

# 概率图模型中的变分近似推理方法

程强<sup>1</sup> 陈峰<sup>1</sup> 董建武<sup>1</sup> 徐文立<sup>1</sup>

**摘要** 概率图模型将图论和概率论相结合,为多个变量之间复杂依赖关系的表示提供了统一的框架,在计算机视觉、自然语言处理和计算生物学等领域有着广泛的应用. 概率推理(包括计算边缘概率和计算最大概率状态等问题)是概率图模型研究及应用的核心问题. 本文主要介绍概率图模型近似推理方法中变分推理的最新研究成果. 在变分近似推理的框架下,系统地归纳了概率图模型推理问题的基本研究思路,综述了目前主要的近似推理方法,并分析了近似算法的单调性、收敛性和全局性等性质. 最后,对概率图模型近似推理方法的研究方向和应用前景作了展望.

**关键词** 概率图模型, 贝叶斯网, 马尔科夫随机场, 近似推理, 变分法, 对偶分解

**引用格式** 程强, 陈峰, 董建武, 徐文立. 概率图模型中的变分近似推理方法. 自动化学报, 2012, 38(11): 1721–1734

**DOI** 10.3724/SP.J.1004.2012.01721

## Variational Approximate Inference Methods for Graphical Models

CHENG Qiang<sup>1</sup> CHEN Feng<sup>1</sup> DONG Jian-Wu<sup>1</sup> XU Wen-Li<sup>1</sup>

**Abstract** Graphical models bring together graph theory and probability theory in a powerful and compact formalism for multivariate statistical modeling. They have been successfully applied to a wide variety of fields, including computer vision, natural language processing, and computational biology. Probabilistic inference, such as finding the most probable assignment and computing the marginals, is the core problem to be solved in the applications of graphical models. In this paper, a detailed survey of the achievements on variational approximate inference methods obtained in recent years is presented. Besides, a discussion of future trend on approximate inference of graphical models is given.

**Key words** Graphical models, Bayesian network, Markov random field (MRF), approximate inference, variational method, dual decomposition

**Citation** Cheng Qiang, Chen Feng, Dong Jian-Wu, Xu Wen-Li. Variational approximate inference methods for graphical models. *Acta Automatica Sinica*, 2012, 38(11): 1721–1734

概率图模型<sup>[1–2]</sup>是由图论和概率论结合而成的描述多元统计关系的模型,它为多个变量之间复杂依赖关系的表示提供了统一的框架,具有紧凑有效、简洁直观的特点. 概率图模型推理问题从上世纪 90 年代开始被关注,到目前为止已经发展了 20 多年. 由于概率图模型对大规模复杂系统简约而紧凑的表示能力,及其学习问题和推理问题的理论价值,概率图模型的研究和应用在这 20 多年中一直蓬勃地发展着.

概率图模型在计算机视觉、自然语言处理、计

算生物学,以及控制理论、组合优化、统计物理学和经济学等领域有着广泛的研究和应用. 概率图模型在计算机视觉领域应用最为广泛,立体匹配<sup>[3]</sup>、图像分割<sup>[4]</sup>、光流估计<sup>[5]</sup>、图像去噪<sup>[6]</sup>、纹理合成<sup>[7]</sup>、图像补全<sup>[8]</sup>、目标识别<sup>[8]</sup>等计算机视觉问题都可以被归结为概率图模型的推理问题. 在信号处理领域,概率图模型主要被用于自然语言处理中的词性标注<sup>[9]</sup>、主题推理<sup>[10]</sup>、通信信号处理中的解码<sup>[11]</sup>等问题. 此外,信号处理中的很多算法都被证明与概率图模型上信度传播推理算法等价,例如低密度奇偶校验码<sup>[12]</sup>、Turbo 解码算法<sup>[11]</sup>、Kalman 滤波算法<sup>[13–14]</sup>等. 此外,生物信息学中的很多问题也可以归结为概率图模型的推理问题,例如基因序列分析<sup>[15–16]</sup>、蛋白质结构设计<sup>[17–18]</sup>等问题.

概率图模型推理问题是具有良序结构的组合优化问题,除了应用领域广泛外,其本身具有重要的研究价值. 概率图模型的近似推理方法大致可分为变分法和采样法两类. 变分法是将推理问题转化为目标函数(自由能)最小化问题,然后利用优化方法来求解最优化问题<sup>[19]</sup>. 采样法是使用从概率图所表示

收稿日期 2011-12-08 录用日期 2012-06-15  
Manuscript received December 8, 2011; accepted June 15, 2012  
国家重点基础研究发展计划(973 计划)(2009CB320602),国家自然科学基金(61071131),北京市自然科学基金(4122040)资助  
Supported by National Basic Research and Development Program of China (973 Program) (2009CB320602), National Natural Science Foundation of China (61071131), and Beijing Natural Science Foundation (4122040)  
本文责任编辑 张化光  
Recommended by Associate Editor ZHANG Hua-Guang  
1. 清华信息科学与技术国家实验室 清华大学自动化系 北京 100084  
1. Tsinghua National Laboratory for Information Science and Technology, Department of Automation, Tsinghua University, Beijing 100084

的联合分布中采样得到的样本来近似边缘概率或者最大概率状态<sup>[20]</sup>. 其中, 变分法是研究和应用较为广泛的一种推理方法. 目前, 变分推理的研究思路是从优化的角度, 特别是对偶优化的角度出发, 从目标函数、约束范围、优化方法三个方面研究高效的近似推理算法. “分解”思想是变分近似的灵魂. 在边缘概率问题的近似求解上, “分解”思想的主要表现是“以简代繁”, 即把多变量的联合概率使用局部变量的边缘概率的合理组合来近似. 在最大后验概率状态问题的近似求解上, “分解”思想的主要表现是“分而治之”, 即把原始复杂问题拆分成一个个简单且容易解决的子问题, 然后再把每个子问题的解整合成原始问题的近似解. 本文将系统地综述基于“分解”思想的变分近似推理方法.

本文的结构安排如下: 第 1 节简要介绍概率图模型及其推理问题; 第 2 节介绍概率图模型的精确推理方法, 这是变分近似的基础; 第 3 节归纳总结概率图模型变分近似推理的发展阶段; 第 4 节和第 5 节分别对边缘概率问题和最大后验概率问题的变分近似推理方法进行了系统综述. 最后, 第 6 节对概率图模型的变分近似推理方法及其应用前景进行了总结和展望.

## 1 概率图模型及其推理问题

### 1.1 概率图模型简介

概率图模型(或简称“图模型”)<sup>[1-2]</sup>是由图论和概率论结合而成的描述多元统计关系的有效模型. 一个概率图模型可以用  $G = (\mathcal{V}, \mathcal{E})$  表示, 其中  $\mathcal{V} = \{1, 2, \dots, N\}$  表示图上的结点集合,  $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$  表示图上的边集合. 每一个结点  $s \in \mathcal{V}$  表示一个变量  $X_s$ , 其状态空间记为  $\mathcal{X}_s$ , 状态空间  $\mathcal{X}_s$  既可以是连续的(例如  $\mathcal{X}_s = \mathbf{R}$ ), 也可以是离散的(例如  $\mathcal{X}_s = \{0, 1, \dots, m-1\}$ ), 离散状态空间概率图模型是研究和应用最为广泛的一种模型, 本文将主要介绍该类图模型上的变分近似推理方法.

常见的概率图模型包括贝叶斯网(Bayesian network)和马尔科夫随机场(Markov random field, MRF)<sup>[1-2]</sup>, 图 1 显示了这两种图模型的基本形式. 贝叶斯网和马尔科夫随机场中的边都表示两个变量之间的依赖关系, 它们分别通过条件概率表和势函数来定量的表示这种依赖关系. 贝叶斯网所表示的联合概率分布可表示为

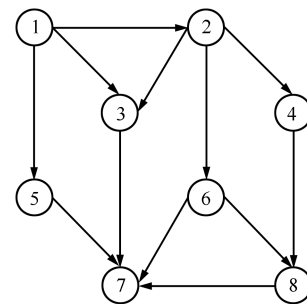
$$p(\mathbf{x}) = \prod_{s \in \mathcal{V}} p_s(x_s | x_{\pi(s)}) \quad (1)$$

其中,  $\mathbf{x} = \{x_1, x_2, \dots, x_m\}$ ,  $x_{\pi(s)}$  表示变量  $x_s$  的父

结点的集合. 马尔科夫随机场使用局部势函数乘积的形式表示联合概率:

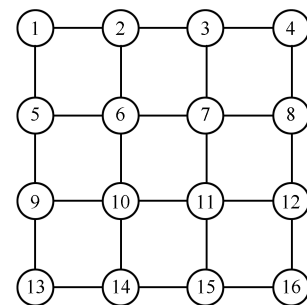
$$p(\mathbf{x}) = \frac{1}{Z_\psi} \prod_{\alpha \in S_\alpha} \psi_\alpha(\mathbf{x}_\alpha) \quad (2)$$

任何有向图模型都可以等价地转化为无向图模型<sup>[2]</sup>, “等价”的意思是指保持其联合概率分布和变量之间的条件独立关系不变. 此外, 在实际应用中, 马尔科夫随机场的表示范围比贝叶斯网的表示范围更加广泛. 因此, 在本文的后续介绍中, 如不特别强调, 概率图模型一般指马尔科夫随机场. 不过, 贝叶斯网能够蕴含比马尔科夫随机场更多的信息量, 例如贝叶斯网中的有向边可以表示两个变量之间的因果关系<sup>[21]</sup>, 而这种因果关系在马尔科夫随机场中无法表示.



(a) 贝叶斯网

(a) Bayesian network



(b) 马尔科夫随机场

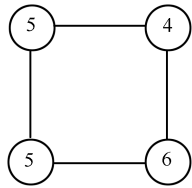
(b) Markov random field

图 1 概率图模型示例

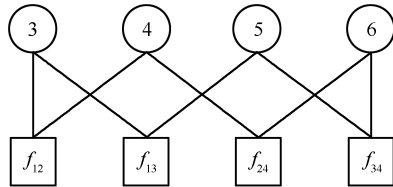
Fig. 1 Illustrations of graphical models

贝叶斯网和马尔科夫随机场都可以用因子图(Factor graph)来表示<sup>[14]</sup>. 图 2 显示了一个  $2 \times 2$  的成对马尔科夫随机场及其因子图. 概率图模型还有基于区域的表示方法, 例如团图(Cluster graph)<sup>[22]</sup>和连接树(Junction tree)<sup>[23]</sup>等. 虽然概率图模型的各种表示方法形式各异, 但是它们都利用因子分解的方式和条件独立性的假设紧凑的表示多个变量的联合分布, 具体表现为: 1) 图结构所表示的变量独立

性与变量之间马尔科夫性一致<sup>[21]</sup>; 2) 其表示的联合概率分布可以用局部函数的乘积来表示.



(a)  $2 \times 2$  的成对马尔科夫随机场  
(a) A  $2 \times 2$  pairwise MRF



(b) (a) 的因子图 (原结点用圆圈表示, 因子结点用方框表示)  
(b) The factor graph of (a) (Variables are denoted by cycles and factors are denoted by squares.)

图 2 因子图示例

Fig. 2 The illustration of a factor graph

如果式 (2) 中的势函数只包含是单个变量的势函数和被同一条边相连的两个变量的势函数, 那么该马尔科夫随机场又被称为成对马尔科夫随机场 (Pairwise MRF), 其表示的联合概率分布如下所示:

$$p(\mathbf{x}) = \frac{1}{Z_\psi} \prod_{s \in \mathcal{V}} \psi_s(x_s) \prod_{(s,t) \in \mathcal{E}} \psi_{st}(x_t, x_s) \quad (3)$$

成对马尔科夫随机场是较常用到的一种概率图模型, 此外, 其他马尔科夫随机场都可以等价地转化为成对马尔科夫随机场<sup>[1]</sup>.

## 1.2 概率推理问题

概率图模型的应用需要解决结构学习、参数学习和概率推理三个问题<sup>[2]</sup>. 结构学习是指在给定样本数据的情况下, 估计出最符合数据间依赖关系的概率图模型结构. 参数学习是指在给定概率图模型结构的情况下, 通过样本数据估计出各个变量之间定量的依赖关系, 如条件概率分布等. 概率推理是求取部分变量或者所有变量的边缘概率或者最大概率状态. 概率推理问题是概率图模型研究及应用的重要问题, 原因之一是概率图模型应用的最终目的是得到需要的推理结果, 这正在概率推理所关心的问题, 原因之二是在很多结构学习和参数学习算法中, 需要概率推理算法作为中间步骤. 概率推理问题包括计算部分变量的边缘概率、计算最大概率状态和

计算边缘化-最大概率状态等问题. 计算部分变量的边缘概率是求解如下问题:

$$p(\mathbf{x}_\beta) = \sum_{\mathbf{x} \setminus \mathbf{x}_\beta} p(\mathbf{x}) \quad (4)$$

计算最大概率状态 (Maximum a posteriori, MAP) 是求解如下问题:

$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} p(\mathbf{x}) \quad (5)$$

计算边缘化-最大概率状态 (Marginal-MAP) 是求解如下问题:

$$\mathbf{x}_\beta^* = \arg \max_{\mathbf{x}_\beta \in \mathcal{X}_\beta} \sum_{\mathbf{x} \setminus \mathbf{x}_\beta} p(\mathbf{x}) \quad (6)$$

上述问题中, 问题 (4) 和问题 (5) 是问题 (6) 较简单的特例, 计算最大概率状态问题是 NP-完全问题, 计算部分变量的边缘概率问题是 #P-完全问题, 计算边缘化-最大概率状态问题是 NP<sup>PP</sup>-完全问题<sup>[2, 24]</sup>. 由于大多数情况下精确求解这些问题是不可实现的, 我们只能采用近似方法来解决这些问题, 本文将从变分的角度综述在解决计算边缘概率问题和计算 MAP 问题上的最新研究成果. 由于针对计算边缘化-最大概率状态问题的研究较少, 本文不对其进行单独介绍, 感兴趣的读者可参考文献 [25-29], 其中, 文献 [27-29] 从变分角度阐述了边缘化-最大概率状态问题.

## 2 精确推理方法

虽然大部分概率图模型上的推理问题不可精确求解, 然而探索各种条件下的精确推理方法仍然是研究人员孜孜不倦的追求. 精确推理的研究思路主要是研究特殊图结构和特殊参数条件下的多项式级复杂度的精确推理方法, 目前, 能够进行精确推理的概率图模型有 (团) 树状概率图<sup>[21]</sup>、单环概率图<sup>[30]</sup>、变量状态为二值的概率图模型<sup>[31-33]</sup> 等.

### 2.1 边缘概率和联合概率的关系

概率图模型上边缘概率和联合概率之间的关系是树状图模型上精确推理的基础, 它们之间存在两种关系: 1) 边缘概率是联合概率的边缘化; 2) 联合概率可以用边缘概率的组合来表示. 边缘概率有两种不同的形式: 1) 基于求和的边缘概率 (和边缘概率, Sum-marginal); 2) 基于求最大的边缘概率 (最大边缘概率, Max-marginal). 基于求和的边缘概率定义为

$$p_s(\mathbf{x}_\beta) = \sum_{\mathbf{x} \setminus \mathbf{x}_\beta} p(\mathbf{x}) \quad (7)$$

基于求最大的边缘概率定义为

$$p_m(\mathbf{x}_\beta) = \max_{\mathbf{x} \setminus \mathbf{x}_\beta} p(\mathbf{x}) \quad (8)$$

由于和边缘概率与最大边缘概率在精确推理上的一致性,下面将以和边缘概率为主介绍精确推理方法,一般情况下,在计算最大边缘概率时,只需要将和边缘概率计算公式中的求和符号( $\sum$ )变为求最大符号( $\max$ )即可。

在树状概率图模型上,联合概率与边缘概率的深层联系是:联合概率可以表示为局部变量边缘概率的乘积的形式<sup>[34-35]</sup>,如下式所示:

$$p(\mathbf{x}) \propto \prod_{s \in \mathcal{V}} p_s(x_s) \prod_{(s,t) \in \mathcal{E}} \frac{p_{st}(x_s, x_t)}{p_s(x_s) p_t(x_t)} \quad (9)$$

上式中,  $p_s(x_s)$  表示变量  $x_s$  的边缘概率,  $p_{st}(x_s, x_t)$  表示变量  $\{x_s, x_t\}$  的联合边缘概率,式(9)中的边缘概率即可以是和边缘概率(式(7)),也可以是最大边缘概率(式(8)).联合概率与边缘概率这种深层的关系,其本质是树上概率的全局约束可以由部分变量概率的局部约束来实现,这个关系使得树状结构概率图模型上的精确推理高效可行,也为近似推理方法的研究提供了崭新思路。

## 2.2 精确推理方法概述

对于树状结构的概率图模型,其精确推理可以通过变量消元法(Variable elimination)<sup>[21]</sup>,以  $O(Nm^2)$  的复杂度来完成<sup>[36]</sup>,其中  $N$  是图模型中变量的数目,  $m$  是状态最多的变量的状态数目.计算和边缘概率的变量消元法又称为和积算法(Sum-product),计算最大边缘概率的变量消元法又称为最大积算法(Max-product),它们的原理都是分配率<sup>[36]</sup>.

计算部分变量的边缘概率本质上“先积后和”的形式,如式(4)所示,该形式的计算复杂度是指数级的.利用树状结构特点和分配率,可以把这种“先积后和”的计算形式转化为“先和后积”的计算形式,从而降低计算复杂度。

和积算法的另一种表示形式是信度传播算法(Belief propagation)<sup>[1, 21-22]</sup>,又被称为消息传递算法(Message passing),它用信息(Message)的形式表示和积算法中的变量消元过程,具有简洁一致的特点.信度传播算法与和积算法在本质上是相同的<sup>[14]</sup>.Yedidia 等证明了信度传播算法可以通过自由能(Free energy)最小化来解释<sup>[22]</sup>.概率图模型中所研究的自由能实质是一种距离的度量,即两个联合概率分布之间的距离度量,这种距离度量可由库

尔贝克-莱布勒散度(Kullback-Leibler divergence, KL 散度)定义和推导<sup>[37]</sup>.两个联合概率分布之间的 KL 散度定义为

$$KL(q||p) = \sum_{\mathbf{x}} q(\mathbf{x}) \log \frac{q(\mathbf{x})}{p(\mathbf{x})} \quad (10)$$

利用拉格朗日乘子法最小化  $p(\mathbf{x})$  和  $q(\mathbf{x})$  之间的 KL 散度即可得到信度传播算法.从自由能优化角度解释信度传播算法为从优化角度研究近似推理方法奠定了基础。

对于非树状概率图模型,直接运用信度传播算法并不能得到精确解,但是可以把非树状图模型转化为团树(又称为邻接树)模型,然后利用团树上的信度传播算法<sup>[22-23, 38]</sup>得到原始图模型的精确推理结果,团树上的信度传播算法的复杂度是团树最大团中变量数目的指数级.Cheng 等提出了扩展外平面图上的迭代和积算法,该算法不需要构造团树,可精确求解树的宽度为 2 的图模型的推理问题<sup>[39]</sup>.此外,部分具有特殊结构或者特殊参数的概率图模型也存在低复杂度的精确推理算法.Weiss 证明了在单环图模型上运行信度传播算法可以得到最大概率状态的精确结果<sup>[30]</sup>.Kolmogorov 等证明了对于次模(Submodular)自由能,TRW(Tree-reweighted message-passing)算法<sup>[35]</sup>可以获得精确的最大概率状态<sup>[31]</sup>.图割法也能在次模性自由能上获得精确解,并且计算速度比 TRW 方法快<sup>[40]</sup>.Kolmogorov 等证明了对于不具有次模性的自由能,其精确最小化是 NP 难问题<sup>[41]</sup>.针对配分函数的求解问题,完美匹配法(Perfect matching)可以精确的求解二值无单变量势函数的平面图的配分函数<sup>[32, 42]</sup>.

虽然精确推理在一些特殊的概率图模型上是可行的,但是,在实际应用中遇到的大部分概率图模型的推理问题是无法多项式时间精确求解的,近似推理方法是解决这些问题的唯一途径,下面将介绍概率图模型上近似推理的主要思路及解决方法。

## 3 变分近似推理的发展阶段

本节主要总结变分近推理的发展阶段,后续章节将重点介绍变分近推理的最新研究成果.变分法的思路是首先将推理问题转化为优化问题,然后利用优化理论来解决推理问题.在解决优化问题的环节使用近似方法,便可以得到原始推理的近似解.根据变分近似推理在不同阶段的研究思路的不同,我们将其大致分为三个研究阶段:信度传播阶段、自由能优化阶段和对偶优化阶段。

1) 第一阶段:信度传播.信度传播算法始于

1990 年左右, 是 Pearl 教授等针对树状贝叶斯网提出的精确推理算法<sup>[21]</sup>, 它也可以被用在非树结构的概率图模型上而得到近似推理结果. 这个阶段的近似推理方法与精确推理方法密切联系, 基本思路是首先通过引入部分变量之间条件独立性的假设, 然后利用精确推理方法得到在不同假设情况下的近似结果. 该阶段的近似推理方法包括: 剪枝算法 (Pruning algorithm)<sup>[43]</sup>、定界条件法 (Bounded conditioning method)<sup>[44-45]</sup> 等. 由于主要使用精确推理方法, 这些方法面临着指数级计算复杂度的问题. 此外, 这些方法多含启发式成分, 近似结果缺乏理论保证.

2) 第二阶段: 自由能优化. 自由能优化法始于 2000 年左右, Yedidia 等从变分法的角度阐述了信度传播算法的内在机理, 证明了无论是树状图模型还是有环图模型上的信度传播算法都是自由能最小化问题的一种解法<sup>[22]</sup>. 之后, 学者们通过构造各种不同形式自由能来提出不同形式的信度传播算法<sup>[46-50]</sup>.

3) 第三阶段: 对偶优化. 对偶优化始于 2005 年左右, 其基本思路是将原始的复杂问题分解成一个个简单的子问题的组合, 然后对每个子问题分别求解, 最后将每个子问题的解组合成原始问题的近似解. Wainwright 等首先创新性地提出了基于树分解和重参数化的近似推理方法 (TRW)<sup>[35]</sup>, 后来该方法被 Kolmogorov 等推广和发展<sup>[40]</sup>, Komodakis 和 Sontag 等明确提出了对偶分解的框架<sup>[51-52]</sup>.

## 4 边缘概率问题的变分近似推理

边缘概率包括和边缘概率 (式 (7)) 与最大边缘概率 (式 (8)), 由于这两个问题在求解上的相似性, 本节以和边缘概率为主介绍概率图模型上边缘概率的近似求解方法, 将和边缘概率计算公式中的求和符号 ( $\sum$ ) 变为求最大符号 (max) 即可近似求解最大边缘概率.

### 4.1 边缘概率问题的变分近似思路

基于变分法的边缘概率的近似求解主要采用“分解”的思路, 这种“分解”思路的具体表现是利用部分变量的边缘概率的组合来近似所有变量的联合概率分布, 本节将详细介绍变分框架下这种分解思路的具体实现.

基于分解的边缘概率的近似求解包括对联合概率的分解、对联合熵的分解、对自由能的分解等, 其中, 后两种分解形式都可以从联合概率分布的分解推导出. 联合概率分解近似的基本思路是用部分变量的边缘概率的合理组合表示所有变量的联合概率,

具体表示如下:

$$q(\mathbf{x}) = \prod_{s \in \mathcal{V}} q_s(x_s)^{\rho_s} \prod_{(s,t) \in \mathcal{E}} \left( \frac{q_{st}(x_{st})}{q_s(x_s) q_t(x_t)} \right)^{\rho_{st}} \quad (11)$$

使用变分方法计算边缘概率是基于自由能最小化方法, 其自由能如式 (10) 所示, 其中  $p(\mathbf{x})$  如式 (3) 所示,  $q(\mathbf{x})$  如式 (11) 所示. 对于边缘概率问题的自由能, 有一种更为广义的定义, 即  $\alpha$  散度, 其定义为

$$D_\alpha(p||q) = \frac{\int \alpha p + (1 - \alpha) q - p^\alpha q^{1-\alpha} dx}{\alpha(1 - \alpha)} \quad (12)$$

其中,  $\alpha \in (-\infty, +\infty)$ ,  $KL(q||p)$  是  $\alpha \rightarrow 0$  时的  $\alpha$  散度,  $KL(p||q)$  是当  $\alpha \rightarrow 1$  时的  $\alpha$  散度.

基于自由能最小化的框架, 通过变换式 (11) 中的参数  $\rho$  ( $\rho = \{\rho_s, \forall s \in \mathcal{V}, \rho_{st}, \forall (s, t) \in \mathcal{E}\}$ ) 和式 (12) 中的参数  $\alpha$ , 最小化下面的问题可以得到各种不同的近似方法.

$$\begin{aligned} & \min D_\alpha(p||q) \\ \text{s.t.} & \sum_{x_t} q_{st}(x_{st}) = q_s(x_s) \\ & \sum_{x_s} q_s(x_s) = 1 \end{aligned} \quad (13)$$

### 4.2 边缘概率问题的变分近似算法

本节将根据边缘概率问题的近似思路, 具体介绍不同参数条件下的近似算法及其性质.

#### 4.2.1 近似算法

基于不同的参数  $\rho$  和  $\alpha$ , 通过优化问题 (13) 可以得到不同的近似推理算法.  $\alpha \rightarrow 0$  时的  $\alpha$  散度是研究和最为广泛的一种自由能, 即  $D_{\alpha \rightarrow 0}(p||q) = KL(q||p)$ . 当  $\alpha \rightarrow 0$ , 且  $\rho$  固定时, 使用拉格朗日乘子法求解问题 (13) 可以得到如下的信度传播算法:

$$M_{ts}^{\frac{\rho_s}{\rho_{ts}}} = \sum_{x_t} \left( \psi_{st;jk}(x_s, x_t)^{\frac{1}{\rho_{st}}} \psi_{t;k}(x_t)^{\frac{1}{\rho_t}} \frac{\prod_{u \in \Gamma(t) \setminus s} M_{ut}}{M_{st}^{\frac{\rho_{st}}{\rho_{st}} - 1}} \right) \quad (14)$$

边缘概率可由迭代定点处的信息来表示, 即

$$q_s(x_s) \propto \psi_{s;j}(x_s)^{\frac{1}{\rho_s}} \prod_{u \in \Gamma(s)} M_{us}(x_s) \quad (15)$$

朴素平均场法 (Naive mean field)<sup>[53]</sup> 使用全分解的方式来近似原始的联合概率, 即  $\rho_{st} = 0$  和  $\rho_s = 1$  (当  $\rho_{st} = 0$ , 式 (14) 不再适用, 其有更加简洁的迭代计算方法<sup>[53]</sup>). 基于平均场思路的近似方法还有结构平均场近似 (Structure mean field, SMF)<sup>[54]</sup> 和树状平均场近似<sup>[55-56]</sup> 等方法, TAP 修正方法 (TAP correction method)<sup>[53, 57]</sup> 和线性响应方法 (Linear response method)<sup>[58]</sup> 在平均场基础上进行扩展或优化而得到更加精确的近似解. 信度传播算法 (Loopy belief propagation, LBP)<sup>[21-22]</sup> 使用 Bethe 自由能分解的方法来近似  $p(\mathbf{x})$ , 即  $\rho_{st} = 1$  和  $\rho_s = 1$ . Wainwright 等提出的权树分解的近似方法是一种凸函数近似, 其相当于  $0 < \rho_{st} \leq 1$  和  $\rho_s = 1$ <sup>[48]</sup>. Wiegerinck 等提出一种分数信度传播方法 (Fractional belief propagation, FBP)<sup>[46]</sup>, 该方法固定  $\rho_s = 1$ , 然后使用线性响应理论 (Linear response theory) 来优化  $\rho_{st}$ , 从而得到经过优化的近似结果. Meshi 等研究了参数  $\rho$  和自由能凸性之间的关系, 通过选择合适的参数构造凸自由能而得到收敛的信度传播算法<sup>[49]</sup>.

当  $\alpha \rightarrow 1$ , 其  $\alpha$  散度为  $KL(p||q)$ , 此时的自由能实质是  $\sum_{\mathbf{x}} p(\mathbf{x}) \log q(\mathbf{x})$ . Minka 在这方面进行了较多的研究, 提出了期望传播算法 (Expectation propagation, EP) 和幂期望传播算法 (Power expectation propagation, PEP) 等近似推理方法<sup>[59-61]</sup>.

当  $\alpha \rightarrow 0$  和  $\alpha \rightarrow 1$  时,  $\alpha$  散度可以被表示为对数的形式, 这种表示形式比较简单, 并且容易优化. 然而, 当  $\alpha$  取其他值时, 这种简化的表示形式不再存在, 优化此时的  $\alpha$  散度比较繁琐. 所以, 对  $\alpha$  取其他值时的  $\alpha$  散度的研究比较少. 图 3 归纳了边缘概率问题的近似方法.

### 4.2.2 算法性质

算法的近似效果与自由能的凸性、算法的收敛性密切相关, 研究人员都希望构造凸的自由能函数来得到比较好的近似效果. Heskes 提出了基于线性约束的凸自由能的充分条件<sup>[47, 50]</sup>, Pakzad 等提出了与之等价且更加简洁的充分条件<sup>[62-63]</sup>. Meshi 等提出了凸化 Bethe 自由能的方法, 使用其参数与 Bethe 自由能参数距离最小的凸自由能来代替 Bethe 自由能, 即求解如下优化问题<sup>[49]</sup>:

$$\min_{\rho} \|\rho - \rho(\text{Bethe})\|$$

其中参数  $\rho$  满足凸自由能的条件, 该方法的效果与 TRW 方法的近似效果相似.

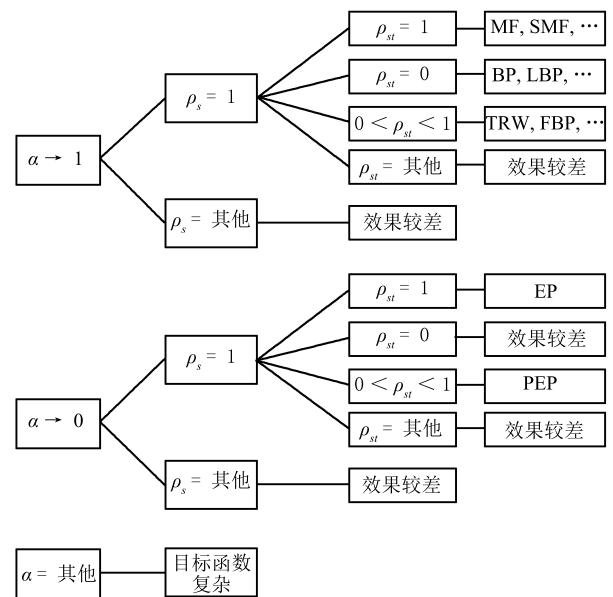


图 3 边缘概率近似方法的归纳  
Fig. 3 The generalization of approximate methods for computing marginals

算法收敛性是影响迭代算法效果优劣的一个重要因素, 然而即使是凸自由能, 其迭代算法也不一定收敛. Tatikonda 等证明了当计算树上 Gibbs 采样算法收敛时, LBP 算法同样收敛, 并给出了 LBP 算法收敛的充分条件<sup>[64]</sup>. Ihler 等研究了信度传播算法的误差, 并给出了比文献 [64] 更强的收敛性条件<sup>[65]</sup>. Heskes 提出了双环算法 (Double-loop algorithm) 来保证解的收敛性<sup>[50]</sup>, 该方法是期望最大化算法 (Expectation-maximization, EM) 的具体实现. 陈峰等提出了递归信度传播算法来解决传统信度传播算法在贝叶斯网中的收敛性问题<sup>[66]</sup>. Meltzer 等提出了收敛的信度传播算法的统一框架—树一致性优化方法 (Tree-consistency bound optimization, TCBO)<sup>[67]</sup>, 该方法可以归纳文献 [40, 50, 68-69] 中的近似推理算法. Hazan 等通过构造凸自由能提出了两种收敛的类似于信度传播的算法<sup>[70]</sup>, 他们把自由能表示为了凸函数之和的形式. 虽然 Bethe 自由能是非凸的, 但由于其近似效果较好, 很多学者研究其最小化的收敛性算法, 例如凹凸过程算法 (Concave-convex procedure, CCCP)<sup>[71]</sup> 和条件概率算法 (Conditioned probability, CP)<sup>[72]</sup>.

上述各种算法的近似效果与概率图模型的结构与参数关系很大. 一般而言, 当参数  $\rho$  满足“有效 (Valid)”条件时<sup>[49]</sup>, 近似效果要比参数  $\rho$  无效时的近似效果好. 相同参数条件下, 相比非凸自由能和非收敛算法, 凸自由能和收敛算法会得到比较好的近似效果.

### 5 MAP 问题变分近似推理方法

前面介绍了边缘概率问题的近似求解方法, 包括和边缘概率与最大边缘概率. 本节将介绍所有变量的最大概率状态 (MAP) 问题的近似思路和方法, 下面首先介绍 MAP 问题与上节的边缘概率问题的关系.

#### 5.1 MAP 问题与边缘概率问题的关系

虽然 MAP 问题与边缘概率问题的近似思路不同, 但它们之间存在着密切的联系, 即可以在一个统一的框架下表示它们.

令  $\theta_t(x_t) = -\log \psi_t(x_t)$ ,  $\theta_{ts}(x_t, x_s) = -\log \psi_{ts}(x_t, x_s)$ , 成对马尔科夫随机场的 MAP 问题可以表示为

$$\min_{\mathbf{x} \in \mathcal{X}} E(\boldsymbol{\theta}, \mathbf{x}) = \langle \boldsymbol{\theta}, \phi(\mathbf{x}) \rangle = \sum_{s \in \mathcal{V}} \boldsymbol{\theta}_s \cdot \mathbf{x}_s + \sum_{(s,t) \in \mathcal{E}} \boldsymbol{\theta}_{st} \cdot \mathbf{x}_{st} \quad (16)$$

其中,  $E(\boldsymbol{\theta}, \mathbf{x})$  称为自由能.

计算边缘概率的自由能和计算 MAP 问题的自由能并不相同, 即  $KL(q||p)$  和  $E(\boldsymbol{\theta}, \mathbf{x})$  不相同, 它们相差了联合概率分布的熵  $H(q(\mathbf{x}))$ <sup>[38]</sup>, 即

$$KL(q||p) = - \sum_{\mathbf{x}} q(\mathbf{x}) \log p(\mathbf{x}) + \sum_{\mathbf{x}} q(\mathbf{x}) \log q(\mathbf{x}) = \langle \boldsymbol{\theta}, q(\mathbf{x}) \rangle - H(q(\mathbf{x})) \quad (17)$$

其中,  $\langle \boldsymbol{\theta}, q(\mathbf{x}) \rangle$  和  $\langle \boldsymbol{\theta}, \phi(\mathbf{x}) \rangle$  是等价的. Meltzer 等通过引入参数  $T$ , 把这两种自由能统一表示为<sup>[67]</sup>

$$F(q(\mathbf{x})) = \langle \boldsymbol{\theta}, q(\mathbf{x}) \rangle - TH(q(\mathbf{x})) \quad (18)$$

当  $T = 1$  时,  $F(q(\mathbf{x}))$  是边缘概率问题的自由能, 当  $T = 0$  时,  $F(q(\mathbf{x}))$  是 MAP 问题的自由能, 这建立了两种自由能之间的联系. 在边缘概率问题的求解算法中, 对  $T$  取极限, 令  $T \rightarrow 0$ , 即可得到 MAP 问题的近似推理算法. 这可以解释前面提到的把和边缘概率计算公式中的求和符号 ( $\sum$ ) 换为求最大符号 ( $\max$ ) 即可得到 MAP 问题的近似推理算法. 接下来, 我们介绍针对 MAP 问题具体特点的近似思路和算法.

#### 5.2 MAP 问题的变分近似思路

问题 (16) 是个组合优化或者整数规划问题, 其精确求解是 NP 难的<sup>[2, 73]</sup>, 目前对该问题主要有两种近似思路: 线性规划松弛和对偶分解.

#### 5.2.1 线性规划松弛

线性规划松弛 (Linear programming relaxation, LPR) 方法是对问题 (16) 的约束域直接进行近似<sup>[1, 74]</sup>:

$$\min_{\boldsymbol{\mu} \in \mathcal{L}(G)} LPR(\boldsymbol{\mu}) = \sum_{s \in \mathcal{V}} \boldsymbol{\theta}_s \cdot \boldsymbol{\mu}_s(x_s) + \sum_{(s,t) \in \mathcal{E}} \boldsymbol{\theta}_{st} \cdot \boldsymbol{\mu}_{st}(x_s, x_t) \quad (19)$$

其中变量  $\boldsymbol{\mu}$  的定义域  $\mathcal{L}(G)$  是线性规划松弛后的局部一致性约束. 图 4 显示了全局一致性约束和局部一致性约束的关系. 局部一致性约束的强弱与约束中最大团 (变量集合) 的大小有关, 通过不断增加团的大小并利用团的一致性约束就可以得到一系列越来越紧的约束. 如果最大团中含有的变量数目为  $t + 1$ , 那么该局部约束称为  $t$  阶局部约束, 表示为  $\mathcal{L}_t(G)$ . 不同阶次约束之间的关系为

$$\mathcal{L}_1(G) \supseteq \mathcal{L}_2(G) \supseteq \mathcal{L}_3(G) \supseteq \dots \supseteq \mathcal{M}(G) \quad (20)$$

最紧的约束是团中包括所有变量, 即全局一致性约束. 对 (团) 树的宽度为  $t$  的任意概率图,  $t$  阶约束  $\mathcal{L}_t(G)$  即可保证其精确解<sup>[1]</sup>. (团) 树的宽度定义为一个树状图模型中最大团所含有的变量数目减 1, 普通树的宽度为 1. 线性规划松弛后的问题可以用线性规划理论中的方法来解决, 例如单纯形法 (Simplex method)<sup>[17]</sup>、内点法 (Interior point methods)<sup>[75-76]</sup> 和椭球法 (Ellipsoid method)<sup>[77]</sup>. 然而, 线性规划松弛方法并没有充分利用图的结构性质, 对于大规模的概率图模型推理问题, 特别对于实际应用问题, 不能快速地得到原始问题的近似解.

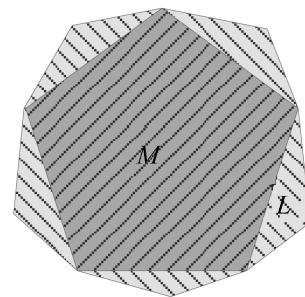


图 4 全局一致性约束和局部一致性约束示意图 (中间深色区域  $M$  是凸的全局一致性约束; 周围浅色区域  $L$  和中间深色区域  $M$  共同表示全局约束的松弛, 即局部一致性约束)  
Fig. 4 The illustration of the marginal polytope and the local consistent constraints (The dark region  $M$  denotes the convex global constraint, and the combination of the light and dark regions  $L$  and  $M$  denotes the local consistent constraint.)

除线性规划松弛外, MAP 问题还可以用二次规划松弛来近似, 二次规划松弛的最优值与 MAP 问题的最优值相同, 然而, 二次规划松弛需要解决目标函数非凸的问题, 相关研究工作见文献 [78–81].

### 5.2.2 对偶分解

对偶分解是 MAP 问题近似推理研究和应用最多的近似方法. 其基本思路是首先把原始复杂的问题分解成一个个容易求解的子问题, 然后对子问题单独求解, 最后把所有子问题的解合成为原始问题的一个近似解. 下面的不等式总是成立:

$$\min_{\mathbf{x}} E(\boldsymbol{\theta}, \mathbf{x}) = \min_{\mathbf{x}, \{\mathbf{x}^{C_i}\}} \sum_{C_i \in \mathcal{C}(G)} E(\boldsymbol{\theta}^{C_i}, \mathbf{x}^{C_i}) \geq \sum_{C_i \in \mathcal{C}(G)} \min_{\mathbf{x}^{C_i}} E(\boldsymbol{\theta}^{C_i}, \mathbf{x}^{C_i}) \quad (21)$$

其中,  $\mathcal{C}(G)$  是覆盖原图  $G$  的一系列子图的集合,  $E(\boldsymbol{\theta}^{C_i}, \mathbf{x}^{C_i})$  是定义在子图  $C_i$  上的 MRF 自由能. 令

$$F(\boldsymbol{\theta}^C) = \sum_{C_i \in \mathcal{C}(G)} \min_{\mathbf{x}^{C_i}} E(\boldsymbol{\theta}^{C_i}, \mathbf{x}^{C_i})$$

那么  $F(\boldsymbol{\theta}^C)$  是原图 MRF 自由能的一个下界, 称为对偶函数. 对偶分解的目标是最大化  $F(\boldsymbol{\theta}^C)$ , 以获得最紧的下界, 即

$$\begin{aligned} \max F(\boldsymbol{\theta}^C) &= \sum_{C_i \in \mathcal{C}(G)} \min_{\mathbf{x}^{C_i}} E(\boldsymbol{\theta}^{C_i}, \mathbf{x}^{C_i}) \\ \text{s.t. } \boldsymbol{\theta}^C &\in \Theta^C \end{aligned} \quad (22)$$

上式中,  $\Theta^C$  是参数  $\boldsymbol{\theta}^C$  的约束域, 定义为

$$\Theta^C = \left\{ \boldsymbol{\theta}^C \mid \sum_{C_i \in \mathcal{C}(G)} \boldsymbol{\theta}_s^{C_i} = \bar{\boldsymbol{\theta}}_s, \sum_{C_i \in \mathcal{C}(G)} \boldsymbol{\theta}_{st}^{C_i} = \bar{\boldsymbol{\theta}}_{st} \right\} \quad (23)$$

直观理解, 该约束要求分解之后所有子问题的参数之和等于原始图模型的参数. 满足  $\boldsymbol{\theta}^C \in \Theta^C$  的参数被称为原始参数  $\bar{\boldsymbol{\theta}}$  的重参数化 (Reparameterization)<sup>[40]</sup> 或等价转化 (Equivalent transformation)<sup>[82]</sup>.

对偶分解的优化问题也可以直接对原始问题使用拉格朗日松弛法 (Lagrangian relaxation) 推导出<sup>[52]</sup>, 该方法通过引入拉格朗日乘子 (Lagrange multipliers) 将有约束的优化问题变为无约束的优化问题. 至此, 我们给出了概率图模型 MAP 问题的求解思路, 即先从对偶角度近似求解推理问题, 然后从线性规划松弛的角度研究近似的精度. 对偶问题和线性规划松弛的关系如图 5 所示. 对于树状结构

的概率图模型, LP 松弛后的界依然是紧的; 对于有环模型, LP 松弛后的界一般比原始问题的真实解小. 但是, 对于具有特殊参数概率图模型, 例如二值次模的自由能优化问题, LP 松弛后的最优值等于原始问题的最优值, 即其松弛的界是紧的 (Tight)<sup>[83]</sup>.

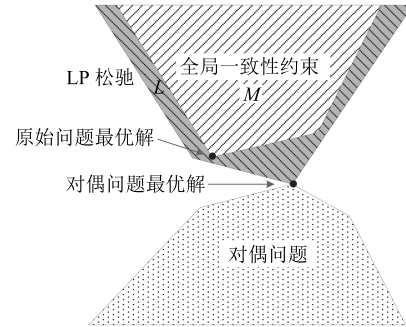


图 5 原始问题、LP 松弛以及对偶问题的关系 (其中 LP 松弛得到原始问题的一个下界, 对偶问题与 LP 松弛的最优解相等)

Fig. 5 Relations between primal problem, LP relaxation and dual problem (The LP relaxation is a lower bound of the original problem, and optimal values of the LP relaxation and its dual problem are equal.)

### 5.3 MAP 问题的变分近似算法

在对偶分解的框架下, MAP 推理问题需要解决如下两个问题: 1) 如何分解原始问题; 2) 如何优化对偶函数. 下面将具体介绍这两个问题的解决方法.

#### 5.3.1 分解方法

在分解环节, 主要是把原始问题分解为易于求解的子问题的组合, 目前存在的分解方式有: 把每个结点和每条边分别作为一个子问题<sup>[68–69, 82, 84]</sup>, 把每棵子树作为一个子问题<sup>[35, 40, 51, 85]</sup>, 把二值次模概率图作为一个子问题<sup>[51, 56, 86]</sup>, 把二值无单变量势函数的平面概率图作为一个子问题<sup>[42]</sup>, 把每个外平面图 (Outer-planar graph) 作为一个子问题<sup>[87]</sup>. 此外, Yarkony 等使用一棵覆盖树 (Covering tree) 来近似原来的有环图模型, 并证明了这种分解方法与子树分解的等价性<sup>[88]</sup>.

单一的分解方式并不能获得很好的近似效果, 因此需要研究多种分解方式融合的方法, 主要解决思路是在树状图分解的基础上增加的高阶约束的分解及优化. Sontag 等提出了环一致性约束 (Cycle relaxation) 的高阶约束增加方法<sup>[89–90]</sup>, 他们将二阶局部一致性约束  $\mathcal{L}_2(G)$  中联合分布  $\tau_{ijk}(x_i, x_j, x_k)$  中的三个变量限制在概率图的同一个环 (cycle) 上, 该方法是二阶局部一致性约束的松弛. Werner 提



出了增加团的约束来收紧成对一致性局部约束的方法<sup>[82]</sup>; Komodakis 等提出了一种通过修复环操作来实现收紧约束的算法<sup>[91]</sup>; Schraudolph 等利用最大权重完美匹配 (Maximum-weight perfect matching) 和 Edmond 的敛簇算法 (Blossom-shrinking algorithm) 来解决 MAP 近似推理问题<sup>[32]</sup>; Globerson 等使用加权平面图分解的方法来近似二值非平面图的配分函数<sup>[42]</sup>; Schraudolph 提出了类似的方法来求解最大可能的状态<sup>[33]</sup>;

对偶分解是对原问题的松弛, 强对偶在一般情况下并不成立, 然而, 下面的定理<sup>[35, 52, 83, 92]</sup> 提供了分解后强对偶成立的充分条件:

**定理 1.** 如果存在一个解  $\tilde{\mathbf{x}}$ , 其中, 对每一个子问题  $E(\theta^{C_i}, \mathbf{x}^{C_i})$ ,  $\tilde{\mathbf{x}}^{C_i} \in \arg \min_{\mathbf{x}^{C_i}} E(\theta^{C_i}, \mathbf{x}^{C_i})$ , 则强对偶成立, 即  $MAP(\theta) = \sum_{C_i \in \mathcal{C}(G)} \min_{\mathbf{x}^{C_i}} E(\theta^{C_i}, \mathbf{x}^{C_i})$ .

### 5.3.2 近似算法

对偶分解得到的对偶函数  $F(\theta^C)$  是非可微凸函数, 因此问题 (21) 是个非光滑凸优化问题. 结合具体的分解方式, 该问题可以使用块坐标下降算法 (Block coordinate descent method, BCD) 和次梯度算法 (Subgradient method) 等方法求解.

块坐标下降法通过依次优化对偶函数的部分子问题 (其他子问题保持不变) 而使得目标函数值不断单调下降, 该算法的应用需要解决如何选择部分子问题进行优化的问题. 利用块坐标下降方法求解对偶问题的算法主要有: 最大-和扩散算法 (Max-sum diffusion algorithm, MSD)<sup>[69]</sup>、最大积线性规划算法 (Max product linear programming, MPLP)<sup>[68, 90]</sup>、树块坐标下降算法 (Tree block coordinate descent algorithm, TBCD)<sup>[85]</sup> 和顺序树重参数化算法 (TRW-S)<sup>[40]</sup>. 这些算法的不同之处在于它们把不同的子问题作为块 (Block) 来进行优化, MSD 算法使用了一个边子问题和与该边相连的一个结点子问题进行优化, MPLP 算法使用了一个边子问题和与该边相连的所有子问题进行优化, TBCD 算法使用了概率图模型一棵生成树上的所有子问题进行优化, TRW-S 算法把许多子树作为一个块, 但每次只优化这些子树共同覆盖的一个结点或者一条边的参数. Meltzer 等总结了各种块坐标下降算法的共同点, 提出了该类算法的统一框架—树一致性界优化算法 (Tree-consistency bound optimization, TCBO)<sup>[67]</sup>. BCD 算法的优点是容易实现及收敛速度快, 然而, 由于对偶函数的非光滑性, BCD 算法可能会收敛到某个次优解 (Get stuck). 不过, 对二值状态的成对 MRF, BCD 算法的不动点即为对偶问题的全局最优解<sup>[31, 68]</sup>. 对非光滑函数进

行光滑处理可以使得 BCD 算法突破次优解, 收敛到全局最优解<sup>[93-94]</sup>. 此外, 还可以使用非光滑优化的竞价算法 (Auction algorithm)<sup>[95]</sup> 等来保证全局最优解.

次梯度法是非光滑优化的另一种常用算法, 它与光滑优化的梯度下降法相似, 是用次梯度来代替原来的梯度<sup>[51]</sup>. 次梯度法的应用需要解决的问题是次梯度的求取及向定义域的投影, MAP 问题的对偶问题使用次梯度法求解的优势是次梯度比较容易的获得, 因为每个子问题的最大概率状态  $\mathbf{x}^*$  即为其次梯度<sup>[51]</sup>. 基于此优势, 次梯度法的应用范围比块坐标下降法更加广泛, 块坐标下降法主要用于树分解的对偶问题, 而次梯度法可以用于任意方式分解的对偶问题. Komodakis 等对次梯度法在概率图模型 MAP 推理问题上的应用进行了系统地阐述<sup>[51]</sup>, Batra 等把原始问题分解为外平面图的组合, 并用次梯度法来求解对偶问题<sup>[87]</sup>. Jancsary 等把原始图模型分解成边和点的集合, 然后利用增益次梯度方法 (Incremental subgradient algorithm) 求解对偶问题, 能得到较快的收敛速度<sup>[84]</sup>. 次梯度法在理论上保证了对偶问题的解的全局最优性, 然而与块坐标下降法相比, 它的缺点是收敛速度比较慢, 并缺少停机准则. 增加次梯度法收敛速度的一种方法是对原非光滑函数进行光滑化处理, 使非光滑优化问题变为光滑优化问题<sup>[96-97]</sup>, 该方法可以把算法的时间复杂度由  $O(1/\varepsilon^2)$  降低到  $O(1/\varepsilon)$  ( $\varepsilon$  为结果精度)<sup>[98-99]</sup>.

除了从对偶问题求解之外, 还可以从原始 LP 松弛的角度进行求解. Sontag 提出了增加  $k$  值环不等式 ( $k$ -ary cycle inequality) 来获得更紧的约束, 并利用 LP 规划算法求解<sup>[74]</sup>. Ravikumar 等提出了基于内环迭代和外环迭代的双环算法 (Double loop algorithm) 来保证全局最优性<sup>[100]</sup>. Hazan 等通过对原始变量 (Primal variables) 和对偶变量 (Dual variables) 进行迭代求解, 以找到问题的一个鞍点 (Saddle point)<sup>[94]</sup>. 最近, Martins 等提出利用乘子变方向法 (Alternating direction method of multipliers, ADMM) 来求解原始 LP 松弛问题<sup>[101]</sup>, 该方法比次梯度法收敛快, 且能保证全局最优解. Meshi 等利用乘子变方向法直接求解对偶问题, 获得了比求解原始问题收敛更快的近似算法<sup>[102]</sup>. 此外, 还有学者从约束规划 (Constraint programming)<sup>[103]</sup> 的角度研究 MAP 问题的近似求解方法, 例如分枝定界法 (Branch and bound)<sup>[104-106]</sup>.

### 5.3.3 近似效果

评价近似推理算法效果的测试数据集包括合成

数据和实际数据,合成数据以网格图 ( $n \times m$ ) 和全连通图为基本结构,利用高斯分布或均匀分布随机生成图模型的参数,通过比较多次近似结果的平均值来比较各种近似算法的优劣. 实际数据主要包括机器视觉数据<sup>1</sup>和计算生物学数据<sup>2</sup>. 总体而言,对于不同参数的合成测试数据,近似算法的近似效果差别较大;对于实际数据,近似算法往往能取得比较理想的近似结果. 例如在立体匹配数据集上,基于树分解 ( $\mathcal{L}_1(G)$ ) 的近似算法能够得到误差低于 1% 的近似效果<sup>[40, 51]</sup>. 通过增加高阶约束及利用团树方法,可以获得该类问题的精确解(时间复杂度为小时量级)<sup>[107]</sup>. 在光流估计、多目标标记等机器视觉问题上,基于树分解和外平面图分解 ( $\subseteq \mathcal{L}_2(G)$ ) 的算法能够取得比较理想的效果<sup>[51, 87]</sup>. 计算生物学数据的测试包括蛋白质结构设计和侧链预测等问题,针对 97 个蛋白质结构设计测试数据集,基于树分解的近似算法可以精确求解其中 2 个问题<sup>[17]</sup>,利用增加三元组约束 ( $\supseteq \mathcal{L}_2(G)$ ) 的方法,可以得到其中 96 个问题的精确解<sup>[90]</sup>. 针对侧链预测问题,基于树分解的算法可以精确求解 90% 以上的问题<sup>[17]</sup>,利用增加三元组约束可以精确求解全部问题 (30/30)<sup>[90]</sup>.

## 6 结论与展望

概率图模型在很多实际问题上都表现出简约而紧凑的表示能力,概率推理是概率图模型研究及应用的核心问题,由于实际应用中绝大部分概率图模型的推理问题无法精确求解,近似方法是解决这些概率推理问题的唯一途径. 本文基于“分解”的思路,系统地总结了最近几年概率图模型变分近似推理的研究进展. 虽然近年来概率图模型变分近似推理的研究取得了实质性的进展,然而更加高效准确的近似推理方法以及针对具体问题的快速推理方法仍然是概率图模型研究和应用的热点和难点问题. 我们将从理论研究和具体应用等方面对概率图模型近似推理的发展前景稍作介绍.

1) 高阶约束的选择问题. 约束是限制近似推理精度及复杂度的关键因素,基于低阶约束的推理算法虽然容易实现,但近似精度欠缺. 增加高阶约束是提高近似精度的唯一方法,然而,高阶约束存在数量众多、计算复杂度高、冗余性等问题,目前尚没有得到很好的解决. 因此,最优的高阶约束的选择问题是目前研究的难点和热点问题.

2) 边缘化-最大概率状态问题. 概率图模型上的边缘化-最大概率状态问题是图模型推理中的难

点问题,目前国内外对该问题尚缺少系统的研究. 研究基于变分法的边缘化-最大概率状态问题的近似解法,提出针对该问题的单调收敛的消息传递算法以及分析算法的各种性质是目前概率图模型推理领域一个重要的研究方向.

3) 近似精度的理论分析问题. 目前不同算法近似效果的比较仍然停留在实验比较层面,从理论上分析各种近似算法的近似精度研究较少. 不同算法的近似效果受模型结构和参数影响很大,这是理论精度分析的困难所在. 然而,研究特定模型参数(例如高斯模型)下的理论近似精度是未来研究的重要方向.

4) 在具体应用方面,概率图的应用领域不仅局限于本文中提到的几个领域,它在计算机视觉中的跟踪和行为识别问题、机器学习的结构学习、生物信息学领域的数据挖掘、电网优化调度、交通网络动态优化、各种益智游戏的推理问题等实际问题上都表现出巨大的应用潜力,例如,以色列巴伊兰大学的 Goldberger 使用概率图模型及信度传播算法解决数独推理问题<sup>[108]</sup>. 概率图模型的具体应用需要解决问题描述及模型构建、专用算法设计、效果评估等问题,这是个多学科交叉及融合的问题.

总之,概率图模型推理问题具有重要的理论研究意义和应用价值,目前在理论研究上和实际应用中仍然存在很多问题亟待解决.

## References

- 1 Wainwright M J, Jordan M. Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 2008, 1(1-2): 1-305
- 2 Koller D, Friedman N. *Probabilistic Graphical Models Principles and Techniques*. Cambridge, MA: MIT Press, 2009
- 3 Tappen M F, Freeman W T. Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters. In: *Proceedings of the 9th IEEE International Conference on Computer Vision*. Beijing, China: IEEE, 2003. 900-906
- 4 Chu Yi-Ping, Zhang Yin, Ye Xiu-Zi, Zhang San-Yuan. Adaptive video segmentation algorithm using hidden conditional random fields. *Acta Automatica Sinica*, 2007, 33(12): 1252-1258  
(褚一平, 张引, 叶修梓, 张三元. 基于隐条件随机场的自适应视频分割算法. *自动化学报*, 2007, 33(12): 1252-1258)
- 5 Roth S, Black M J. Fields of experts: a framework for learning image priors. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. San Diego, USA: IEEE, 2005. 860-867
- 6 Freeman W T, Pasztor E C, Carmichael O T. Learning low-level vision. *International Journal of Computer Vision*, 2000, 40(1): 25-47

<sup>1</sup><http://vision.middlebury.edu/>

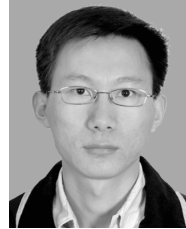
<sup>2</sup><http://jmlr.csail.mit.edu/papers/volume7/yanover06a/>

- 7 Kwatra V, Schödl A, Essa I, Turk G, Bobick A. Graphcut textures: image and video synthesis using graph cuts. *ACM Transactions on Graphics*, 2003, **22**(3): 277–286
- 8 Komodakis N, Tziritas G. Image completion using global optimization. In: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006. 442–452
- 9 Huang X D, Acero A, Hon H W. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. New Jersey: Prentice Hall PTR, 2001
- 10 Rush A M, Sontag D, Collins M, Jaakkola T. On dual decomposition and linear programming relaxations for natural language processing. In: Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing. Massachusetts, USA: Association for Computational Linguistics, 2010. 1–11
- 11 McEliece R J, MacKay D J C, Cheng J F. Turbo decoding as an instance of Pearl’s “belief propagation” algorithm. *IEEE Journal on Selected Areas in Communications*, 1998, **16**(2): 140–152
- 12 Gallager R G. Low density parity check codes. *IEEE Transactions on Information Theory*, 1962, **8**(1): 21–28
- 13 Kalman R E. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 1960, **82**(1): 35–45
- 14 Kschischang F R, Frey B J, Loeliger H A. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 2001, **47**(2): 498–519
- 15 McAuliffe J D, Pachter L, Jordan M I. Multiple-sequence functional annotation and the generalized hidden Markov phylogeny. *Bioinformatics*, 2004, **20**(12): 1850–1860
- 16 Siepel A, Haussler D. Combining phylogenetic and hidden Markov models in biosequence analysis. *Journal of Computational Biology*, 2004, **11**(2–3): 413–428
- 17 Yanover C, Meltzer T, Weiss Y. Linear programming relaxations and belief propagation — an empirical study. *Journal of Machine Learning Research*, 2006, **7**: 1887–1907
- 18 Yanover C, Schueler-Furman O, Weiss Y. Minimizing and learning energy functions for side-chain prediction. *Journal of Computational Biology*, 2008, **15**(7): 899–911
- 19 Jordan M I, Ghahramani Z, Jaakkola T, Saul L K. An introduction to variational methods for graphical models. *Machine Learning*, 1999, **37**(2): 183–233
- 20 Gilks W R, Richardson S, Spiegelhalter D J. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall, 1996
- 21 Pearl J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. San Mateo: Morgan Kaufmann, 1988
- 22 Yedidia J S, Freeman W T, Weiss Y. Constructing free energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 2005, **51**(7): 2282–2312
- 23 Aji S M, McEliece R J. The generalized distributive law. *IEEE Transactions on Information Theory*, 2000, **46**(2): 325–343
- 24 Park J D, Darwiche A. Complexity results and approximation strategies for MAP explanations. *Journal of Artificial Intelligence Research*, 2004, **21**(1): 101–133
- 25 Doucet A, Godsill S J, Robert C P. Marginal maximum a posteriori estimation using Markov chain Monte Carlo. *Statistics and Computing*, 2002, **12**: 77–84
- 26 Huang J B, Chavira M, Darwiche A. Solving MAP exactly by searching on compiled arithmetic circuits. In: Proceedings of the 21st National Conference on Artificial Intelligence. Menlo Park, USA: AAAI, 2006. 1143–1148
- 27 Liu Q, Ihler A. Variational algorithms for marginal MAP. In: Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence. Barcelona, Spain: AUAI, 2011. 453–462
- 28 Jiang J R, Rai P, Hal D. Message-passing for approximate MAP inference with latent variables. In: Proceedings of the Conference on Neural Information Processing Systems. Granada, Spain: MIT Press, 2011. 1197–1205
- 29 Cheng Q, Chen F, Dong J W, Xu W L. Approximating the sum operation for marginal-MAP inference. In: Proceedings of the 26th AAAI Conference on Artificial Intelligence. Toronto, Canada, 2012, 1882–1887
- 30 Weiss Y. Correctness of local probability propagation in graphical models with loops. *Neural Computation*, 2000, **12**(1): 1–41
- 31 Kolmogorov V, Wainwright M. On the optimality of tree-reweighted max-product message passing. In: Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence. Edinburgh, Scotland: AUAI, 2005. 316–323
- 32 Schraudolph N N, Kamenetsky D. Efficient exact inference in planar Ising models. In: Proceedings of the Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2009. 1417–1424
- 33 Schraudolph N N. Polynomial-time exact inference in NP-hard binary MRFs via reweighted perfect matching. In: Proceedings of the 13th International Conference on Artificial Intelligence and Statistics. Sardinia, Italy: JMLR W & CP, 2010. 717–724
- 34 Cowell R G, Dawid P, Lauritzen S L, Spiegelhalter D J. *Probabilistic Networks and Expert Systems*. Berlin: Springer, 1999
- 35 Wainwright M J, Jaakkola T S, Willsky A S. MAP estimation via agreement on trees: message-passing and linear programming. *IEEE Transactions on Information Theory*, 2005, **51**(11): 3697–3717
- 36 Wainwright M J, Jaakkola T S, Willsky A S. Tree-based reparameterization framework for analysis of sum-product and related algorithms. *IEEE Transactions on Information Theory*, 2003, **49**(5): 1120–1146
- 37 Cover T M, Thomas J A. *Elements of Information Theory*. New York: Wiley-Interscience, 2006
- 38 Aji S M, McEliece R J. The generalized distributive law and free energy minimization. In: Proceedings of the 39th Allerton Conference on Communication, Control, and Computing. Monticello, USA: IEEE, 2001. 672–681

- 39 Cheng Q, Chen F, Xu W L, Wang S. Recursive sum-product algorithm for generalized outer-planar graphs. *Information Processing Letters*, 2012, **112**(11): 449–456
- 40 Kolmogorov V. Convergent Tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, **28**(10): 1568–1583
- 41 Kolmogorov V, Zabih R. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, **26**(2): 147–159
- 42 Globerson A, Jaakkola T. Approximate inference using planar graph decomposition. In: Proceedings of the 2007 Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2007. 473–480
- 43 Van Engelen R A. Approximating Bayesian belief networks by arc removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(8): 916–920
- 44 D'Ambrosio B. Incremental probabilistic inference. In: Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence. Washington, USA: AUAI, 1993. 301–308
- 45 Poole D. Average-case analysis of a search algorithm for estimating prior and posterior probabilities in Bayesian networks with extreme probabilities. In: Proceedings of the 13th International Joint Conference on Artificial Intelligence. Chambéry, France: AAAI, 1993. 606–612
- 46 Wiergerinck W, Heskes T. Fractional belief propagation. In: Proceedings of the 2003 Conference on Neural Information Processing Systems. Whistler, Canada: MIT Press, 2003. 438–445
- 47 Heskes T. On the uniqueness of loopy belief propagation fixed points. *Neural Computation*, 2004, **16**(11): 2379–2413
- 48 Wainwright M J, Jaakkola T S, Willsky A S. A new class of upper bounds on the log partition function. *IEEE Transactions on Information Theory*, 2005, **51**(7): 2313–2335
- 49 Meshi O, Jaimovich A, Globerson A, Friedman N. Convexifying the Bethe free energy. In: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal, Canada: AUAI, 2009. 402–410
- 50 Heskes T. Convexity arguments for efficient minimization of the Bethe and Kikuchi free energies. *Journal of Artificial Intelligence Research*, 2006, **26**(1): 153–190
- 51 Komodakis N, Paragios N, Tziritas G. MRF energy minimization and beyond via dual decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, **33**(3): 531–552
- 52 Sontag D, Globerson A, Jaakkola T. Introduction to dual decomposition for inference. *Optimization for Machine Learning*. Cambridge, USA: MIT Press, 2011. 1–37
- 53 Opper M, Saad D. *Advanced Mean Field Methods: Theory and Practice*. Cambridge, USA: MIT Press, 2001
- 54 Xing E P, Jordan M I, Russell S. A generalized mean field algorithm for variational inference in exponential families. In: Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence. Acapulco, Mexico: AUAI, 2003. 538–591
- 55 Wiergerinck W. Variational approximations between mean field theory and the junction tree algorithm. In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence. Stanford, USA: AUAI, 2000. 626–633
- 56 Saul L K, Jordan M I. Exploiting tractable substructures in intractable networks. In: Proceedings of the 1997 Conference on Neural Information Processing Systems. Denver, USA: MIT Press, 1997. 486–492
- 57 Plefka T. Convergence condition of the TAP equation for the infinite-ranged Ising spin glass model. *Journal of Physics A: Mathematical and General*, 1982, **15**(6): 1971–1978
- 58 Leisink M A R, Kappen H J. Learning in higher order Boltzmann machines using linear response. *Neural Networks*, 2000, **13**(3): 329–335
- 59 Minka T P. A Family of Algorithms for Approximate Bayesian Inference [Ph.D. dissertation], Massachusetts Institute of Technology, Cambridge, MA, USA, 2001
- 60 Minka T P. Expectation propagation for approximate Bayesian inference. In: Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence. Seattle, USA: AUAI, 2001. 362–369
- 61 Minka T, Qi Y. Tree-structured approximations by expectation propagation. In: Proceedings of the 2003 Conference on Neural Information Processing Systems. Whistler, Canada: MIT Press, 2003. 1–8
- 62 Pakzad P, Anantharam V. Belief propagation and statistical physics. In: Proceedings of the 2002 Conference on Information Sciences and Systems. Princeton, USA: IEEE, 2002. 1–3
- 63 Pakzad P, Anantharam V. Estimation and marginalization using the Kikuchi approximation methods. *Neural Computation*, 2005, **17**(8): 1836–1873
- 64 Tatikonda S C, Jordan M I. Loopy belief propagation and Gibbs measures. In: Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence. Alberta, Canada: AUAI, 2002. 493–500
- 65 Ihler A T, Fisher J W, Willsky A S, Chickering M. Loopy belief propagation: convergence and effects of message errors. *Journal of Machine Learning Research*, 2006, **6**(1): 905–936
- 66 Chen Feng, Liu Hong, Xu Wen-Li. Recursive belief propagation — belief propagation in a well-order. *Acta Automatica Sinica*, 2010, **36**(8): 1091–1098  
(陈峰, 刘红, 徐文立. 递推信度传播算法 — 按良序的信度传播. 自动化学报, 2010, **36**(8): 1091–1098)
- 67 Meltzer T, Globerson A, Weiss Y. Convergent message passing algorithms: a unifying view. In: Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal, Canada: AUAI, 2009. 393–401
- 68 Globerson A, Jaakkola T. Fixing max-product: convergent message passing algorithms for MAP LP-relaxations. In: Proceedings of the 2008 Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2008. 553–560
- 69 Werner T. A linear programming approach to max-sum problem: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, **29**(7): 1165–1179

- 70 Hazan T, Shashua A. Convergent message-passing algorithms for inference over general graphs with convex free energies. In: Proceedings of the 2008 Conference on Uncertainty in Artificial Intelligence. Helsinki, Finland: AUAI, 2008. 264–273
- 71 Yuille A L. CCCP algorithms to minimize the Bethe and Kikuchi free energies: convergent alternatives to belief propagation. *Neural Computation*, 2002, **14**(7): 1691–1722
- 72 Pretti M, Pelizzola A. Stable propagation algorithm for the minimization of the Bethe free energy. *Journal of Physics A: Mathematical and General*, 2003, **36**: 11201–11211
- 73 Shimony S E. Finding MAPs for belief networks is NP-hard. *Journal of Artificial Intelligence*, 1994, **68**(2): 399–410
- 74 Sontag D A. Approximate Inference in Graphical Models Using LP Relaxations [Ph. D. dissertation], Massachusetts Institute of Technology, Cambridge, MA, USA, 2010
- 75 Alizadeh F. Interior point methods in semidefinite programming with applications to combinatorial optimization. *Journal on Optimization*, 1995, **5**(1): 13–51
- 76 Mitchell J E. Cutting plane methods and subgradient methods. *Tutorials in Operations Research*, 2005. 34–61
- 77 Bland R G, Goldfarb D, Todd M J. The ellipsoid method: a survey. *Operations Research*, 1981, **29**(6): 1039–1091
- 78 Ravikumar P, Lafferty J. Quadratic programming relaxations for metric labeling and Markov random field MAP estimation. In: Proceedings of the 23rd International Conference on Machine Learning. New York, USA: ACM, 2006. 737–744
- 79 van de Ven J, Fabio R. Distributed anytime MAP inference. In: Proceedings of the 2011 Conference on Uncertainty in Artificial Intelligence. Barcelona, Spain: AUAI, 2011. 708–716
- 80 Peng J, Hazan T, Srebro N, Xu J B. Approximate inference by intersecting semidefinite bound and local polytope. In: Proceedings of the 15th International Conference on Artificial Intelligence and Statistics. Canaries, Spanish: JMLR W & CP, 2012. 868–876
- 81 Kumar A, Zilberstein S, Toussaint M. Message-passing algorithms for MAP estimation using DC programming. In: Proceedings of the 15th International Conference on Artificial Intelligence and Statistics. Canaries, Spanish: JMLR W & CP, 2012. 656–664
- 82 Werner T. High-arity interactions, polyhedral relaxations, and cutting plane algorithm for soft constraint optimisation (MAP-MRF). In: Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition. Anchorage, USA: IEEE, 2008. 1–8
- 83 Weiss Y, Yanover C, Meltzer T. MAP estimation, linear programming and belief propagation with convex free energies. In: Proceedings of the 2007 Conference on Uncertainty in Artificial Intelligence. Vancouver, Canada: AUAI, 2007. 416–425
- 84 Jancsary J, Matz G, Trost H. An incremental subgradient algorithm for approximate MAP estimation in graphical models. In: Proceedings of the 2010 Conference on Neural Information Processing Systems. Vancouver, Canada, 2010
- 85 Sontag D, Jaakkola T. Tree block coordinate descent for MAP in graphical models. In: Proceedings of the 2009 International Conference on Artificial Intelligence and Statistics. Florida, USA: JMLR W & CP, 2009. 544–551
- 86 Osokin A, Vetrov D, Kolmogorov V. Submodular decomposition framework for inference in associative Markov networks with global constraints. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEE, 2011. 1889–1896
- 87 Batra D, Gallagher A C, Parikh D, Chen T. Beyond trees: MRF inference via outer-planar decomposition. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 2496–2503
- 88 Yarkony J, Fowlkes C, Ihler A. Covering trees and lower-bounds on quadratic assignment. In: Proceedings of the 2010 IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, USA: IEEE, 2010. 887–894
- 89 Sontag D, Jaakkola T. New outer bounds on the marginal polytope. In: Proceedings of the 2007 Conference on Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2007. 1393–1400
- 90 Sontag D, Meltzer T, Globerson A, Jaakkola T, Weiss Y. Tightening LP relaxations for MAP using message passing. In: Proceedings of the 2008 Conference on Uncertainty in Artificial Intelligence. Helsinki, Finland: AUAI, 2008. 503–510
- 91 Komodakis N, Paragios N. Beyond pairwise energies: Efficient optimization for higher-order MRFs. In: Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Miami, USA: IEEE, 2009. 2985–2992
- 92 Geoffrion A. Lagrangean relaxation for integer programming. *Approaches to Integer Programming*, 1974, **2**: 82–114
- 93 Johnson J. Convex Relaxation Methods for Graphical Models: Lagrangian and Maximum Entropy Approaches [Ph. D. dissertation], Massachusetts Institute of Technology, Cambridge, MA, USA, 2008
- 94 Hazan T, Shashua A. Norm-product belief propagation: primal-dual message-passing for approximate inference. *IEEE Transactions on Information Theory*, 2010, **56**(12): 6294–6316
- 95 Bertsekas D. Auction algorithms for network flow problems: a tutorial introduction. *Computational Optimization and Applications*, 1992, **1**(1): 7–66
- 96 Johnson J K, Malioutov D M, Willsky A S. Lagrangian relaxation for map estimation in graphical models. In: Proceedings of the 2007 Annual Allerton Conference on Communication, Control, and Computing. Illinois, USA: IEEE, 2007. 1–10
- 97 Gimpel K, Smith N A. Softmax-margin CRFs: training log-linear models with cost functions. In: Proceedings of the 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics-Human Language Technologies. Los Angeles, USA: Association for Computational Linguistics, 2010. 733–736

- 98 Savchynskyy B, Kappes J, Schmidt S, Schnörr C. A study of Nesterov's scheme for Lagrangian decomposition and MAP labeling. In: Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition. Colorado Springs, USA: IEEE, 2011. 1817–1823
- 99 Jojic V, Gould S, Koller D. Accelerated dual decomposition for MAP inference. In: Proceedings of the 2010 International Conference on Machine Learning. Haifa, Israel: ACM, 2010. 503–510
- 100 Ravikumar P, Agarwal A, Wainwright M J. Message-passing for graph-structured linear programs: proximal projections, convergence and rounding schemes. In: Proceedings of the 25th International Conference on Machine Learning. Helsinki, Finland: ACM, 2008. 800–807
- 101 Martins A L, Figueiredo M A T, Aguiar P M Q, Smith N A, Xing E P. An augmented Lagrangian approach to constrained MAP inference. In: Proceedings of the 28th International Conference on Machine Learning. Bellevue, USA: ACM, 2011. 1–8
- 102 Meshi O, Globerson A. An alternating direction method for dual MAP LP relaxation. In: Proceedings of the 2011 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Athens, Greece: Springer, 2011. 470–483
- 103 Rossi F, Van Beek P, Walsh T. *Handbook of Constraint Programming*. Amsterdam: Elsevier, 2006
- 104 Canzar S, Toussaint N C, Klau G W. An exact algorithm for side-chain placement in protein design. *Optimization Letters*, 2011, **5**(3): 393–406
- 105 Marinescu R, Dechter R. AND/OR branch-and-bound search for combinatorial optimization in graphical models. *Artificial Intelligence*, 2009, **173**(16–17): 1457–1491
- 106 Sun M, Telaprolu M, Lee H, Savarese S. Efficient and exact MAP-MRF Inference using branch and bound. In: Proceedings of the 2012 International Conference on Artificial Intelligence and Statistics. Canaries, Spanish: JMLR W & CP, 2012. 1134–1142
- 107 Meltzer T, Yanover C, Weiss Y. Globally optimal solutions for energy minimization in stereo vision using reweighted belief propagation. In: Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China: IEEE, 2005. 428–435
- 108 Goldberger J. Solving Sudoku Using Combined Message Passing Algorithms, Technical Report T R-BIU-ENG-2007-05-03, Bar-Ilan University, Israel, 2007



**程 强** 清华大学自动化系博士研究生。2009 年获得南京大学自动化专业学士学位。主要研究方向为概率图模型近似推理方法。E-mail: cheng-q09@mails.tsinghua.edu.cn

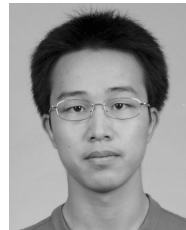
(**CHENG Qiang** Ph.D. candidate in the Department of Automation, Tsinghua University. He received his bachelor degree in automation from Nanjing University in 2009. His research interest covers approximate inference methods in graphical models.)



**陈 峰** 清华大学自动化系副教授。主要研究方向为概率图模型和视频图像分析。本文通信作者。

E-mail: chenfang@mail.tsinghua.edu.cn

(**CHEN Feng** Associate professor in the Department of Automation, Tsinghua University. His research interest covers visual surveillance, and image and video processing. Corresponding author of this paper.)



**董建武** 清华大学自动化系博士研究生。2010 年获得中南大学自动化专业学士学位。主要研究方向为概率图模型及其应用。

E-mail: djw10@mails.tsinghua.edu.cn

(**DONG Jian-Wu** Ph.D. candidate in the Department of Automation, Tsinghua University. He received his bachelor degree in automation from Central South University in 2010. His research interest covers graphical models and their applications.)



**徐文立** 清华大学自动化系教授。主要研究方向为计算机视觉, 图像视频处理, 机器人和自动控制。

E-mail: xuwl@mail.tsinghua.edu.cn

(**XU Wen-Li** Professor in the Department of Automation, Tsinghua University. His research interest covers computer vision, image and video processing, robotics, and automatic control.)