NeuroBE: Escalating Neural Network Approximations of Bucket Elimination

Sakshi Agarwal, Kalev Kask, Alex Ihler, Rina Dechter
{sakshia1, kkask, ihler, dechter} @ics.uci.edu

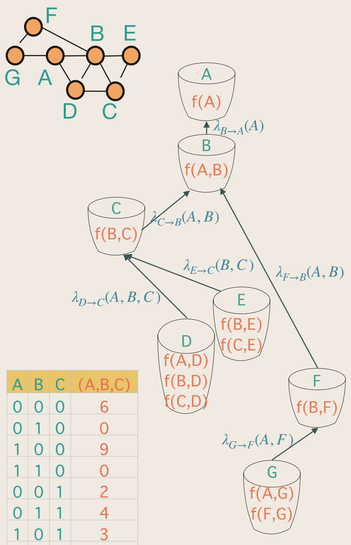
uai2022

Summary

We are using the power of neural networks to approximate the universal inference scheme of bucket elimination to compute the partition function.

$$Z = \sum_{x} \prod_{\alpha} f_\alpha(x_\alpha)$$

Bucket Elimination (BE)

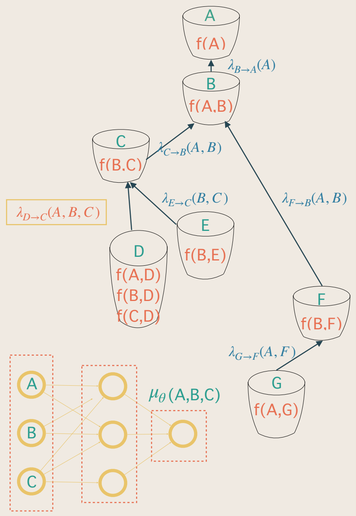


A	B	C	(A,B,C)
0	0	0	6
0	1	0	0
1	0	0	9
1	1	0	0
0	0	1	2
0	1	1	4
1	0	1	3
1	1	1	1

Time and space complexity exponential in induced width => not scalable

[Dechter, 1999]

Deep BE

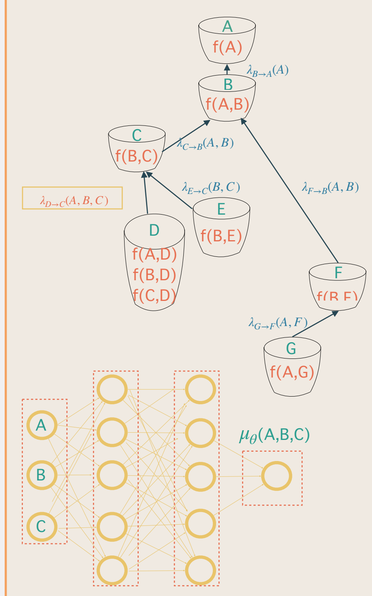


NN across buckets have : 1) fixed architecture; 2) fixed no.of samples; 3) uniform training loss

Complexity:
Time: $O(n(|T_{NN}| + N \,.\, r \,.\, k^{i+1}))$
Space: $O(n(k^i + N + |NN|))$

[Razhegi et al, 2021]

NeuroBE



NNs across buckets have:
1. varying NN architecture size
2. varying training sample size
3. weighted training loss

Complexity:
Time: $O(n(|T_{NN}| + N(w) \,.\, r \,.\, k^{i+1}))$
Space: $O(n(k^i + N(w) + |NN(w)|))$

Notation

n : number of variables
r : number of functions
k : domain size
i : i-bound
w : width
N : #train samples
$|NN|$: NN size
$|T_{NN}|$: NN evaluation time

Utilizing message size

- Processing each bucket (width w) eliminating X with constants : L, b, η (same across buckets)
- Feed-forward NN: L Layers, #b*w hidden-units.
  NN size, |NN(w)| [Barlett et al, 2019] $\propto (L * b * w)^2 log(b * w)$
- Sample complexity, [Vapnik, 1999]
  $$N(w) = \eta * (L * b * w)^2 log(b * w)$$
- Generate a sample set $\{(s, \mu^*_{norm}(s))\}$ of size N(w) as:
  i. Sample a configuration s uniformly, calculate :
  $$(\mu^*(s)) = \sum_{X} \prod_{f_\alpha \in B} f_\alpha(s)$$
  ii. Normalize (μ^*) as $\mu^*_{norm}(s) = \frac{\mu^*(s) - \beta}{\gamma - \beta}$
  where, γ and β are max, min μ^* values in training dataset $(\mathcal{D}_{Train})$, respectively.

Loss Function

- Minimize the Importance mean square error as loss over mini-batches $\mathcal{D}_i \in \mathcal{D}_{Train}$, NN outputs μ(S), weight function W(S).

$$I.m.s.e = \frac{1}{\#\mathcal{D}_i} \sum_{s \in \mathcal{D}_i} (\mu^*_{norm}(s) - \mu(s))^2 * W(s),$$

where, $W(s) = \frac{\mu^*(s)}{\sum_{s' \in \mathcal{D}_{Train}} \mu^*(s')}$

Performance Evaluation

- Local Bucket Error : $log\mu^* - log\mu$
- Global Bucket Error : $log\lambda - log\mu$
- Global Error : $error = |log_{10}Z^* - log_{10}\hat{Z}|$
- Comparing against WMB, DBE

Benchmarks

Grid: easy and hard instances, without-deterministic, 12 instances;

Pedigree: hard, with-deterministic, 7 instances;

DBN: medium-without determinism, 6 instances

Problem Description						WMB		DBE (#h=100, N=320k)				NeuroBE (#h=w, N(w,η=4), N_{min} = 19k)					
i-bound=20					ref Z			statistics over 10 runs				Statistics on resources			statistics on error over 10 runs		
Id	name	k	#v	w		error	#NB	avg error	min error	stdev	time	h_{max}	N_{avg}	N_{max}	avg error	min error	std	time (h)
1	grid4040f10	2	1600	55	5490	215.45	308	**97.1**	**11.81**	65.15	11.8	55	60k	182k	**24**	**16.51**	8	5.4
2	grid4040f5	2	1600	55	2800	84.92	308	**39.9**	**6.28**	34.96	11.7	55	60k	182k	**21.2**	**18.87**	6.4	5.3
3	grid4040f2	2	1600	55	1220	25.24	308	**7.34**	**1.2**	5.4	11.5	55	60k	182k	**5.1**	**3.46**	1.8	5.2
4	grid4040f15	2	1600	55	8200	338.2	308	**83.46**	**41.78**	34.2	13.4	55	60k	182k	**22.4**	**5.38**	13.3	5.4
5	grid4040f10w	2	1600	114	5637	297.7	376	100.5	**6.4**	82.15	21.2	114	209k	900k	**51.05**	**29.72**	18.9	13.7
6	grid4040f5w	2	1600	114	2819	136.99	376	78.2	**72.62**	5.6	**21.3**	114	209k	900k	**16.6**	**12.28**	3.92	13.9
7	grid4040f2w	2	1600	114	1231	32	376	**15.12**	**0.92**	20.52	18.2	114	209k	900k	25.78	**12.61**	17.9	11.8
8	grid4040f15w	2	1600	114	8230	657.03	376	**220.91**	**95.23**	192.2	17.7	114	209k	900k	103.9	79.59	27.8	**13.7**

Problem Description						WMB		DBE (#h=100, N=320k)				NeuroBE (#h=3w, N(w,η=10), N_{min} = 49k)					
i-bound=20					ref Z			statistics on error over 10 runs				Statistics on resources			statistics on error over 10 runs		
Id	name	k	#v	w		error	#NB	avg error	min error	stdev	time	h_{max}	N_{avg}	N_{max}	avg error	min error	std	time (h)
1	pedigree13	3	888	33	-31.18	6.4696	127	5.32	2.62	3.06	11.4	96	218k	706k	1.11	0.76	0.25	7.5
2	pedigree41	5	885	32	-76.04	4.1497	92	4.27	3.25	0.734	9.9	93	190k	658k	0.47	0.153	0.21	6.2
3	pedigree51	5	871	35	-77.27	9.7624	120	**23.92**	**9.23**	**12.74**	13.1	102	259k	809k	3.51	1.96	0.89	10.4
4	pedigree34	5	922	33	-64.23	7.0762	106	5.91	1.57	5.98	11.1	96	211k	706k	0.65	0.23	0.29	6.96
5	pedigree7	4	867	34	-64.82	6.0012	108	11.26	5.18	7.8	10.5	99	350k	900k	1.75	1.21	0.7	10.7
6	pedigree19	5	693	28	-59.020	2.5809	43	6.054	5.41	0.92	9.35	71	149k	482k	2.61	1.91	0.6	5

Problem Description						WMB		DBE (#h=100, N=320k)				NeuroBE (#h=3w, N(w,η=30), N_{min} = 147k)					
i-bound=20					ref Z			statistics on error over 10 runs				Statistics on resources			statistics on error over 10 runs		
Id	name	k	#v	w		error	#NB	avg error	min error	stdev	time	h_{max}	N_{avg}	N_{max}	avg error	min error	std	time (h)
1	rbm20	2	40	21	58.53	0.0007	20	0.22	0.034	0.17	1.05	60	147k	147k	0.23	0.002	0.21	0.45
2	rbm21	2	42	22	63.15	6.39	22	0.48	**0.27**	0.19	1.25	63	163k	164k	0.78	0.133	0.8	0.46
3	rbm22	2	40	21	66.55	8.65	24	0.47	**0.14**	0.35	1	66	180k	182k	0.32	0.035	0.26	0.57
4	rbm-ferro20	2	44	23	151.16	0.005	20	1.33	0.29	1.21	1.03	60	147k	147k	1.58	0.38	1.15	0.44
5	rbm-ferro21	2	42	22	152.62	1.98	22	3.43	**0.83**	1.89	1.17	63	163k	164k	2.79	0.125	2.57	0.48
6	rbm-ferro22	2	44	23	166.11	0.517	24	6.52	3.86	1.5	1.3	66	180k	182k	4.69	1.09	3.2	0.59