

计算机视觉中高阶能量项的优化

史利民, 余 淼, 胡占义

(中国科学院自动化研究所模式识别国家重点实验室 北京 100190)
(slm. sx@126. com)

摘 要: 如何极小化含高阶项的能量代价函数, 近 2 年引起了计算机视觉界的高度重视, 相关工作在计算机视觉三大国际会议 (ICCV, ECCV, CVPR) 及主流刊物上有不少报道. 文中对此进行了一些简单介绍, 以期对读者有一些帮助.

关键词: 马尔可夫随机场; 能量最小化; 高阶能量项

中图法分类号: TP391

A Short Introduction to the Minimization of Higher Order Potentials in Computer Vision

Shi Limin, Yu Miao, and Hu Zhanyi

(National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190)

Abstract: Recently, how to minimize the higher order potentials becomes a hot topic in computer vision, which is witnessed by the numerous publications in recent computer vision conferences (ICCV, ECCV, CVPR) as well as in major computer vision journals. This note is a short introduction to its current development, and could be of reference to the interested reader.

Key words: Markov random fields; energy minimization; higher order potentials

在计算机视觉中, 很多底层视觉问题都可以归结为求解一个马尔可夫随机场最大后验概率模型. 根据 Markov-Gibbs 等价关系, 此模型经常用一个能量极小化模型

$$E(x) = \theta_{\text{const}} + \sum_{p \in V} \theta_p(x_p) + \sum_{c \in C} \theta_c(x_c)$$

表述. 其中 c 称为 clique, C 是 clique 集合, $|c|$ 称为 clique c 的阶, 表示 c 中包含节点的个数. 当 $|c| \geq 3$ 时, 该模型即称为高阶能量模型.

目前较成熟的算法主要集中在求解当 $|c| = 2$ 时的 pairwise 模型. 求解这类模型的方法大致可分为 2 类: 从组合优化入手的 graph cuts 算法和以消息传递为途径的置信传播 (belief propagation, BP) 和重置权树消息传播 (tree-reweighted message passing, TRW) 算法^[1]. 近几年来, 随着各种有效算法的不断

提出和完善, 如 α -expansion, Loop BP, TRW-S 等, 这类概率模型被广泛地应用于计算机视觉的各个领域, 诸如图像分割、立体匹配、图像去噪、视频合成等. 可以说, 这类模型和求解算法的出现, 在上述领域内产生了变革性的推动作用; 其效率和效果都远远超出了之前算法所能达到的程度.

显而易见的是, 这种 pairwise 模型无法表达更为复杂的统计信息及先验知识. 众多的实验表明^[2-21], 通过在能量函数中引入高阶项, 可以更合理、精确地描述诸多底层视觉问题, 从而使实验结果比传统的低阶能量函数得到的结果有很大的改善. 基于此, 寻求更复杂的高阶能量函数的优化算法, 成为了一个自然的选择.

“世上没有免费的午餐”在引入高阶项的同时,

也带来了求解的困难. Kolmogorov 等^[22]指出, 对于任意一个函数 $f: L^n \rightarrow \mathbb{R}$ 当且仅当它的所有二元投影函数都是 submodular, 函数 f 就是 submodular; 当且仅当函数 f 为 submodular, f 可以在多项式时间内求解; L 是 label 的个数, n 是变量个数. 这一结论在理论上给出了能量函数可用 graph cuts 求解的充分必要条件, 并为高阶能量函数的优化提供了可行途径. 但问题是多数高阶能量函数都非 submodular, 特别是计算机视觉中许多重要的高阶约束, 如共面性约束等^[2]. 寻求高阶能量函数的普适求解算法并不是一件容易的事情. 目前关于高阶能量函数优化的研究主要集中在构造有用的 submodular 高阶能量函数类和针对一些特殊的 nonsubmodular 高阶能量函数设计有效求解算法上.

Kohli 等^[2]利用文献^[22]的结论, 构造了一类特殊的任意阶次的 submodular 能量函数, 该类能量函数可利用 α - β swap 或 α -expansion 在多项式时间内求解. 作为该类能量函数的一个特例, 文献^[2]给出了 p^n Potts model 作为 Potts model 的推广. 文献^[3]通过引入一个截断参数 Q , 将 p^n Potts model 推广为 Robust p^n model, 并给出了用 move making 进行近似求解的算法. Rother 等^[5]通过截断的方式得到 nonsubmodular 项的近似 submodular 表达, 然后对其进行最小化. 尽管该算法求解时每次迭代并不能保证得到全局最优解, 但在 nonsubmodular 项较少时可以得到比较好的近似解.

鉴于具体视觉应用的能量函数的高阶项具有“稀疏性”, 即高阶项仅有少量的 label 可以使其获得较小的能量, 这一性质可以用来缩小变量的搜索空间, 从而大大降低求解的复杂度. Rother 等^[6]利用高阶项的这个特性, 通过线性包络约束变量取值, 仅需添加少量的辅助变量就可将高阶能量项转化为等价的二次函数 QPBF (quadratic pseudo-Boolean function), 从而可以用 message passing 或 graph cuts 进行优化. Kohli 等^[7]提出了高阶能量项的一种新的表示方法——将其表示为线性函数的上下包络, 从而为高阶能量函数及其对应的 Gibbs 能量模型的求解提供了一个统一的框架. 其线性函数的下包络可以通过添加少量的辅助节点将其转化为 pairwise 能量项; 而上包络求解则较为困难, 文献^[7]提出了一种新的 message passing 算法对其进行求解. Lan 等^[8]也从“稀疏性”出发, 利用 BP 算法来求解高阶能量函数; 其采用了一种自适应状态空间来控制复杂度, 从而大大减少了消息计算和消息传

递, 使用 BP 算法来求解高阶能量函数成为可能. Potetz^[9]针对一类高阶能量优化问题, 使用变量替换将高阶项转换成低阶; 其同样利用“稀疏性”的特点, 通过自适应直方图约束变量搜索空间, 提高了 BP 算法的运算效率, 使算法复杂度由原来的随着 clique 规模增大呈指数增长, 降低为线性增长. Tarlow 等^[10]构造了 2 类新的高阶能量函数, 通过精心设计消息更新策略, 从而得到了更加高效的 BP 求解算法.

带约束的二标记问题在图像分割等领域有着重要的应用, 但其中很多都是 NP-hard 问题, 文献^[11-12]分别考虑了其中的 2 类问题. 文献^[11]将带有连接性约束的 MAP-MRF 问题转换为线性规划松弛问题, 并基于多面体组合理论构造了相应的求解算法. 虽然该算法是针对一类特殊的高阶能量函数提出的, 但其思想也可以推广到求解任意高阶能量项的问题. 文献^[12]针对带有标记数量约束的二标记问题, 基于图分解和动态规划构造了 DD (decompose dynamic) 算法, 它可以在多项式时间内得到原问题的 ϵ 近似解.

QPBO (quadratic pseudoboolean optimization)^[13]作为一种二标记问题求解算法, 不仅能够有效地求得 submodular 问题的最优解, 对于 nonsubmodular 问题也能得到部分变量的最优解. Kolmogorov 等^[13]通过对强联通子图进行拓扑排序来选择最小割, 使其标记尽可能多的节点. Rother 等^[14]给出了 2 种方法: QPBOP (quadratic pseudoboolean optimization probing) 和 QPBOI (quadratic pseudoboolean optimization improve). QPBOP 利用“试探”的方式, 在保持解的最优性的同时逐步确定剩余未知变量; QPBOI 从任意的能量开始, 通过利用 QPBO 算法迭代, 保证每次迭代能量值不增. 文献^[15]利用分层的 OPBOI 求解基于条件随机场构造的高阶能量函数, 从而使得光流计算从传统的局部光流一致性假设提升到局部仿射不变性假设. Woodford 等^[16]针对含有二阶平滑项的能量优化问题, 结合 QPBO 和推广的 α -expansion 算法构造了效率较高的求解算法.

Ladicky 等^[17]提出了关联层次网络模型 (associative hierarchical network), 其可视为关联马尔可夫网络模型 (associative Markov network) 在高阶情况下的一个自然推广. 该模型将多个层级下 (如像素级、分割区域级) 的特征计算融合到同一个模型框架中. Russell 等^[18]针对这一模型构造了一

种基于 graph cuts 的快速近似求解算法,其对于非常高阶(甚至上万阶)的 clique 也适用。

近年来提出的基于对偶分解的能量优化框架^[19-20]也是一个值得关注的方向。该框架将一个复杂的高阶 MRF 问题的拉格朗日对偶问题分解成相对简单的子问题(slave 问题),通过协调各子问题得到原问题的近似解(master 问题)。文献^[21]将这一优化框架用来处理高阶能量优化问题,并针对一类高阶能量函数(pattern based energy)构造了基于消息传递的有效算法。该框架为高阶能量优化提供了一种可能的有效途径。

到目前为止,尽管高阶 MRF 优化方面的研究已经取得一些进展,但总的来说这方面的工作还刚刚起步。已构造的 submodular 高阶能量函数形式都比较简单,还远远不能达到对很多视觉底层问题进行有效描述的要求;而所谓的 nonsubmodular 有效求解算法也都是针对少数特殊的 nonsubmodular 问题给出的。已提出的一些高阶能量优化问题的求解框架也不够完善,所以一般性的高阶能量函数优化问题仍然是一个极具挑战性但具有重要意义的研究方向,并已引起了越来越多的研究者的重视,在 2011 年 CVPR 上亦有多篇相关文章被录用^[23-27]。期待在不久的将来,针对高阶 MRF 能量优化能够出现进一步突破性的工作。

参考文献(References):

- [1] Szeliski R, Zabih R, Scharstein D, *et al.* A comparative study of energy minimization methods for Markov random fields with smoothness-based priors [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2008, 30(6): 1068-1080
- [2] Kohli P, Kumar M, Torr P. P3 & Beyond: solving energies with higher order cliques [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2007: 1-8
- [3] Kohli P, Ladicky L, Torr P. Graph cuts for minimizing robust higher order potentials [R]. Oxford: Oxford Brookes University, 2008
- [4] Kohli P, Ladicky L, Torr P. Robust higher order potentials for enforcing label consistency [J]. *International Journal of Computer Vision*, 2009, 82(3): 302-324
- [5] Rother C, Kumar S, Kolmogorov V, *et al.* Digital tapestry [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2005: 589-596
- [6] Rother C, Kohli P, Feng W, *et al.* Minimizing sparse higher order energy functions of discrete variables [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2009: 1382-1389
- [7] Kohli P, Kumar M P. Energy minimization for linear envelope MRFs [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2010: 1863-1870
- [8] Lan X, Roth S, Huttenlocher D P, *et al.* Efficient belief propagation with learned higher-order markov random fields [M] // *Lecture Notes in Computer Science*. Heidelberg: Springer, 2006, 3952: 269-282
- [9] Potetz B. Efficient belief propagation for vision using linear constraint nodes [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2007: 1-8
- [10] Tarlow D, Givoni I, Zemel R. Hop-map: efficient message passing with high order potentials [OL]. [2011-03-29]. http://www.cs.toronto.edu/~zemel/documents/TarlowGivoni-Zemel_AISTATS2010.pdf
- [11] Nowozin S, Lampert C H. Global connectivity potentials for random field models [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2009: 818-825
- [12] Lim Y, Jung K, Kohli P. Energy minimization under constraints on label counts [M] // *Lecture Notes in Computer Science*. Heidelberg: Springer, 2010, 6312: 535-551
- [13] Kolmogorov V, Rother C. Minimizing nonsubmodular functions with graph cuts—a review [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2007, 29(7): 1274-1279
- [14] Rother C, Kolmogorov V, Lempitsky V S, *et al.* Optimizing binary MRFs via extended roof duality [C] // *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*. Los Alamitos: IEEE Computer Society Press, 2007: 1-8
- [15] Glocker B, Heibel T H, Navab N, *et al.* TriangleFlow: optical flow with triangulation-based higher-order likelihoods [M] // *Lecture Notes in Computer Science*. Heidelberg: Springer, 2010, 6313: 272-285
- [16] Woodford O J, Torr P H S, Reid I D, *et al.* Global stereo reconstruction under second order smoothness priors [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2009, 31(12): 2115-2128
- [17] Ladicky L, Russell C, Kohli P, *et al.* Associative hierarchical CRFs for object class image segmentation [C] // *Proceedings of IEEE International Conference on Computer Vision*. Arlington: AUAI Press, 2009: 739-746
- [18] Russell C, Ladicky L, Kohli P, *et al.* Exact and approximate Inference in associative hierarchical [C] // *Proceedings of Conference on Uncertainty in Artificial Intelligence*. Arlington: AUAI Press, 2010: 8-11

- [19] Komodakis N, Paragios N, Tziritas G. MRF optimization via dual decomposition; message-passing revisited [C] // Proceedings of the 11th International Conference on Computer Vision. Arlington: AUAI Press, 2007: 1-8
- [20] Komodakis N, Paragios N, Tziritas G. MRF energy minimization and Beyond via dual decomposition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(3): 531-552
- [21] Komodakis N, Paragios N. Beyond pairwise energies: efficient optimization for higher-order MRFs [C] // Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2009: 2985-2992
- [22] Kolmogorov V, Zabih R. What energy functions can be minimized via graph cuts? [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(2): 147-159
- [23] Komodakis N. Efficient training for pairwise or higher order MRFs via dual decomposition [OL]. [2011-03-29]. <http://www.cvpapers.com/cvpr2011.html>
- [24] Gallagher A, Batra D, Parikh D. Inference for order reduction in MRFs [OL]. [2011-03-29]. http://ttic.uchicago.edu/~dparikh/Publications/GallagherBatraParikh_CVPR2011_ORI.pdf
- [25] Kim T, Nowozin S, Kohli O, *et al.* Variable grouping for energy minimization [OL]. [2011-03-29]. <http://www.nowozin.net/sebastian/papers/kim2011variablegrouping.pdf>
- [26] Bleyer M, Rother C, Kohli P, *et al.* Object stereo-joint stereo matching and object segmentation [OL]. [2011-03-29]. <http://www.cvpapers.com/cvpr2011.html>
- [27] Batra D, Kohli P. Making the right moves: guiding alpha-expansion using local primal-dual gaps [OL]. [2011-03-29]. http://ttic.uchicago.edu/~dbatra/publications/assets/bk_cvpr11.pdf