The Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)

Variational BEJG Solvers for Marginal-MAP Inference with Accurate Approximation of B-Conditional Entropy

Igor Kiselev
University of Waterloo
200 University Avenue West, Waterloo, Ontario N2L 3G1, Canada
ipkiselev@cs.uwaterloo.ca

Introduction

We developed novel variational MMAP inference algorithms and proximal convergent solvers, where we can improve the approximation accuracy while better preserving the original MMAP query by designing such a dual variational objective function that an upper bound approximation is applied only to the entropy of decision variables. A task of marginal-MAP (MMAP) inference is a "mixed" problem, where "max" or B-variables are decision variables to be optimized while all other variables are "sum" variables to be marginalized out. Since MMAP inference is computationally intractable ($\mathsf{NP}^{\mathsf{PP}}$-complete), approximate variational inference techniques such as upper bounding or convex relaxations methods are of particular interest. To address the scalability issues of solving MMAP problems on complex loopy graphs with multi-variable factors, the community has been making significant progress in developing approximate variational algorithms. Unfortunately, most current approximation approaches for MMAP have focused on pair-wise models and their extensions to the models with high-order interactions often result in simplification of the original MMAP query and greater approximation errors.

We propose four novel contributions. (1) Our novel BEJG objective function for models with high-order interactions decomposes the intractable MMAP problem into a linear combination of tractable sub-problems and ensures that B-variables are maximized jointly within each sub-problem. This design ensures that all and only B-variables remain in our upper bound of the intractable entropy of B-variables (B-entropy), which is completely removed now from the joint entropy as required by the original MMAP query. (2) Our novel BEJG variational algorithms (generalized and zero-temperature) solve MMAP exactly on any junction graph of a special AB-tree structure and provide accurate approximate results on general graphs with high-order interactions. (3) Our novel proximal solvers (BEJG and Generalized Bethe) for MMAP on models with multi-variable factors demonstrate improved convergence properties and are based on the proximal point approach, where we convert the direct optimization of the MMAP variational objective function into the sequence of proximal minimization problems with the property that after a finite number of iterations the sequence of associated solutions is converging to the MMAP solution. (4) Additionally, we scale our variational approach to solve POMDP problems and show that we can find the optimal controller by MMAP inference in a single-$\mathcal{DBN}$ generative model of a special cascading structure. We are the first to our knowledge who proposed BEJG variational algorithms and proximal algorithms to solve MMAP on the models with high-order interactions.

BEJG Variational Approach for MMAP

Consider an equivalent junction graph (JG) representation of a Bayesian Network $G = (C, S, \mathcal{E})$ with clusters C, separators S ($S_{kl} \subseteq \{C_k \cap C_l\}$) and edges $\mathcal{E}$. We ensure that this JG satisfies the running intersection property and family preservation property. We introduce types of clusters and separators by the type of variables in their scope (A and/or B-variables): ($C = \{C^A \cup C^B \cup C^{AB}\}$ and $\mathcal{E} = \{\mathcal{E}^A \cup \mathcal{E}^B \cup \mathcal{E}^{AB}\}$). Each cluster potential $\psi_{C_k}(x)$ equals to the product of factors associated with this cluster. We use the form of the original joint distribution factorized via cluster potentials $\psi_{C_k}(x)$, which we generalize by including "dummy" separator functions $b_{S_{kl}} = 1, \forall kl \in \mathcal{E}$. We also introduce parameters $\theta_{C_k}(x) = \ln \psi_{C_k}(x)$, $\theta_{S_{kl}}(x) = \ln b_{S_{kl}}(x)$ to obtain the convex exponential form:

$$p(x) = \exp[\textstyle\sum_{k \in C} \theta_{C_k}(x) - \sum_{kl \in \mathcal{E}} \theta_{S_{kl}}(x)] \,.$$

We derive our novel BEJG variational approximation for MMAP by (1) representing the MMAP problem with its exact dual variational form (Liu and Ihler 2013), and (2) designing a tractable BEJG objective function with a special upper bound approximation for B-entropy, which is a linear combination of all AB, B cluster and separator entropies with B-variables only.

Theorem 1 *The optimization of BEJG free energy F^{MMAP}_{BEJG} over the marginals from a convex set L provides an accurate MMAP estimate on general junction graphs with high-order interactions:*

$$\max_{\tau \in L} F^{MMAP}_{BEJG} = \mathbb{E}_{\,\tau}[\theta(x)] + \sum_{kl \in \mathcal{E}^{AB}} H_{S_{kl}}(x^B; \tau) + \sum_{k \in C \setminus C^B} H_{C_k}(\tau) - \sum_{kl \in \mathcal{E} \setminus \mathcal{E}^B} H_{S_{kl}}(\tau) - \sum_{k \in C^{AB}} H_{C_k}(x^B; \tau) \quad (1)$$

, *where set of locally consistent marginals* $L := \{\tau : \tau \geq 0; \sum_{\sim x_{S_{kl}}} \tau_{C_k}(x) = \tau_{S_{kl}}(x), \forall k, kl\}$*; the expected energy via cluster and separator marginals is* $\mathbb{E}_{\tau}[\theta(x)] = \sum_{k \in C} \sum_{x \in C_k} \tau_{C_k}(x) \cdot \theta_{C_k}(x) - \sum_{kl \in \mathcal{E}} \sum_{x \in S_{kl}} \tau_{S_{kl}}(x) \cdot \theta_{S_{kl}}(x)$*;* $H_{C_k}(\tau) = -\sum_{x \in C_k} \tau_{C_k}(x) \cdot \ln \tau_{C_k}(x)$ *is a cluster entropy via cluster marginals;* $H_{S_{kl}}(\tau) = -\sum_{x \in S_{kl}} \tau_{S_{kl}}(x) \cdot \ln \tau_{S_{kl}}(x)$ *is a separator entropy.*

We also provide special conditions when our BEJG approximation scheme gives exact MMAP solutions. Thus, we: (1) specify a special structure of a junction graph with high order interactions (denoted as AB-tree JG) on which MMAP inference is tractable, (2) prove that our BEJG objective function equals to the exact MMAP dual function on any AB-tree JG. Importantly, our BEJG objective function is distinct from JGBP approximate objective (Liu and Ihler 2013) since our special approximation for the B-conditional entropy $H(X_A \mid X_B, q)$ ensures that B-entropy is completely removed from the joint entropy as required by the exact MMAP dual form (Liu and Ihler 2013).

We designed generalized and zero-temperature BEJG variational algorithms for MMAP, which fixed points correspond to stationary points of BEJG variational problem (1).

Algorithm 1 Zero-temperature BEJG algorithm

for iteration $n = 1, 2, \dots$ pass messages $\mu_{k \to l}(x_{S_{kl}})$ from a cluster C_k to a cluster $C_l, \forall kl \in \mathcal{E}$:

$$C_k^A \overset{S^A}{\to} C_l : \mu_{k \to l}(x_{S_{kl}}) = \sum_{\sim x_{S_{kl}}} \alpha(x_{C_k}) \tag{2}$$

, where $\alpha(x_{C_k}) := \frac{\psi_{C_k}(x_{C_k})}{b_{S_{kl}}(x_{S_{kl}})} \cdot \prod_{j \in N(k) \setminus l} \mu_{j \to k}(x_{S_{kl}})$

$$C_k^{AB,B} \overset{S^B}{\to} C_l : \mu_{k \to l}(x_{S_{kl}}) = \max_{\sim x^B_{S_{kl}}} \Big\{ \sum_{x^A_{C_k}} \alpha(x_{C_k}) \Big\}$$

$$C_k^{AB} \overset{S^A}{\to} C_l : \mu_{k \to l}(x_{S_{kl}}) = \sum_{\sim x_{S_{kl}}} \Big\{ \alpha(x_{C_k}) \Big\} \cdot \delta[X^*_{C^B_k}]$$

, where δ is the indicator function, s.t. $\delta = 1$ if $x^B_{C_k} = x^*_B$ and $\delta = 0$ otherwise, $\forall x^B_{C_k} \in X^B_{C_k}$; and local MMAP is:

$$X^*_{C^B_k} = \arg\max_{X^B_{C_k}} \sum_{x^A_{C_k}} \psi_{C_k}(x) \prod_{j \in N(k)} \mu_{j \to k}(x_{S_{kl}})$$

$$C_k^{AB} \overset{S^{AB}}{\to} C_l : \mu_{k \to l}(x_{S_{kl}}) = \max_{\sim x^B_{S_{kl}}} \beta(x^B_{C_k}) \cdot \sum_{\sim x_{S_{kl}}} \frac{\alpha(x_{C_k})}{\beta(x^B_{C_k})}$$

, where $\beta(x^B_{C_k}) = \sum_{x^A_{C_k}} \psi_{C_k}(x) \prod_{j \in N(k) \setminus l} \mu_{j \to k}(x_{S_{kl}})$

At convergence, decode optimal MMAP configurations X^*_B with a Trace-MMAP procedure

Results. We compare our variational and proximal algorithms to existing variational methods, including JGBP (Liu and Ihler 2013), Weighted mini-bucket (Dechter and Rish 2003), and Dual decomposition (Ping, Liu, and Ihler 2015) methods. Here, we use (1) simulated models of various complexity generated from synthetic distributions, (2) Bayesian networks from known UAI benchmarks, and (3) MMAP problems encoding non-pairwise sequential decision-making problems. We show that our solvers outperform other variational methods for many of reported cases (cf. Table 1 and appendix). Additionally, we demonstrate the important real-life application of the proposed variational approaches to solve complex tasks of policy optimization and sequential decision-making problems (complex MMAP instances) (Kiselev and Poupart 2014), (Kiselev 2018).

Problem	size: # vars-factors - "B" variables		
Ising	64-176-32	81-225-41	400-1160-200
Solver	$\ln p(X^*_B)$ Error= (Exact/Best - Solver Estimate)		
BEJG	0	0.49146587	2.10468011
WMB	5.87321002	7.58984151	–
DD	4.27607773	4.88179131	27.92131600
GBProx	1.90346121	0	1.040683100
JG	1.33533977	10.6139733	36.51320000
BEProx	0	0.00711234	0

Problem	size: # vars-factors-"B" vars			
Hidden	40-79-20	70-139-35	60-119-30	50-99-25
Solver	$\ln p(X^*_B)$ Error			
BEJG	0	1.184743	0	0.31548
WMB	0	2.346445	1.566266	0.69414
DD	0	0.356240	2.161253	0.33595
GBProx	0	1.371855	0	0
JG	0	0	2.155431	3.51697
BEProx	0	0.020891	0	0

Problem	size: # vars-factors-"B" vars			
Bayesian	60-60-30	59-59-30	58-58-29	57-56-28
Solver	$\ln p(X^*_B)$ Error			
BEJG	0.035210	0	0	0.083177
WMB	2.780708	4.471639	4.562611	4.668145
DD	2.738148	3.292984	3.218876	3.324410
GBProx	0.559616	0	1.370763	0
JG	0.111067	3.555348	0.698378	0.090003
BEProx	0	0	0	0

Table 1: Accuracy comparison of zero-temperature BEJG algorithm (*BEJG*), BEJG Proximal solver (*BEProx*) and GB Proximal solver (*GBProx*) to existing methods in terms of log-MMAP error $\ln p(X^*_B)$ = (Exact/Best - Solver Result).

References

Dechter, R., and Rish, I. 2003. Mini-buckets: A general scheme for bounded inference. *J. ACM* 50(2):107–153.

Kiselev, I., and Poupart, P. 2014. Policy optimization by marginal-map probabilistic inference in generative models. In Lomuscio, and Scerri., eds., *AAMAS 2014*. IFAAMAS.

Kiselev, I. 2018. Variational inference techniques for sequential decision making in generative models. In *NIPS 2018, Workshop on Infer to Control*.

Liu, Q., and Ihler, A. 2013. Variational algorithms for marginal map. *J. Mach. Learn. Res.* 14(1):3165–3200.

Ping, W.; Liu, Q.; and Ihler, A. T. 2015. Decomposition bounds for marginal map. In *NIPS 2015*, 3267–3275.