Inference on diagrams in the category of Markov kernels (Extended abstract)

Grégoire Sergeant-Perthuis, Nils Ruet

Inference on diagrams in the category of Markov kernels (Extended abstract)

Grégoire Sergeant-Perthuis* Nils Ruet †

Graphical models are widely used families of probability distributions that capture conditional independence relations between a collection of variables $X_i, i \in I$; celebrated examples are Hidden Markov models and Bayesian Networks [1]. Graphical models are built from directed and undirected graphs $G = (I, A)$ where nodes $i \in I$ are uniquely identified with the variables X_i. Inference in graphical models ultimately boils down to inference for an undirected graphical model, achieved through the Belief Propagation algorithm [2]. Such Inference constitutes a specific instance of variational inference as it revolves around a free energy termed the Bethe free energy [2]. Adopting a variational inference perspective for graphical models has facilitated the extension of the Belief Propagation algorithm to encompass broader classes of probability distributions, enabling the accommodation of interactions among more than 2 variables in contrast to traditional graphical models (factor graphs [3]); this is achieved through the introduction of the Kikuchi free energies [4]. Let us denote $\mathbf{Mes}^f, \mathbf{Kern}^f$, the categories with objects finite measurable spaces and respectively with morphisms measurable maps and the second Markov Kernels (stochastic matrices). $\mathbf{Mes}^f$ can be seen as a subcategory of $\mathbf{Kern}^f$. As exhibited in [5–7], what underlies variational inference for those classes of probability distributions are presheaves from a finite poset to $\mathbf{Mes}^f$, which morphisms are epimorphisms. We will call them the 'graphical' presheaves. Our contribution is to extend the Generalized Belief Propagation [5] to any presheaf from a finite poset to $\mathbf{Kern}^f$. This work is contained in Chapter 9 of [8] and Appendix 1 of unpublished [9], where we consider the more general problem of optimizing a collection of cost functions over a presheaf of signals.

1. Motivation and related work

Consider a collection of agents represented by vertices $i \in I$ that can communicate their beliefs to neighboring vertices ∂i through undirected edges $e \in A$. Each agent has its own representation of its environment, denoted by E_i. They can share their beliefs with neighboring nodes $j \in \partial i$ through a measurable map $f^i_e : E_i \to E_e$. Graphical models and their extensions do not allow us to account for such heterogeneity in the way each agent models their environment. Such setting is better captured by cellular sheaves [10] and applications [11], important examples of which are Sheaf Neural Networks [12], are limited to functors from the poset associated to a graph ($i \leq e \iff i \in e$) to the category of finite vector spaces $\mathbf{Vect}^f$. We are interested in the more general case where beliefs transfer through a hierarchy, i.e. a poset, and we provide an algorithm for inference in such case where Sheaf Neural Networks can't be used; by convention, we consider presheaves instead of functors: 'orders' are given top-down. More generally, cellular sheaves are restricted to the face poset of a cell complex and hence don't apply to all hierarchies and therefore not to our case.

2. Free energy for poset shaped diagrams in $\mathbf{Kern}^f$ and message passing algorithm

Definition 1 (Graphical presheaves)**.** Let I be a finite set and $\mathscr{A} \subseteq \mathscr{P}(I)$ be a sub-poset of the powerset of I. Let $E_i, i \in I$ are finite sets. For $a \in \mathscr{A}$ $E_a := \prod_{i \in a} E_i$, let $F(a) := E_a$, and for $b \subseteq a$, let $F^a_b : E_a \to E_b$

*LCQB, Sorbone université, Paris, France, gregoire.sergeant-perthuis@sorbonne-universite.fr

†CIAMS, Université Paris-Saclay,Orsay, France

Submitted to:
ACT 24

be the projection map from $\prod_{i\in a} E_i$ to $\prod_{i\in b} E_i$. F is called a graphical presheaf from $\mathscr{A}$ to $\mathbf{Mes}^f$.

For $\mathscr{A}$ a finite poset, the 'zeta-operator' of $\mathscr{A}$, denoted ζ, from $\bigoplus_{a\in\mathscr{A}}\mathbb{R}$ to $\bigoplus_{a\in\mathscr{A}}\mathbb{R}$ is defined as, for any $\lambda\in\bigoplus_{a\in\mathscr{A}}\mathbb{R}$ and any $a\in\mathscr{A}$, $\zeta(\lambda)(a)=\sum\limits_{b\leq a}\lambda_b$. ζ is invertible [13], we denote μ its inverse; its matrix expression $(\mu(a,b), b\leq a)$ defines the Möbius function of $\mathscr{A}$. Let F be a presheaf from $\mathscr{A}$ to $\mathbf{Kern}^f$; $F^a_b: F(a)\to F(b)$ is denoted element-wise as $F^a_b(\omega_b|\omega_a)$, with $\omega_b\in F(b), \omega_a\in F(a)$. It induces a presheaf $\tilde{F}$ from $\mathscr{A}$ to $\mathbf{Vect}^f$, where $\tilde{F}^a_b:\mathbb{P}(F(a))\to\mathbb{P}(F(b))$ is the linear map that sends probability distributions $p\in\mathbb{P}(F(a))$ to $F^a_b\circ p$. Following [5], we introduce a free energy $\mathscr{F}(Q)=\sum_{a\in\mathscr{A}}c(a)\left(\mathbb{E}_{Q_a}[H_a]-S(Q_a)\right)$; $c(a)=\sum_{b\geq a}\mu(b,a)$ is the generalization of the inclusion-exclusion formula associated to $\mathscr{A}$. $S(Q_a)=-\sum_{\omega_a\in F(a)}Q_a(\omega_a)\ln Q_a(\omega_a)$ is the entropy of Q_a. We propose to solve $\inf_{Q\in\lim\tilde{F}}\mathscr{F}(Q)$. $\tilde{F}^*$ is the functor obtained by dualizing the morphisms $\tilde{F}^a_b$, i.e. $\tilde{F}^{*,b}_a:\tilde{F}(b)^*\to\tilde{F}(a)^*$ sends linear maps $l_b:\tilde{F}(b)\to\mathbb{R}$ to $l_b\circ\tilde{F}^a_b:\tilde{F}(a)\to\mathbb{R}$.

For a functor G from $\mathscr{A}$ to $\mathbb{R}$-vector spaces, we define μ_G as, for any $a\in\mathscr{A}$ and $v\in\bigoplus_{a\in\mathscr{A}}G(a)$, $\mu_G(v)(a)=\sum_{b\leq a}\mu(a,b)G^b_a(v_b)$. Let us define the function $FE:\prod_{a\in\mathscr{A}}\mathbb{P}(E_a)\to\prod_{a\in\mathscr{A}}\mathbb{R}$ as $FE(Q)=(\mathbb{E}_{Q_a}[H_a]-S_a(Q_a), a\in\mathscr{A})$, which sends a collection of probability measures over $\mathscr{A}$ to their Gibbs free energies. For any $Q\in\prod_{a\in\mathscr{A}}\mathbb{P}(E_a)$, let us denote $\mathrm{d}_Q FE$ as the differential of FE at the point Q.

Theorem 1. *Let $\mathscr{A}$ be a finite poset, let F be a presheaf from $\mathscr{A}$ to $\mathbf{Kern}^f$. Let $H_a: F(a)\to\mathbb{R}$ be a collection of (measurable) functions. The critical points of $\mathscr{F}$ are the $Q\in\lim\tilde{F}$ such that,*

$$\mu_{\tilde{F}^*}d_Q FE|_{\lim\tilde{F}}=0 \tag{1}$$

The message-passing algorithm we consider is Algorithm 1; it specializes to the General Belief Propagation for graphical presheaves. For two elements $a,b\in\mathscr{A}$, such that $b\leq a$, two types of messages are considered: top-down messages $m_{a\to b}\in\mathbb{R}^{F(b)}$ and bottom-up messages $n_{b\to a}\in\mathbb{R}^{F(a)}$.

Algorithm 1: Message passage algorithm for presheaves from $\mathscr{A}$ to $\mathbf{Kern}^f$

Data: Initialization: $(m^0_{a\to b}\in\mathbb{R}^{F(b)}, b,a\in\mathscr{A}$ s.t. $b\leq a)$, a poset $\mathscr{A}$, a presheaf $F:\mathscr{A}\to\mathbf{Kern}^f$;

for $t\leq T$ **do**

    for $a\in\mathscr{A}, b\in\mathscr{A}$ *such that* $b\leq a$ **do**

        $\forall\omega_a\in F(a),\quad n_{b\to a}(\omega_a)\leftarrow\prod_{\substack{c:b\leq c\\ c\nleq a}}\sum_{\omega'_b\in F(b)}m_{c\to b}(\omega'_b)\cdot F^a_b(\omega'_b|\omega_a)$

    end

    for $a\in\mathscr{A}, b\in\mathscr{A}$ *such that* $b\leq a$ **do**

        $b_a=e^{-H_a}\prod_{\substack{b\in\mathscr{A}:\\ b\leq a}}n_{b\to a}$

        $p_a=\frac{b_a}{\sum_{\omega_a}b_a(\omega_a)}$

        $m_{a\to b}\leftarrow m_{a\to b}\cdot\frac{\tilde{F}^a_b(p_a)}{p_b}$

    end

end

A criterion to stop the algorithm is when the beliefs do not change, i.e., when $p^{t+1}_a\approx p^t_a$. The fixed points of the previous message-passing algorithm correspond to critical points of $\mathscr{F}$ over $\lim F$ (Corollary of Theorem 2.2 [9]v2). Theorem 1 differs from a similar characterization of critical points of a free energy for specifications in [14] by the fact that the $\mu_{\tilde{F}^*}$ and $\lim\tilde{F}$ are applied to the same presheaf $\tilde{F}$ and not two different presheaves/functors (G,F).

References

[1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, 1988.

[2] J. S. Yedidia, W. T. Freeman, and Y. Weiss, *Understanding belief propagation and its generalizations*, p. 239–269. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003.

[3] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[4] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Generalized belief propagation," in *Proceedings of the 13th International Conference on Neural Information Processing Systems*, NIPS'00, (Cambridge, MA, USA), p. 668–674, MIT Press, 2000.

[5] J. Yedidia, W. Freeman, and Y. Weiss, "Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms," *IEEE Transactions on Information Theory*, vol. 51, pp. 2282–2312, July 2005.

[6] O. Peltre, "Message passing algorithms and homology," 2020. Ph.D. thesis, Link.

[7] O. Peltre, "A homological approach to belief propagation and Bethe approximations," in *International Conference on Geometric Science of Information*, pp. 218–227, Springer, 2019.

[8] G. Sergeant-Perthuis, *Intersection property, interaction decomposition, regionalized optimization and applications.* PhD thesis, Université de Paris, 2021. Link.

[9] G. Sergeant-Perthuis, "Regionalized optimization," 2022. `https://arxiv.org/abs/2201.11876`.

[10] J. Curry, *Sheaves, cosheaves and applications*. PhD thesis, The University of Pennsylvania, 2013. arXiv:1303.3255.

[11] J. Hansen and R. Ghrist, "Distributed optimization with sheaf homological constraints," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 565–571, 2019.

[12] C. Bodnar, F. D. Giovanni, B. P. Chamberlain, P. Lio, and M. M. Bronstein, "Neural sheaf diffusion: A topological perspective on heterophily and oversmoothing in GNNs," in *Advances in Neural Information Processing Systems* (A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, eds.), 2022.

[13] G.-C. Rota, "On the foundations of combinatorial theory I. Theory of Möbius functions," *Probability theory and related fields*, vol. 2, no. 4, pp. 340–368, 1964.

[14] G. Sergeant-Perthuis, "Compositional statistical mechanics, entropy and variational inference," 2024. `https://arxiv.org/abs/2403.16104`.