Global Sensitivity Analysis of MAP inference in Selective Sum-Product Networks

Julissa Villanueva Llerena
Institute of Mathematics and Statistics
University of São Paulo , Brazil
jgville@ime.usp.br

Denis Deratani Mauá
Institute of Mathematics and Statistics
University of São Paulo, Brazil
ddm@ime.usp.br

1 Introduction

Sum-Product Networks (SPN) are deep probabilistic models that have exhibited state-of-the-art performance in several machine learning tasks [1, 2, 3, 4, 5, 6, 7, 8]. As with many other probabilistic models, performing Maximum-A-Posteriori (MAP) inference is NP-hard in SPNs [9, 10]. A notable exception is selective SPNs [11, 9], that allows MAP inference in linear time. Due to the high number of parameters, SPNs learned from data can produce unreliable and overconfident inference. This effect can be partially detected by performing a Sensitivity Analysis (SA) of the model predictions to changes in the parameters [12]. In this work, we develop efficient algorithms for global quantitative analysis of MAP inference in selective SPNs. In particular, we devise a polynomial-time procedure to decide whether a given MAP configuration is *robust* with respect to changes in the model parameters. Experiments with real-world datasets show that this approach can discriminate easy- and hard-to-classify instances, often more accurately than criteria based on the probabilities induced by the model.

2 MAP Inference in Sum-Product Networks

An SPN $\mathtt{S}$ is a rooted weighted acyclic directed graph with indicator variables as leaves, and sum and product as internal nodes. We assume that SPN are *complete*, *decomposable* and *normalized* [1]. To ensure linear-time MAP inference, we also assume that SPNs are *selective* [11] (i.e., that at most one child sub-network of any sum node evaluates to nonzero at any realization). For example, the evaluation of the selective SPN in Figure 1(a) at $X_0 = 0$, $X_1 = 0$ and $X_2 = 1$ is $\mathtt{S}(0, 0, 1) = 0.4(0 \times 0.3 \times 0.6) + 0.6(0.8 \times 0.9 \times 1) = 0.432$.

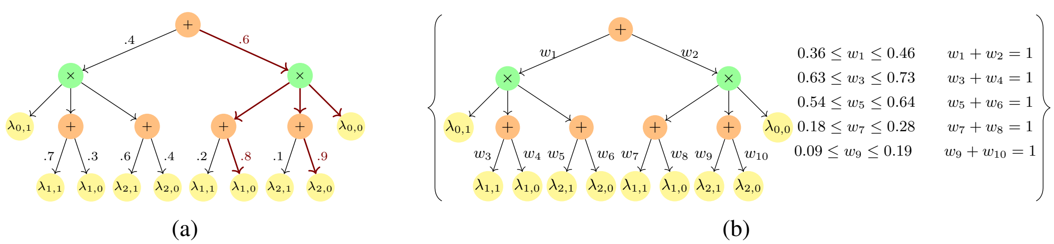

Figure 1: (a) Selective SPN. (b) CSPNs obtained by 0.1-contamination of the SPN in (a).

An SPN $\mathtt{S}$ induces a probability measure $\mathbb{P}_{\mathtt{S}}$ over the domain of its scope. Thus, given n SPN $\mathtt{S}$ with scope $\mathbf{X}, \mathbf{E}$ and *evidence* $\mathbf{e} \in val(\mathbf{E})$, we define the set of **Maximum-A-Posteriori** (MAP) inference instantiations as:

$$\mathbf{x}^* \in \arg\max_{\mathbf{x} \in val(\mathbf{X})} \mathbb{P}(\mathbf{x}|\mathbf{e}) = \arg\max_{\mathbf{x} \in val(\mathbf{X})} \mathtt{S}(\mathbf{x}, \mathbf{e})\,. \quad (1)$$

The problem is solvable in linear in the size of the network for selective SPNs by the simple Max-Product algorithm, that consists in replacing sum operations with maximizations and then evaluating the corresponding Max-Product network [11, 9]. The corresponding MAP instantiation is obtained by backtracking the solutions of the maximizations from the root toward the leaves. The highlighted subnetwork in Figure 1(a) contains the arcs selected by Max-Product.

3 Global Sensitivity Analysis of MAP Inferences in SPNs

In order to investigate the sensitivity of an inference w.r.t.] perturbations in the model parameters, we consider a set of SPNs $\{\mathtt{S}_{\mathbf{w}} : \mathbf{w} \in \mathcal{C}\}$, where each $\mathtt{S}_{\mathbf{w}}$ is parameterized by weights **w** and share the same network structure. We also assume that $\mathcal{C}$ is given by the Cartesian product of sets $\mathcal{C}_i$, one for each sum node i. Two common approaches to obtaining $\mathcal{C}_i$ is ϵ-contamination:

$$\mathcal{C}_i = \{(1-\epsilon)\mathbf{w}_i + \epsilon\mathbf{v} : v_j \geq 0, \sum_j v_j = 1\}\,, \tag{2}$$

where $\epsilon \in (0, 1)$, $\mathbf{w}_i$ are the weights associated to sum node i and j ranges over the children of i, and the **Imprecise Dirichlet Model** (IDM) [13]:

$$\mathcal{C}_i = \{\mathbf{w}_i : w_{ij} = \frac{N_{ij} + s \cdot v_i}{s + \sum_j N_{ij}}, v_j \geq 0, \sum_j v_j = 1\}\,, \tag{3}$$

where N_{ij} denote the number of times that the edge $i \to j$ is nonzero over the dataset, and s is the prior strength s of a Multinomial-Dirichlet model.

In this work, we perform global SA by verifying whether a MAP inference $\mathbf{x}^*$ is the single maximizer of each SPN in a set $\{\mathtt{S}_{\mathbf{w}} : \mathbf{w} \in \mathcal{C}\}$ given evidence $\mathbf{e}$, that is, whether:

$$\max_{\mathbf{x} \neq \mathbf{x}^*} \max_{\mathbf{w} \in \mathcal{C}} \left(\frac{\mathtt{S}_{\mathbf{w}}(\mathbf{x}, \mathbf{e})}{\mathtt{S}_{\mathbf{w}}(\mathbf{x}^*, \mathbf{e})} \right) < 1\,. \tag{4}$$

The following algorithm computes the left-hand side of the inequality below:

$$V^i = \begin{cases} 1 & \text{if } i \text{ is consistent,} \\ 0 & \text{if } i \text{ is inconsistent,} \\ \prod_j V^j & \text{if } i \text{ is a product node,} \\ \max\left\{ \max_{j \in ch(i), j \neq k} \left(\max_{\mathbf{w}_i \in \mathcal{C}_i} \frac{w_{ij} \max_{\mathbf{x}} \mathtt{S}^j_{\mathbf{w}_j}(\mathbf{x}, \mathbf{e})}{w_{ik} \mathtt{S}^k_{\mathbf{w}_k}(\mathbf{x}^*, \mathbf{e})} \right),\ V^k \right\} & \text{if } i \text{ is a sum node,} \end{cases} \tag{5}$$

where k is the active child of i for $\mathtt{S}$ at MAP configuration and the evidence $(\mathbf{x}^*, \mathbf{e})$.

4 Experiments and Results

We perform SA of MAP inferences on 8 benchmark multilabel classification domains. We assess the ability to distinguish between reliable and unreliable classifications by measuring both exact match (i.e., whether all labels are correct) and accuracy (which rewards relevant predictions while discounting for irrelevant predictions), on instances deemed robust or non-robust by each method. We compare our approach against a baseline procedure that considers a classification robust if the difference between the most probable and the second most probable configurations exceeds a given threshold p. We select the best values $\epsilon*$, $s*$ and $p*$ for resp. ϵ-contamination, IDM and difference of probability criteria using a validation dataset, an then perform a SA for each multilabel prediction in the test dataset. The results are in the table below.

Dataset		Accuracy			Exact Match		
		Robust	¬*Robust*	ΔAcc	*Robust*	¬*Robust*	ΔEM
Arts	$\epsilon*$	**0.88**	0.196	0.634	**0.833**	0.143	**0.69**
	$s*$	0.107	0.351	-0.244	0.089	0.247	-0.158
	$p*$	0.81	0.159	**0.651**	0.75	0.107	0.643
Business	$\epsilon*$	0.751	0.582	0.169	0.617	0.598	0.019
	$s*$	**0.781**	0.641	0.14	**0.642**	0.469	0.173
	$p*$	0.762	0.581	**0.181**	0.62	0.392	**0.228**
Emotions	$\epsilon*$	0.595	0.41	0.185	0.238	0.163	0.075
	$s*$	**0.686**	0.413	**0.273**	**0.308**	0.16	**0.148**
	$p*$	0.574	0.391	0.183	0.176	0.176	0
Flags	$\epsilon*$	0.917	0.468	0.449	0.5	0.118	0.382
	$s*$	0.917	0.468	0.449	0.5	0.1	**0.4**
	$p*$	0.917	0.468	0.449	0.5	0.118	0.382

Dataset		Accuracy			Exact Match		
		Robust	¬*Robust*	ΔAcc	*Robust*	¬*Robust*	ΔEM
Health	$\epsilon*$	**0.667**	0.557	0.11	0.5	0.416	0.084
	$s*$	0.637	0.482	**0.155**	0.537	0.304	**0.233**
	$p*$	0.655	0.552	0.103	**0.552**	0.409	0.143
Human	$\epsilon*$	0.203	-	-	0.146	-	-
	$s*$	0.203	-	-	0.146	-	-
	$p*$	**0.211**	0.198	0.013	**0.155**	0.14	0.015
Plant	$\epsilon*$	0.331	0.217	0.114	0.324	0.213	0.111
	$s*$	**0.367**	0.162	**0.205**	**0.362**	0.157	**0.205**
	$p*$	0.345	0.212	0.132	0.338	0.208	0.13
Scene	$\epsilon*$	0.857	0.277	0.58	0.857	0.212	0.645
	$s*$	**0.929**	0.293	0.636	**0.929**	0.293	0.636
	$p*$	0.923	0.276	**0.647**	0.923	0.211	**0.712**

One sees that the SA obtained with IDM achieve the highest accuracy and exact math in 5 of the 8 domains, followed by the ϵ-contamination (3 out of 8). Probability difference outperforms the other methods w.r.t. accuracy only for Human, where the our methods consider all instances robust. Yet the difference in accuracy (or exact match scores) for robust and non-robust is very small, showing that the two sets perform indeed very similar (but perhaps it would be more sensible to classify all instances as non-robust). The IDM-based SA perform particularly poorly in the Arts domain, where the accuracy of the robust portion is significantly inferior to the accuracy in the non-robust portion.

References

[1] Hoifung Poon and Pedro Domingos. Sum-product networks: A new deep architecture. In *Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence*, pages 337–346, 2011.

[2] Mohamed R Amer and Sinisa Todorovic. Sum product networks for activity recognition. *IEEE transactions on pattern analysis and machine intelligence*, pages 800–813, 2016.

[3] Kaiyu Zheng, Andrzej Pronobis, and Rajesh PN Rao. Learning graph-structured sum-product networks for probabilistic semantic maps. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*, 2018.

[4] Andrzej Pronobis, Francesco Riccio, and Rajesh PN Rao. Deep spatial affordance hierarchy: Spatial knowledge representation for planning in large-scale environments. In *ICAPS 2017 Workshop on Planning and Robotics*, 2017.

[5] Andrzej Pronobis and Rajesh PN Rao. Learning deep generative spatial models for mobile robots. In *Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on*, pages 755–762, 2017.

[6] Wei-Chen Cheng, Stanley Kok, Hoai Vu Pham, Hai Leong Chieu, and Kian Ming A Chai. Language modeling with sum-product networks. In *Fifteenth Annual Conference of the International Speech Communication Association*, 2014.

[7] Julissa Villanueva and Denis Deratani Mauá. On using sum-products networks for multi-label classification. In *Proceedings of the Sixth Brazilian Conference on Intelligent Systems*, pages 25–30, 2017.

[8] Robert Peharz, Antonio Vergari, Karl Stelzner, Alejandro Molina, Martin Trapp, Kristian Kersting, and Zoubin Ghahramani. Probabilistic deep learning using random sum-product networks. *CoRR*, abs/1806.01910, 2018.

[9] Robert Peharz, Robert Gens, Franz Pernkopf, and Pedro Domingos. On the latent variable interpretation in sum-product networks. *Journal of Transactions on Pattern Analysis and Machine Intelligence*, pages 1–14, 2016.

[10] Diarmaid Conaty, Denis Deratani Mauá, and Cassio Polpo de Campos. Approximation complexity of maximum a posteriori in sum-product networks. In *Proceedings of the Thirty-Third conference on Uncertainty in artificial intelligence*, pages 322–331, 2017.

[11] Robert Peharz, Robert Gens, and Pedro Domingos. Learning selective sum-product networks. In *LTPM workshop*, 2014.

[12] Denis Deratani Mauá, Diarmaid Conaty, Fabio Gagliardi Cozman, Katja Poppenhaeger, and Cassio Polpo de Campos. Robustifying sum-product networks. *International Journal of Approximate Reasoning*, pages 163–180, 2018.

[13] Peter Walley. Inferences from multinomial data: learning about a bag of marbles. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):3–34, 1996.