{(t+1)}(s_i) = \frac{1}{Z^{a \rightarrow i}} \sum_{s_j \in D, j \in V(a) \setminus i} \delta_{(s_i, s_j) \in Q_a} \prod_{b \in V(j) \setminus a} \eta_{b \rightarrow j}^{(t)}(s_j) \quad (4)$$

当公式(4)收敛时,可以根据固定点 $\eta_{a \rightarrow i}^*(s_i)$ 得到变量 i 取值 s_i 的边缘概率分布 $p(s_i)$ 的近似值 $b_i(s_i)$. 具体如下:

$$p(s_i) \approx b_i(s_i) = \frac{1}{Z^i} \prod_{a \in V(i)} \eta_{a \rightarrow i}^*(s_i) \quad (5)$$

其中, $Z^i = \sum_{s_j \in D} \prod_{a \in V(i)} \eta_{a \rightarrow i}^*(s_i)$ 为归一化因子. 对于每个变量 i , 计算熵:

$$S_i = -\sum_{s_j \in D} b_i(s_j) \log(b_i(s_j)) \quad (6)$$

取出熵值 S_i 最小的变量 i^* , 将变量 i^* 的取值固定到分量 $s_i^* = \max_{s_i} b_{i^*}(s_i)$ 上.

文献[22]中给出了求解 $RB(k, n, \alpha, r, c, p)$ 模型实例的基于变量熵的 BP 算法, 在该算法中, 将变量分为 3 种类型.

- A 类型: 已经被固定赋值的变量.
- B 类型: 与 A 类型的变量出现在同一个约束中的变量.
- C 类型: 其他变量.

具体算法如下:

输入: 二元 RB 模型的一个实例, 最大迭代次数 t_{\max} 和精度 ε .

输出: 实例的解或者算法不收敛.

Step 1. $t \leftarrow 0$, 对于因子图中的每条边 (a, i) , 随机初始化信息 $\eta_{a \rightarrow i}^{(0)}(s_i) \in [0, 1]$.

Step 2. $t \leftarrow t+1$, 根据公式(4)计算 $\eta_{a \rightarrow i}^{(t)}(s_i)$.

Step 3. 如果 $|\eta_{a \rightarrow i}^{(t)}(s_i) - \eta_{a \rightarrow i}^{(t-1)}(s_i)| < \varepsilon$, 则令 $\eta_{a \rightarrow i}^*(s_i) = \eta_{a \rightarrow i}^{(t)}(s_i)$, 并转到 step 5; 如果 $t > t_{\max}$, 则输出不收敛.

Step 4. 返回到 Step 2.

Step 5. 利用公式(5)计算每个变量的边缘概率分布 $b_i(s_i)$.

Step 6. 对于每个变量 i , 计算熵 $S_i = -\sum_{s_j \in D} b_i(s_j) \log(b_i(s_j))$, 挑选熵最小的变量 $i^* = \arg \min_{i \in V} S_i$, 将 i^* 的取值固定

到分量 $s_i^* = \arg \max_{s_i} b_{i^*}(s_i)$.

Step 7. $t \leftarrow 0$, 对于因子图中的每条边 (a, i) , 按如下方式初始化信息 $\eta_{a \rightarrow i}^{(0)}(s_i)$:

- (1) 对于 A 类型的变量, 不做处理.
- (2) 对于 B 类型的变量, 如果 i 的取值 s_i 与固定变量 j^* 的取值 s_j^* 有 $(s_i, s_j^*) \in Q_a$, 则令 $\eta_{a \rightarrow i}^{(0)}(s_i) = 0$; 否则, $\eta_{a \rightarrow i}^{(0)}(s_i) = 1$.
- (3) 对于 C 类型的变量, 随机初始化信息 $\eta_{a \rightarrow i}^{(0)}(s_i) \in [0, 1]$.

Step 8. $t \leftarrow t+1$:

- (1) 对于 A 类型的变量, 不做处理.
- (2) 对于 B 类型的变量, $\eta_{a \rightarrow i}^{(t)}(s_i) = \eta_{a \rightarrow i}^{(t-1)}(s_i)$.
- (3) 对于 C 类型的变量, 对于因子图中的每条边 (a, i) , 根据公式(4)计算 $\eta_{a \rightarrow i}^{(t)}(s_i)$.

Step 9.

- (1) 对于 A 类型的变量, 不做处理.
- (2) 对于 B 类型和 C 类型的变量, 如果 $|\eta_{a \rightarrow i}^{(t)}(s_i) - \eta_{a \rightarrow i}^{(t-1)}(s_i)| < \varepsilon$, 则令 $\eta_{a \rightarrow i}^*(s_i) = \eta_{a \rightarrow i}^{(t)}(s_i)$, 并转到 Step 11; 如果 $t > t_{\max}$, 则输出不收敛.

Step 10. 返回到 Step 8.

Step 11. 对于自由变量(没有被固定的变量), 利用公式(5)计算每个自由变量的边缘概率分布 $b_i(s_i)$.

Step 12. 对于每个自由变量 i , 计算熵 $S_i = -\sum_{s_j \in D} b_i(s_j) \log(b_i(s_j))$, 挑选熵最小的变量 $i^* = \arg \min_{i \in V} S_i$, 将 i^* 的取值

$$\text{固定到分量 } s_i^* = \arg \max_{s_i \in D} b_i^*(s_i).$$

Step 13. 返回到 Step 7,直到所有变量都被赋值.

Step 14. 对于每个约束 a ,如果变量的取值都有 $(s_i^*, s_j^*) \notin Q_a$,则输出这组赋值作为解;否则,输出算法不收敛.

注意到:在算法的迭代过程中,消息更新对 3 种类型的变量分别做不同的处理.对于 A 类型变量的消息不做处理,也即,关于 A 类型的变量在算法迭代中消息保持不变;对于 B 类型变量的消息,在相继两次迭代中也保持不变;对于 C 类型变量的消息,是利用公式(4)进行更新.从某种意义上说,在基于变量熵的 BP 算法的消息迭代过程中,A 类型变量和 B 类型变量可能加速或延迟算法收敛的时间,但并不对算法是否收敛产生影响.因此,算法是否收敛完全取决于消息更新方程式(4).实验结果表明,该算法求解 $RB(k, n, \alpha, r_c, p)$ 模型实例时非常有效,几乎能够求解接近相变点的实例^[22].然而对于算法的性能,至今缺少系统的理论解释.BP 算法是一类消息迭代算法,算法的收敛性直接影响算法的性能.对算法收敛性分析,有助于算法性能的进一步优化,对实际应用中信息传播算法的设计具有指导意义.本文分析了基于变量熵的 BP 算法求解 $RB(k, n, \alpha, r_c, p)$ 模型实例的收敛性,给出了算法收敛的一个概率条件.

2 主要结论及证明

基于变量熵的置信传播算法求解 $RB(k, n, \alpha, r_c, p)$ 模型实例时不是总有效,常表现为不收敛.事实上,对于二元 $RB(k, n, \alpha, r_c, p)$ 模型实例,有如下结论:

定理 2. 在二元 $RB(k, n, \alpha, r_c, p)$ 模型中,如果 $p \in (0, n^{-2\alpha})$,则基于变量熵的 BP 算法在该模型产生的随机实例集上高概率收敛.

在二元 $RB(k, n, \alpha, r_c, p)$ 模型中,每个变量的取值为 $s_i \in D = \{s_1, s_2, \dots, s_{d_a}\}$.相应地,BP 算法中的每个消息可表示为向量形式 $\eta_{a \rightarrow i}^{(t)} = (\eta_{a \rightarrow i}^{(t)}(s_1), \eta_{a \rightarrow i}^{(t)}(s_2), \dots, \eta_{a \rightarrow i}^{(t)}(s_{d_a}))$.

为了方便处理,我们对消息更新方程式(4)取对数,即 $\lambda_{a \rightarrow i}^{(t+1)} = \log \eta_{a \rightarrow i}^{(t+1)}$,则公式(4)表示为

$$\lambda_{a \rightarrow i}^{(t+1)}(s_i) = \log \left(\sum_{s_j \in D, j \in V(a) \setminus i} \delta_{(s_i, s_j) \in Q_a} h^{a \setminus i}(s_j) \right) \tag{7}$$

$$h^{a \setminus i}(s_j) = \exp \left(\sum_{b \in V(j) \setminus a} \lambda_{b \rightarrow j}^{(t)}(s_j) \right) \tag{8}$$

每个消息 $\lambda_{a \rightarrow i}^{(t+1)}$ 是 $V^{a \rightarrow i} = \mathbb{R}^{|D|}$ 中的向量.对于 $s \in D$,用 $\lambda_{a \rightarrow i}^{(t+1)}(s)$ 表示 $\lambda_{a \rightarrow i}^{(t+1)}$ 的某个分量的值.定义全局向量空间 $V = \bigoplus_{i \in a \in F} V^{a \rightarrow i}$,其中 \bigoplus 表示向量的方向和.定义函数 $f: V \rightarrow V$,假设算法中的消息更新并行执行.自然,可以用函数 f 的迭代过程来刻画算法中的消息更新过程.其中,消息更新方程式(7)可表示为函数 f 的分量.

设 $(V, \|\cdot\|)$ 是一个有限尺度的实向量赋范空间,由范数 $\|\cdot\|$ 诱导出 V 上的一个距离: $d(v, w) = \|v - w\|$.若存在 $0 \leq k < 1$,对于任意 $x, y \in V$,有 $d(f(x), f(y)) \leq kd(x, y)$ 成立,则称映射 $f: V \rightarrow V$ 是一个关于 d 的压缩映射.由代数知识,如果 $f: V \rightarrow V$ 是一个关于完备度量空间 (V, d) 的压缩映射,那么 f 有唯一固定点 $x_\infty \in V$,且与初始信息无关.也即,对于任意 $x \in V$,有序列 $x, f(x), f^2(x), \dots, x_\infty$,称 f 收敛到 x_∞ .

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. □

根据中值定理, 如果函数 $f: V \rightarrow V$ 可微, 那么对于任意 $x, y \in V$, 有公式(9)成立:

$$\|f(y) - f(x)\| \leq \|y - x\| \cdot \sup_{z \in [x, y]} \|f'(z)\| \tag{9}$$

其中, $[x, y] = \{cx + (1-c)y : c \in [0, 1]\}$.

因此, 如果 $\sup_{v \in V} \|f'(v)\| < 1$, 那么 f 是压缩函数, 收敛于固定点 $x_\infty \in V$, 且与初始信息无关.

更进一步地, 如果 $\sup_{v \in V} \|f'(v)\| < 1$, 则 BP 算法收敛, 且与初始消息无关. 因此, 关键问题是构造范数 $\|f'(v)\|$.

赋范空间 $(V, \|\cdot\|)$ 可诱导一个关于映射 $A: V \rightarrow V$ 的矩阵范数: $\|A\| = \sup_{\substack{v \in V \\ \|v\| \leq 1}} \|Av\|$.

ℓ_1 -范数定义为 $\|A\|_1 = \max_{j \in \{1, 2, \dots, N\}} \sum_{i=1}^N |A_{ij}|$, ℓ_∞ -范数定义为 $\|A\|_\infty = \max_{i \in \{1, 2, \dots, N\}} \sum_{j=1}^N |A_{ij}|$.

对于子向量空间 $V^{a \rightarrow i}$, 设其上的范数表示为 $\|\cdot\|_{a \rightarrow i}$. 向量空间 V 上的范数定义为

$$\|v\| = \sum_{a \rightarrow i} \|v^{a \rightarrow i}\|_{a \rightarrow i} \tag{10}$$

设 $\lambda \in V, f'(\lambda): V \rightarrow V$ 为 $f: V \rightarrow V$ 在 λ 处的导函数.

在 ℓ_1 -范数下, 则 $\|f'(\lambda)\|_1 = \max_{b \rightarrow j} \sum_{a \rightarrow i} \|f'(\lambda)_{a \rightarrow i, b \rightarrow j}\|_{a \rightarrow i}^{b \rightarrow j}$, 而 $\|f'(\lambda)_{a \rightarrow i, b \rightarrow j}\|_{a \rightarrow i}^{b \rightarrow j} = \sup_{\substack{v \in V^{b \rightarrow j} \\ \|v\|_{b \rightarrow j} \leq 1}} \|f'(\lambda)_{a \rightarrow i, b \rightarrow j} v\|_{a \rightarrow i}$.

对于二元 $RB(k, n, \alpha, r_c, p)$ 模型实例的每一个约束 $a \in F$, 假设 $\forall i \in a, \forall s_i \in D, \exists s_j \in D: \delta_{(s_i, s_j) \in Q_a} > 0$, 那么方程式(7)的导数为

$$\frac{\partial \lambda_{a \rightarrow i}^{(t+1)}(s_i)}{\partial \lambda_{b \rightarrow j}^{(t)}(s_j)} = 1_{V(j) \setminus a}(b) 1_{V(a) \setminus i}(j) \times \frac{\delta_{(s_i, s_j) \in Q_a} h^{a \setminus i}(s_j)}{\sum_{s'_j \in D} \delta_{(s_i, s'_j) \in Q_a} h^{a \setminus i}(s'_j)} \tag{11}$$

其中, 标识函数 $1_X(x) = \begin{cases} 1, & \text{if } x \in X \\ 0, & \text{o.w.} \end{cases}$. 因此,

$$\|f'(\lambda)\|_1 = \max_{b \rightarrow j} \sum_{a \rightarrow i} 1_{V(j) \setminus a}(b) 1_{V(a) \setminus i}(j) B_{a \rightarrow i, b \rightarrow j}(h^{a \setminus i}(\lambda)) \tag{12}$$

其中, $B_{a \rightarrow i, b \rightarrow j}(h^{a \setminus i}(\lambda)) = \left\| \frac{\delta_{(s_i, s_j) \in Q_a} h^{a \setminus i}(\lambda)}{\sum_{s'_j \in D} \delta_{(s_i, s'_j) \in Q_a} h^{a \setminus i}(\lambda)} \right\|_{a \rightarrow i}^{b \rightarrow j}$. 因此, 只要满足 $\sup_{\lambda \in F} \|f'(\lambda)\|_1 < 1$, BP 算法就收敛.

定义矩阵 A 中的元素为

$$A_{a \rightarrow i, b \rightarrow j} = \sup_{h^{a \setminus i} > 0} B_{a \rightarrow i, b \rightarrow j}(h^{a \setminus i}(\lambda)) \tag{13}$$

假设 $\|\cdot\|_{a \rightarrow i}$ 是 $V^{a \rightarrow i} = \mathbb{R}^{|D|}$ 上的 l_∞ 范数, 由代数知识, $A_{a \rightarrow i, b \rightarrow j}$ 可表示为如下形式^[29]:

$$A_{a \rightarrow i, b \rightarrow j} = N(\delta, i, j) = \sup_{\substack{s_i \neq s'_i \\ s_i, s'_i \in D}} \sup_{\substack{s_j \neq s'_j \\ s_j, s'_j \in D}} \frac{\sqrt{\delta_{(s_i, s_j) \in Q_a} \delta_{(s'_i, s'_j) \in Q_a}} - \sqrt{\delta_{(s'_i, s_j) \in Q_a} \delta_{(s_i, s'_j) \in Q_a}}}{\sqrt{\delta_{(s_i, s_j) \in Q_a} \delta_{(s'_i, s'_j) \in Q_a}} + \sqrt{\delta_{(s'_i, s_j) \in Q_a} \delta_{(s_i, s'_j) \in Q_a}}} \tag{14}$$

根据函数的压缩映射原理及公式(9)、公式(12)~公式(14), 有如下推论:

推论 1. 设 $\forall i \in a, \forall s_i \in D, \exists s_j \in D: \delta_{(s_i, s_j) \in Q_a} > 0$. 若 $\max_{b \rightarrow j} \sum_{a \rightarrow i} 1_{V(j) \setminus a}(b) 1_{V(a) \setminus i}(j) N(\delta, i, j) < 1$, 则 BP 算法收敛到唯一固定点, 且与初始信息无关.

定义矩阵 M 的元素为 $M_{a \rightarrow i, b \rightarrow j} = 1_{V(j) \setminus a}(b) 1_{V(a) \setminus i}(j) N(\delta, i, j)$, $\lambda_1, \dots, \lambda_n$ 是矩阵 M 的特征值, $\lambda(M) = \{\lambda_1, \dots, \lambda_n\}$ 为矩阵 M 的谱, $\rho(M) = \sup\{|\lambda_i| : \lambda_i \in \lambda(M)\}$ 为矩阵 M 的谱半径.

定理 3. 设函数 $f: V \rightarrow V$ 可微, 进一步假设 $|f'(x)| \leq M$, M 是一个不依赖于 x 的非负矩阵. 如果 $\rho(M) < 1$, 那么对于任意 $x \in V, f$ 收敛到唯一固定点 x_∞ . 也即 $x, f(x), f^2(x), \dots, x_\infty$ (注意, $|f'(x)| \leq M$ 表示由 $f'(x)$ 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:

$$\|(f^n)'(x)\| = \left\| \prod_{i=1}^n f'(f^{i-1}(x)) \right\| = \left\| \prod_{i=1}^n f'(f^{i-1}(x)) \right\| \leq \left\| \prod_{i=1}^n f'(f^{i-1}(x)) \right\| \leq \left\| \prod_{i=1}^n M \right\| = \|M^n\| \quad (15)$$

由 Gelfand 引理^[35], $\lim_{n \rightarrow \infty} \|M^n\|^{1/n} = \rho(M)$. 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, $\|(f^N)'(x)\| < 1$. 所以, $\|(f^N)(x)\|$ 是压缩函数.

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

因此, 根据定理 3 及公式(11), 有如下结论:

推论 2. 设 $\forall i \in a, \forall s_i \in D, \exists s_j \in D: \delta_{(s_i, s_j) \in Q_a} > 0$. 如果矩阵 M 的谱半径 $\rho(M)$ 严格小于 1, 则 BP 算法收敛到唯一固定点, 且与初始信息无关.

对于一个具体的二元 $RB(k, n, \alpha, r_c, p)$ 模型的随机实例 F , 依据 $RB(k, n, \alpha, r_c, p)$ 模型实例的生成方法, 每个约束 $I \in F$ 有 $n^{2\alpha}p$ 个不协调赋值和 $n^{2\alpha}(1-p)$ 个协调赋值. 也即, 有 $n^{2\alpha}p$ 个 $\delta_{(s_i, s_j) \in Q_I} = 0$ 的赋值和 $n^{2\alpha}(1-p)$ 个 $\delta_{(s_i, s_j) \in Q_I} = 1$ 的赋值. 下面给出 $\forall i \in I, \forall s_i \in D, \exists s_j \in D: \delta_{(s_i, s_j) \in Q_I} > 0$ 高概率成立的 p 的取值界, 有如下引理:

引理 2. 设 $p = 1 - \frac{c}{n^\alpha}$, c 为常数. 则 $c = \alpha$ 是 $\forall i \in I, \forall s_i \in D, \exists s_j \in D: \delta_{(s_i, s_j) \in Q_I} > 0$ 高概率成立的临界点.

证明: 设约束 I 的一个赋值为 $(s_i, s_j) \in D \times D$, 所有的赋值构成赋值矩阵 S , 其元素为 $s_{ij} = (s_i, s_j)$. 位于矩阵 S 的不同行、不同列元素构成一个排列 $s_{1i_1}, s_{2i_2}, \dots, s_{d_n i_{d_n}}$, 这样的排列共有 $d_n!$ 个. 如果在 $n^{2\alpha}(1-p)$ 个协调赋值中至少含有一个排列 $s_{1i_1}, s_{2i_2}, \dots, s_{d_n i_{d_n}}$ 中的元素, 则都有 $\forall i \in I, \forall s_i \in D, \exists s_j \in D: \delta_{(s_i, s_j) \in Q_I} > 0$ 成立. 设随机变量 x 表示在 $n^{2\alpha}(1-p)$ 个协调赋值中含有元素 $s_{1i_1}, s_{2i_2}, \dots, s_{d_n i_{d_n}}$ 的事件发生的数目, 则

$$\begin{aligned} \lim_{n \rightarrow \infty} E(x) &= \lim_{n \rightarrow \infty} \frac{n^\alpha! \binom{n^{2\alpha} - n^\alpha}{(1-p)n^{2\alpha} - n^\alpha}}{\binom{n^{2\alpha}}{(1-p)n^{2\alpha}}} \\ &= \lim_{n \rightarrow \infty} \frac{n^\alpha! ((1-p)n^{2\alpha})((1-p)n^{2\alpha} - 1) \dots ((1-p)n^{2\alpha} - n^\alpha + 1)}{n^{2\alpha} (n^{2\alpha} - 1) \dots (n^{2\alpha} - n^\alpha + 1)} \\ &= \lim_{n \rightarrow \infty} \frac{n^\alpha! ((1-p)) \left((1-p) - \frac{1}{n^{2\alpha}} \right) \dots \left((1-p) - \frac{n^\alpha - 1}{n^{2\alpha}} \right)}{\left(1 - \frac{1}{n^{2\alpha}} \right) \left(1 - \frac{2}{n^{2\alpha}} \right) \dots \left(1 - \frac{n^\alpha - 1}{n^{2\alpha}} \right)} \\ &= \lim_{n \rightarrow \infty} n^\alpha! (1-p)^{n^\alpha} \\ &\approx \lim_{n \rightarrow \infty} \frac{n^{\alpha \left(\frac{n^\alpha + 1}{2} \right)}}{e^{n^\alpha}} \cdot (1-p)^{n^\alpha + 1} \\ &\leq \lim_{n \rightarrow \infty} \frac{n^{\alpha(n^\alpha + 1)}}{e^{n^\alpha}} \cdot (1-p)^{n^\alpha + 1} \quad \left(\text{由于 } \alpha > \frac{1}{2} \right) \end{aligned} \quad (16)$$

因此, 当 $c > \alpha$ 时, $\lim_{n \rightarrow \infty} E(x) = 0$. 由一阶矩方法可知, $\Pr(x \geq 1) = 0$.

不妨设 $v_1 = (s_i, s_j), v_2 = (s_k, s_l)$ 是约束 I 的两个赋值, $\Pr(\langle v_1, v_2 \rangle)$ 表示赋值 v_1 和 v_2 同时满足约束 I 的概率.

- 如果赋值 v_1 和 v_2 相同, 则 $\Pr(\langle v_1, v_2 \rangle) = \frac{\binom{d_n^k - 1}{|Q_I|}}{\binom{d_n^k}{|Q_I|}} = 1 - p$;

- 如果赋值 v_1 和 v_2 不同,则 $\Pr(\langle v_1, v_2 \rangle) = \binom{d_n^k - 2}{|Q_i|} / \binom{d_n^k}{|Q_i|} = (1-p)^2 + O\left(\frac{1}{n^{2\alpha}}\right)$.

设 x_i, x_j 是 $n^{2\alpha}(1-p)$ 个协调赋值中分别含有元素 $s_{1i_1}, s_{2i_2}, \dots, s_{d_n i_{d_n}}$ 和 $s_{1j_1}, s_{2j_2}, \dots, s_{d_n j_{d_n}}$ 的两组赋值,所以,

$$\begin{aligned} E(x^2) &= \sum_{i=j} E(x_i, x_j) + \sum_{i \neq j} E(x_i, x_j) \\ &= E(x) + ((n^\alpha!)^2 - n^\alpha!) \left[(1-p)^2 + O\left(\frac{1}{n^{2\alpha}}\right) \right]^{n^\alpha} \\ &= E(x) + ((n^\alpha!)^2 - n^\alpha!) (1-p)^{2n^\alpha} + O\left(\frac{1}{n^{2\alpha}}\right) \end{aligned} \tag{17}$$

进而有:

$$\lim_{n \rightarrow \infty} \frac{E(x^2)}{E^2(x)} = \lim_{n \rightarrow \infty} \frac{E(x) + ((n^\alpha!)^2 - n^\alpha!) (1-p)^{2n^\alpha}}{E^2(x)} = \lim_{n \rightarrow \infty} \left(\frac{1}{E(x)} + 1 - \frac{1}{n^\alpha!} \right) \tag{18}$$

当 $c < \alpha$ 时, $\lim_{n \rightarrow \infty} E(x) = \infty$. 因此,

$$\lim_{n \rightarrow \infty} \frac{E(x^2)}{E^2(x)} \leq \lim_{n \rightarrow \infty} \left[\frac{1}{E(x)} + 1 - \frac{1}{n^\alpha!} \right] = 1 + o(1) \tag{19}$$

由二阶矩方法, $\Pr(x=0)=0$.

所以, $c=\alpha$ 是 $\forall i \in I, \forall s_i \in D, \exists s_j \in D: \delta_{(s_i, s_j) \in Q_i} > 0$ 高概率成立的临界点. □

引理 2 表明, 当 $p < 1 - en^{-\alpha}$ 时, 则 $\forall i \in I, \forall s_i \in D, \exists s_j \in D: \delta_{(s_i, s_j) \in Q_i} > 0$ 高概率成立. 注意到, 在 $RB(k, n, \alpha, r_c, p)$ 模型中的可满足相变点为 $p_c = 1 - e^{-\frac{\alpha}{c}}$, 而 $1 - e^{-\frac{\alpha}{c}} < 1 - en^{-\alpha}$. 根据定理 1 和引理 2, 在 $RB(k, n, \alpha, r_c, p)$ 模型的可满足区域, 总有 $\forall i \in I, \forall s_i \in D, \exists s_j \in D: \delta_{(s_i, s_j) \in Q_i} > 0$ 高概率成立. 因此, 推论 2 可用作为判定 BP 算法在 $RB(k, n, \alpha, r_c, p)$ 模型的可满足实例上是否高概率收敛的充分条件. 我们有如下推论:

推论 3. 设 $p_c = 1 - e^{-\frac{\alpha}{c}}$, 其中, $\alpha > \frac{1}{k}, r_c > 0$ 均为常数, 且满足 $ke^{-\frac{\alpha}{c}} \geq 1$. 当 $p < p_c$ 时, 对于二元 $RB(k, n, \alpha, r_c, p)$ 模型实例 F , 如果矩阵 M 的谱半径 $\rho(M)$ 严格小于 1, 则 BP 算法高概率收敛, 且与初始信息无关 (注意, 这里的高概率指引理 2 中的高概率).

注意到, 矩阵 M 中的元素由 $N(\delta, i, j)$ 决定, 为此给出 $N(\delta, i, j)=0$ 高概率成立的条件.

引理 3. 设 $p=n^{-c}, c$ 为常数. 当 $c > 2\alpha$ 时, $N(\delta, i, j)=0$ 高概率成立.

证明: 设 x 表示约束 I 不协调赋值的数目的随机变量. 由马尔可夫不等式得: $\Pr\{x \geq 1\} \leq E[x]$. 而

$$E[x] = n^{2\alpha} p = n^{2\alpha} n^{-c} = n^{2\alpha - c} \tag{20}$$

当 $c > 2\alpha$ 时, $\lim_{n \rightarrow \infty} E(x) = 0$. 所以, $\Pr(x \geq 1) = 0$, 进一步有 $\Pr(x < 1) = 1$. 因此, 根据公式(14)有 $\Pr\{N(\delta, i, j)=0\} = 1$. □

在矩阵分析中, 我们知道: 关于非负方阵 M 的谱半径 $\rho(M)$, 如果存在一个矩阵范数 $\|\cdot\|$ 使得 $\|M\| < 1$, 那么 $\lim_{n \rightarrow \infty} M^n = 0$; 如果 $\lim_{n \rightarrow \infty} M^n = 0$ 当且仅当 $\rho(M) < 1$. 下面给出 BP 算法在二元 $RB(k, n, \alpha, r_c, p)$ 模型上高概率收敛的条件.

定理 4. 设 $p=n^{-c}, c$ 为常数. 当 $c > 2\alpha$ 时, BP 算法在二元 $RB(k, n, \alpha, r_c, p)$ 模型实例集上高概率收敛.

证明: 由引理 2 可知: 当 $p_1 < 1 - en^{-\alpha}$ 时, 高概率地有 $\forall i \in I, \forall s_i \in D, \exists s_j \in D: \delta_{(s_i, s_j) \in Q_i} > 0$ 成立. 而引理 3 中, 当 $p_2 < n^{-2\alpha}$ 时, 高概率地有 $\|M\|_\infty = \max_{i \in \{1, 2, \dots, N\}} \sum_{j=1}^N |M_{ij}| < 1$. 所以, 高概率地有 $\rho(M) < 1$ 成立. 容易看出, $p_2 \leq p_1$. 因此, 当 $p < n^{-2\alpha}$ 时,

BP 算法在二元 $RB(k, n, \alpha, r_c, p)$ 模型实例集上高概率收敛. 至此, 定理 2 得证. □

定理 2 表明, 如果 $p \in (0, n^{-2\alpha})$, 则基于变量熵的 BP 算法在二元 $RB(k, n, \alpha, r_c, p)$ 模型产生的随机实例集上高概率收敛.

上述有关证明过程及结论容易扩展到多元 RB 模型实例上.

3 数值实验及分析

为了对本文定理 2 中的结论作出实验验证,数值实验中选用 RB 模型实例.在 RB 模型中,我们选择了两组不同的参数,分别为 $\alpha=0.8, r_c=3$ 和 $\alpha=0.6, r_c=1$,用来生成随机 CSP 实例.对于每组参数取 5 个不同规模 n 的 CSP 随机实例,见表 1 和表 2(表 1 中,取 $\alpha=0.8, r_c=3$,对于不同的问题规模 n ,相应的 d_n 和 m 的值;表 2 中,取 $\alpha=0.6, r_c=1$,对于不同的问题规模 n ,相应的 d_n 和 m 的值).在算法中,取 $t_{\max}=10^3, \varepsilon=10^{-4}$.实验运行的软、硬件环境为: Intel Core i7-2600 CPU,主频 3.40GHZ,内存 4.00GB,操作系统 WIN7,采用 VC++6.0 集成开发工具,使用 C 和 MATLAB 语言编程.

Table 1 First data of numerical experiment

表 1 数值实验中的第 1 组数据

n	α	d_n	r_c	m
20	0.8	11	3	180
40	0.8	19	3	443
60	0.8	26	3	737
80	0.8	33	3	1 052
100	0.8	40	3	1 382

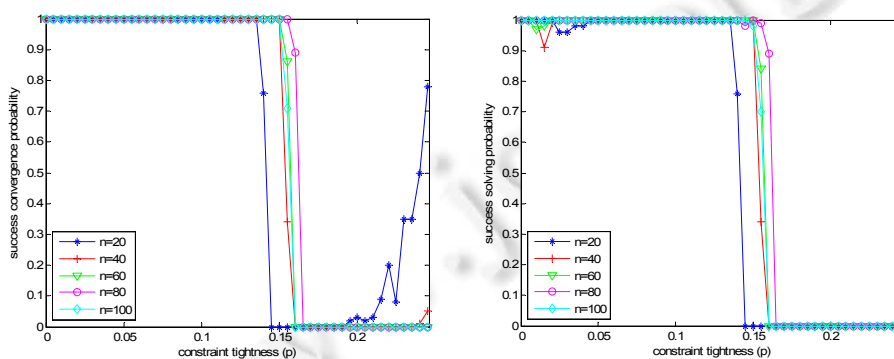
Table 2 Second data of numerical experiment

表 2 数值实验中的第 2 组数据

n	α	d_n	r_c	m
100	0.6	16	1	461
150	0.6	20	1	752
200	0.6	24	1	1 060
250	0.6	27	1	1 380
300	0.6	31	1	1 711

在 $RB(k, n, \alpha, r_c, p)$ 模型中,可满足相变点 $p_c = 1 - e^{-\frac{\alpha}{r_c}}$. 对于第 1 组数据, $p_c=0.234$. 对于第 2 组数据, $p_c=0.451$. 数值实验中,每个数据点由 100 个 $RB(k, n, \alpha, r_c, p)$ 模型的随机实例构成. 第 1 组数据和第 2 组数据的运行结果分别如图 2 和图 3 所示.

图 2(a)是基于变量熵的 BP 算法在第 1 组数据上成功收敛的概率随着参数 p 变化的情形. 当 $p < 0.145$ 时,算法高概率收敛. 当 $p > 0.145$ 时,算法从收敛突变到几乎不收敛. 并且在 $p > 0.195$ 时,算法成功收敛的概率逐渐增加. 我们发现:在 $0.160 < p < 0.195$ 区间内,算法几乎高概率不收敛,即算法失效. 事实上,此区间恰好为 $RB(k, n, \alpha, r_c, p)$ 模型的难解区间,实例的求解难度随着变量规模的增加成指数形式增长.



(a) BP 算法成功收敛的运行结果

(b) BP 算法有效求解的运行结果

Fig.2 Results of the first data

图 2 第 1 组数据的运行结果

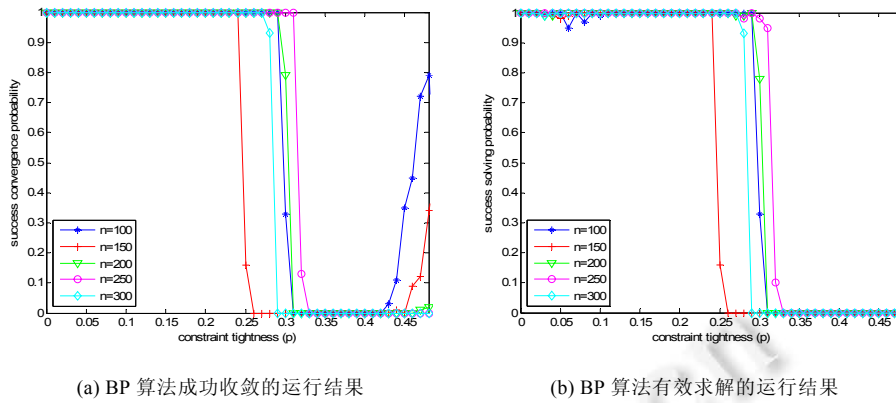


Fig.3 Results of the second data

图3 第2组数据的运行结果

图 2(b)是基于变量熵的 BP 算法在第 1 组数据上能够有效求解实例的概率随着参数 p 变化的情形.当 $p < 0.145$ 时,算法几乎能够高概率地给出实例的解.当 $p > 0.145$ 时,算法不能给出实例的解:一方面,算法本身不收敛;另一方面,算法收敛但不能正确给出实例的解.

对比图 2(a)与图 2(b),我们发现,当 $p > 0.195$ 时,对于部分实例该算法收敛,但未能正确给出实例的解.

对于第 2 组数据,基于变量熵的 BP 算法的运行结果如图 3 所示.当 $p < 0.245$ 时,算法高概率地收敛.当 $p > 0.245$ 时,算法从收敛突变到几乎不收敛.并且在 $p > 0.420$ 时,算法成功收敛的概率逐渐增加.在 $0.325 < p < 0.420$ 区间内,算法几乎高概率不收敛,即算法失效,如图 3(a)所示.

图 3(b)是算法在第 2 组数据上能够解的运行结果.当 $p < 0.245$ 时,算法几乎能够高概率地给出实例的解.当 $p > 0.245$ 时,算法不能给出实例的解.我们发现,当 $p > 0.420$ 时,算法在部分实例收敛,但未能给出实例的解.

为了进一步分析和验证本文结论,并与其他结论进行比较,我们选取 RB 模型中的参数见表 3.对于表 3 中的第 1 组数据 $(n, \alpha, d_n, r_c, m) = (20, 0.4, 3, 4, 240)$,根据定理 2 可计算出算法收敛的条件 p 的取值范围为 $(0, 0.091)$.同理,对于第 2 组数据 $(n, \alpha, d_n, r_c, m) = (40, 0.4, 4, 7, 1033)$ 和第 3 组数据 $(n, \alpha, d_n, r_c, m) = (60, 0.4, 5, 10, 2457)$, p 的取值范围分别为 $(0, 0.052)$ 和 $(0, 0.038)$,这 3 组数据对应实例的相变点 p_c 分别为 0.095, 0.056, 0.039,具体参数值见表 4.不难看出,在定理 2 的收敛性判定条件中, p 的取值上界几乎接近相变点 p_c .进一步说明,对于接近相变点的复杂因子图结构的二元 RB 模型实例,定理 2 的判定条件几乎有效.针对表 3 中的第 1 组数据,选取 p 为 0.078,由 $RB(k, n, \alpha, r_c, p)$ 模型产生 1 000 个二元 CSP 实例进行数值实验,本文能够有效判定的概率为 0.87,文献[29]中能够有效判定的概率为 0.91;同样,对于表 3 中的第 2 组数据和第 3 组数据,本文能够有效判定的概率为 1.00 和 0.93,而文献[29]中分别为 0.84 和 0.79.表 4 中给出了本文结论与文献[29]中结论的对比结果(每组参数生成 1 000 个 RB 模型实例).

Table 3 Values of parameter in RB model

表 3 RB 模型中的相关参数值

n	α	d_n	r_c	m
20	0.4	3	4	240
40	0.4	4	7	1 033
60	0.4	5	10	2 457

Table 4 Comparison of relevant experimental results

表 4 相关实验结果比较

$(0, n^{-2\alpha})$	p_c	p	本文结果	文献[29]结果
$(0, 0.091)$	0.095	0.078	0.87	0.91
$(0, 0.052)$	0.056	0.049	1.00	0.84
$(0, 0.038)$	0.039	0.037	0.93	0.79

当问题规模 n 增大时,理论收敛区间逐渐减小.原因在于, $RB(k,n,\alpha,r_c,p)$ 模型是一个具有增长定义域的随机 CSP 实例产生模型,不协调赋值的数目与参数 p 及问题规模 n 有关.

用难解实例测试算法性能及相关结论是非常必要的,我们使用国际上常用的由 Lecoutre 提供的 benchmark 案例进行测试^[34].数值实验中选取了两组实例集 frb30-15 和 frb40-19 进行测试,结果表明,基于变量熵的置信传播算法在这些实例上基本不收敛.事实上,在这些实例集上,定理 2 中的条件已经不能满足.也即,定理 2 中判定算法收敛的条件失效.当然,不能仅凭某些局部的实验数据就断定某种方法的优劣,每种方法都有其优越性和局限性.

4 结束语

本文分析了基于变量熵的置信传播算法在二元 $RB(k,n,\alpha,r_c,p)$ 模型随机实例集上的收敛性,利用置信传播算法的信息更新函数在实向量空间中的压缩性原理,及更新函数的系数矩阵的谱半径等理论,证明了当 $p \in (0, n^{-2\alpha})$ 时,该算法在二元 $RB(k,n,\alpha,r_c,p)$ 模型产生的随机 CSP 实例集上高概率收敛,给出了置信传播算法能够有效求解二元 $RB(k,n,\alpha,r_c,p)$ 模型实例的理论解释.我们希望这些方法的引入可以提供一种新的思路,用来分析其他信息传播算法的收敛性及性能.进一步工作是:(1) 扩展基于变量熵的 BP 算法在 $RB(k,n,\alpha,r_c,p)$ 模型实例集上的收敛区间;(2) 分析基于变量熵的 BP 算法在 $RB(k,n,\alpha,r_c,p)$ 模型实例集上的有效性,也即,算法收敛后能否给出问题的有效解.

References:

- [1] Martin D, Alan F, Michael M. A probabilistic analysis of randomly generated binary constraint satisfaction. *Theory Computer Science*, 2003,290(3):1815–1828. [doi: 10.1016/S0304-3975(02)00317-1]
- [2] Creignou N, Auade H. The SAT-UNSAT transition for random constraint satisfaction problems. *Discrete Mathematics*, 2009,309(8): 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(1-2):3–67. [doi: 10.1016/S0304-3975(01)00149-9]
- [4] Tsang E. A glimpse of constraint satisfaction. *Artificial Intelligence Review*, 1999,13(3):215–277. [doi: 10.1023/A:1006558104682]
- [5] Xu K, Li W. The SAT phase transition. *Science in China, Series E*, 1999,42(5):494–501. [doi: 10.1007/BF02917402]
- [6] Fan Y, Shen J. On the phase transitions of random k -constraint satisfaction problems. *Artificial Intelligence*, 2011,175(3-4): 914–927. [doi: 10.1016/j.artint.2010.11.004]
- [7] Achlioptas D, Naro A, Peres Y. Rigorous location of phase transitions in hard optimization problems. *Nature*, 2005,435(7043): 759–764. [doi: 10.1038/nature03602]
- [8] Mertens S, Mézard M, Zecchina R. Threshold values of random k -SAT from the cavity method. *Random Structure Algorithms*, 2006,28(3):340–373. [doi:10.1002/rsa.20090]
- [9] 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]
- [10] Aurell E, Gordon U, Kirkpatrick S. Comparing beliefs, surveys, and random walks. *Advances in Neural Information Processing Systems*, 2004,17(1):1–8.
- [11] Mezard M, Parisi G. The cavity method at zero temperature. *Journal of Statistical Physics*, 2003,111(1-2):1–34. [doi: 10.1023/A:102221005097]
- [12] Mezard M, Zecchina R. Random k -satisfiability problem: From an analytic solution to an efficient algorithm. *Physical Review E*, 2002,66(5):No.056126. [doi:10.1103/PhysRevE.66.056126]
- [13] Maneva E, Mossel E, Wainwright M. A new look at survey propagation and its generalizations. *Journal of the ACM*, 2007,54(4): 1089–1098. [doi:10.1145/1255443.1255445]
- [14] Braunstein A, Mezard M, Zecchina R. Survey propagation: An algorithm for satisfiability. *Random Structures and Algorithms*, 2005,27(2):201–226. [doi:10.1002/rsa.20057]
- [15] Braunstein A, Zecchina R. Survey and belief propagation on random k -SAT. *Lecture Notes in Computer Science*, 2004,2919(1): 519–528. [doi:10.1007/978-3-540-24605-3_38]

- [16] Yedidia JS, Freeman WT, Weiss Y. Understanding belief propagation and its generalizations. *Artificial Intelligence*, 2003,8(1): 239–269.
- [17] Mezard M, Parisi G, Zecchina R. Analytic and algorithmic solution of random satisfiability problems. *Science*, 2002,297(5582): 812–815. [doi: 10.1126/science.1073287]
- [18] Xu K, Li W. Exact phase transitions in random constraint satisfaction problems. *Journal of Artificial Intelligence Research*, 2000, 12(1):93–103.
- [19] Liu T, Lin XX, Wang CY, Su KL, Xu K. Large hinge width on sparse random hyper graphs. In: *Proc. of the 22nd Int'l Joint Conf. on Artificial Intelligence*. Barcelona: AAAI, 2011. 611–616. [doi: 10.5591/978-1-57735-516-8/IJCAI11-109]
- [20] 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]
- [21] 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]
- [22] Zhao CY, Zheng ZM. A belief propagation algorithm based on variable entropy for constraint satisfaction problems. *Science China Information Sciences*, 2012,42(9):1170–1180 (in Chinese).
- [23] Kschischang F, Frey B, Loeliger H. Factor graph and the sum-product algorithm. *Information Theory*, 2001,47(1):498–519. [doi: 10.1109/18.910572]
- [24] Weiss Y. Correctness of local probability propagation in graphical models with loops. *Neural Computation*, 2000,12(1):1–41. [doi: 10.1162/089976600300015880]
- [25] 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]
- [26] 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.
- [27] Heskes T. On the uniqueness of loopy belief propagation fixed points. *Neural Computation*, 2004,16(11):2379–2413. [doi: 10.1162/0899766041941943]
- [28] Ihler AT, Fisher JW, Willsky AS. Loopy belief propagation: Convergence and effects of message errors. *Machine Learning Research*, 2005,6(1):905–936.
- [29] Mooij JM, Kappen HJ. Sufficient conditions for convergence of the sum-product algorithm. *IEEE Trans. on Information Theory*, 2007,53(12):4422–4437. [doi: 10.1109/TIT.2007.909166]
- [30] Shi XQ, Schonfeld D, Tuninetti D. Message error analysis of loopy belief propagation for the sum-product algorithm. *Computer and Information Science*, 2010,1009(2):1–30. [doi: 10.1109/ICASSP.2010.5495075]
- [31] Gamarnik D, Shah D, Wei Y. Belief propagation for min-cost network flow: Convergence and correctness. *Operations Research*, 2012,60(2):410–428. [doi: 10.1287/opre.1110.1025]
- [32] 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. [doi: 10.7155/jgaa.00310]
- [33] 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. [doi: 10.1109/TIT.2012.2231464]
- [34] Lecoute. <http://www.cril.univ-artois.fr/~lecoute/benchmarks.html>
- [35] Dieudonné J. *Foundations of modern analysis*. Vol.10-I, New York: Academic Press, 1969.

附中文参考文献:

- [22] 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:信息科学*,2012,42(9):1170–1180.



王晓峰(1980—),男,甘肃会宁人,博士,讲师,CCF 专业会员,主要研究领域为人工智能,机器学习,可计算性,计算复杂性.



许道云(1959—),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为人工智能,可计算性,计算复杂性.