Learning the Structure of Probabilistic Sentential Decision Diagrams

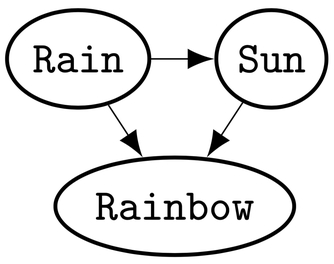
Yitao Liang, Jessa Bekker, Guy Van den Broeck

Ucla KU LEUVEN STAR AI RESEARCH LAB UCLA

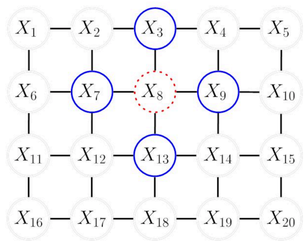
Background

Tractable vs. Intractable Representation

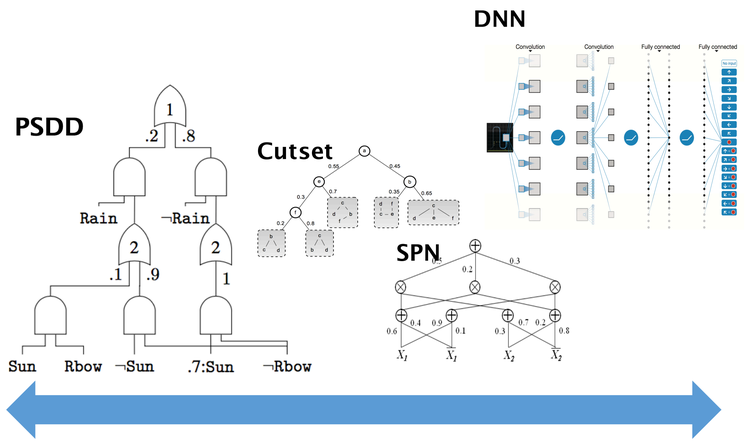
Bayesian networks



Markov networks



Tradeoff: Properties vs. Expressiveness



Strong Properties — Expressiveness

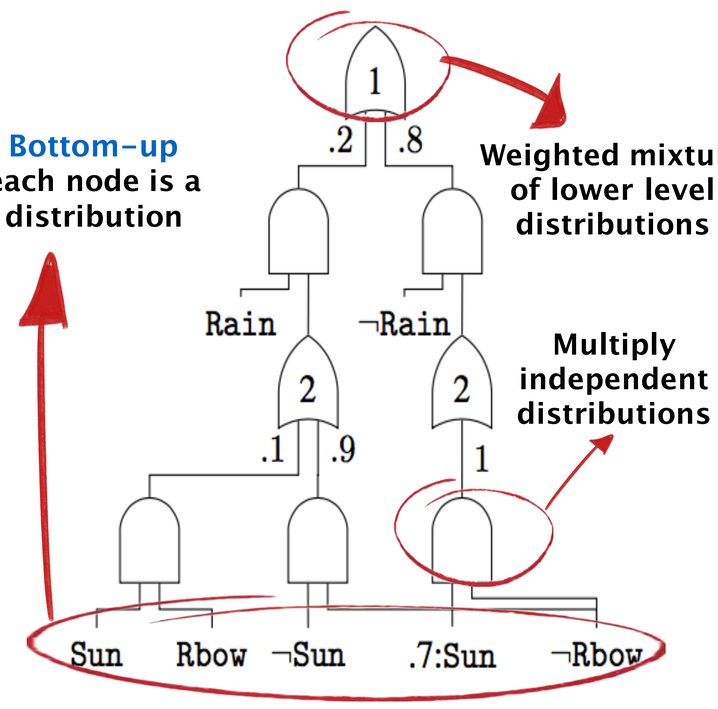
PSDD

The Most Powerful Circuit Representation

- Closed-form parameter estimation
- Efficient multiplication
- Fast MPE inference
- Etc.

Structure learning

Bottom-up
each node is a distribution



Weighted mixture of lower level distributions

Multiply independent distributions

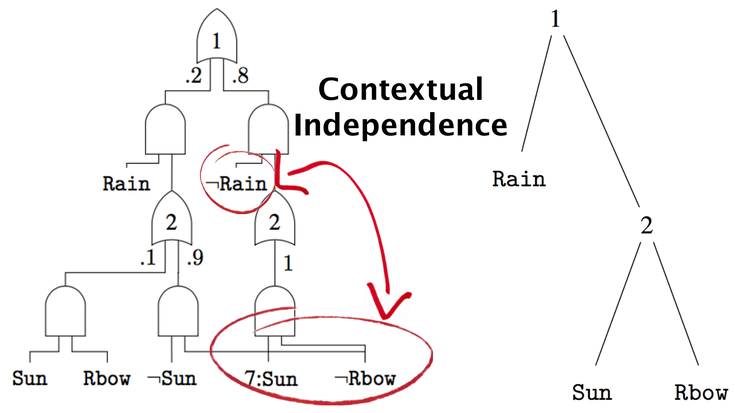
$$\Pr(\texttt{Rain}) = 0.2,$$

$$\Pr(\texttt{Sun} \mid \texttt{Rain}) = \begin{cases} 0.1 \text{ if } \texttt{Rain} \\ 0.7 \text{ if } \neg\texttt{Rain} \end{cases}$$

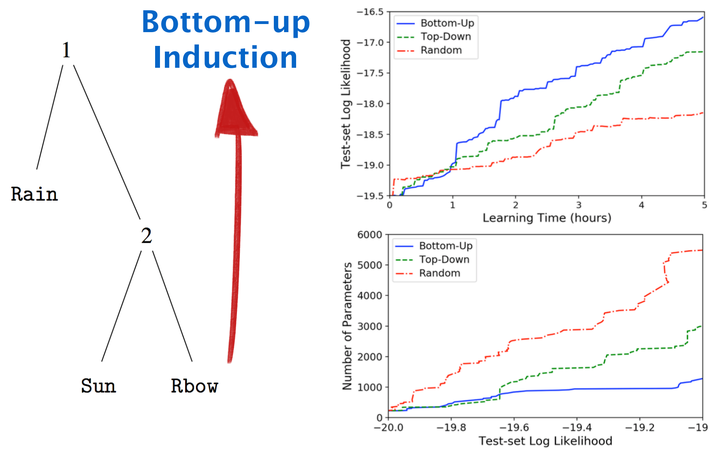
$$\Pr(\texttt{Rbow} \mid \texttt{R}, \texttt{S}) = \begin{cases} 1 \text{ if } \texttt{Rain} \wedge \texttt{Sun} \\ 0 \text{ otherwise} \end{cases}$$

Are PSDDs amenable to tractable learning

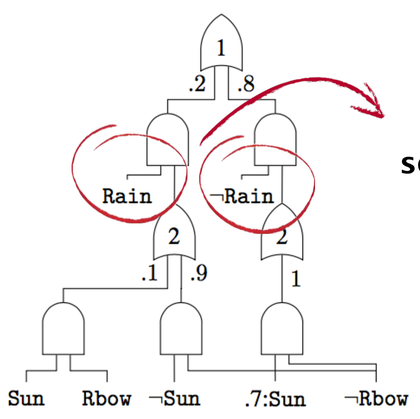
Contextual Independence & Vtree



Bottom-up Induction

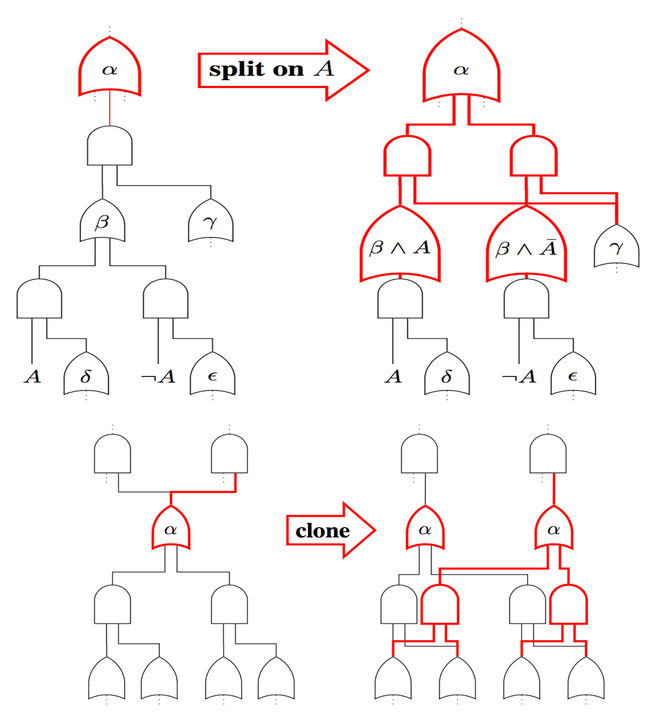


Determinism



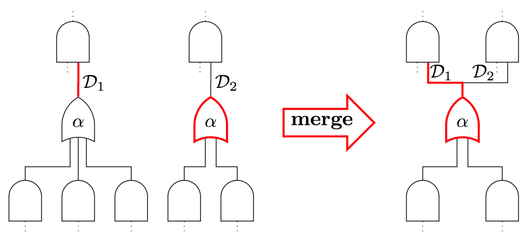
Branching over sentences on prime variables

Search for Structure (LearnPSDD)



$$\text{score} = \frac{\ln \mathcal{L}(r' \mid \mathcal{D}) - \ln \mathcal{L}(r \mid \mathcal{D})}{\text{size}(r') - \text{size}(r)}$$

Future Work (MergePSDD)



Experiments

Roadmap

1. Vtree learning
2. Construct the most naïve PSDD
3. LearnPSDD (search for better structure)

LearnPSDD

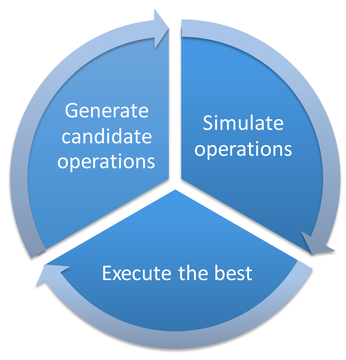
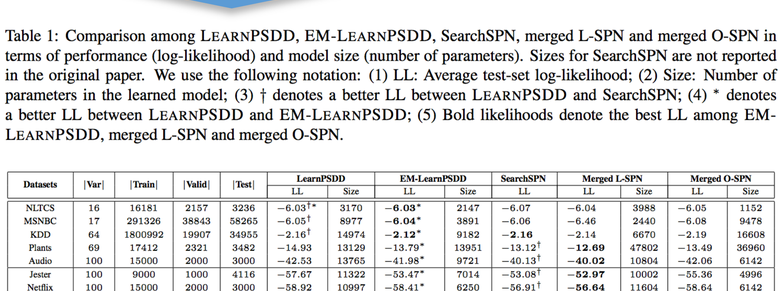


Table 1: Comparison among LearnPSDD, EM-LearnPSDD, SearchSPN, merged L-SPN and merged O-SPN in terms of performance (log-likelihood) and model size (number of parameters). Sizes for SearchSPN are not reported in the original paper. We use the following notation: (1) LL: Average test-set log-likelihood; (2) Size: Number of parameters in the learned model; (3) † denotes a better LL between LearnPSDD and SearchSPN; (4) * denotes a better LL between LearnPSDD and EM-LearnPSDD; (5) Bold likelihoods denote the best LL among EM-LearnPSDD, merged L-SPN and merged O-SPN.

Datasets	|Var|	|Train|	|Valid|	|Test|	LearnPSDD LL	LearnPSDD Size	EM-LearnPSDD LL	EM-LearnPSDD Size	SearchSPN LL	Merged L-SPN LL	Merged L-SPN Size	Merged O-SPN LL	Merged O-SPN Size
NLTCS	16	16181	2157	3236	−6.03†*	3170	**−6.03***	2147	−6.07	−6.04	3988	−6.05	1152
MSNBC	17	291326	38843	58265	−6.05†	8977	**−6.04***	3891	−6.06	−6.46	2440	−6.08	9478
KDD	64	1800992	19907	34955	−2.16†	14974	**−2.12***	9182	**−2.16**	−2.14	6670	−2.19	16608
Plants	69	17412	2321	3482	−14.93	13129	−13.79*	13951	−13.12†	**−12.69**	47802	−13.49	36960
Audio	100	15000	2000	3000	−42.53	13765	−41.98*	9721	−40.13†	**−40.02**	10804	−42.06	6142
Jester	100	9000	1000	4116	−57.67	11322	−53.47*	7014	−53.08†	**−52.97**	10002	−55.36	4996
Netflix	100	15000	2000	3000	−58.92	10997	−58.41*	6250	−56.91†	**−56.64**	11604	−58.64	6142
Accidents	111	12758	1700	2551	−34.13	10489	−33.64*	6752	−30.02†	**−30.01**	13322	−30.83	6846
Retail	135	22041	2938	4408	−11.13	4091	**−10.81***	7251	−10.97†	−10.87	2162	−10.95	3158
Pumsb-Star	163	12262	1635	2452	−34.11	10489	−33.67*	7965	−28.69†	**−24.11**	17604	−24.34	18338
DNA	180	1600	400	1186	−89.11*	6068	−92.67	14864	−81.76†	**−85.51**	4320	−87.49	1430
Kosarek	190	33375	4450	6675	−10.99†	11034	−10.81*	10179	−11.00	**−10.62**	5318	−10.98	6712
MSWeb	294	29441	32750	5000	−10.18†	11389	−9.97*	14512	−10.25	**−9.90**	16484	−10.06	12770
Book	500	8700	1159	1739	−35.90	15197	−34.97*	11292	−34.91†	**−34.76**	11998	−37.44	11916
EachMovie	500	4524	1002	591	−56.43*	12483	−58.01	16074	−53.28†	**−52.07**	15998	−58.05	19846
WebKB	839	2803	558	838	−163.42	10033	−161.09*	18431	−157.88†	**−153.55**	20134	−161.17	10046
Reuters-52	889	6532	1028	1530	−94.94	10585	−89.61*	9546	−86.38†	**−83.90**	46232	−87.49	28334
20NewsGrp.	910	11293	3764	3764	−161.41	12222	−161.09*	18431	−153.63†	**−154.67**	43684	−161.46	29016
BBC	1058	1670	225	330	−260.83	10585	**−253.19***	20327	−252.13†	−253.45	21160	−260.59	8454
AD	1556	2461	327	491	−30.49*	9666	−31.78	9521	−16.97†	−16.77	49790	**−15.39**	31070

Compare with O-SPN: smaller size in 14, better LL in 11, win on both in 6

Compare with L-SPN: smaller size in 14, better LL in 6, win on both in 2

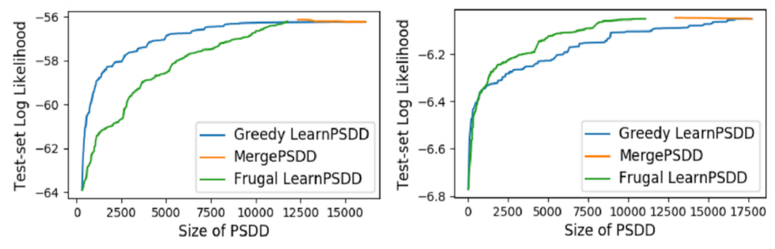


EM/Bagging

Datasets	Var	LearnPSDD Ensemble	Best-to-Date
NLTCS	16	−5.99†	−6.00
MSNBC	17	−6.04†	−6.04†
KDD	64	−2.11†	−2.12
Plants	69	−13.02	−11.99†
Audio	100	−39.94	−39.49†
Jester	100	−51.29	−41.11†
Netflix	100	−55.71†	−55.84
Accidents	111	−30.16	−24.87†
Retail	135	−10.72†	−10.78
Pumsb-Star	163	−26.12	−22.40†
DNA	180	−88.01	−80.03†
Kosarek	190	−10.52†	−10.54
MSWeb	294	−9.89	−9.22†
Book	500	−34.97	−30.18†
EachMovie	500	−58.01	−51.14†
WebKB	839	−161.09	−150.10†
Reuters-52	889	−89.61	−80.66†
20NewsGrp.	910	−155.97	−150.88†
BBC	1058	−253.19	−233.26†
AD	1556	−31.78	−14.36†

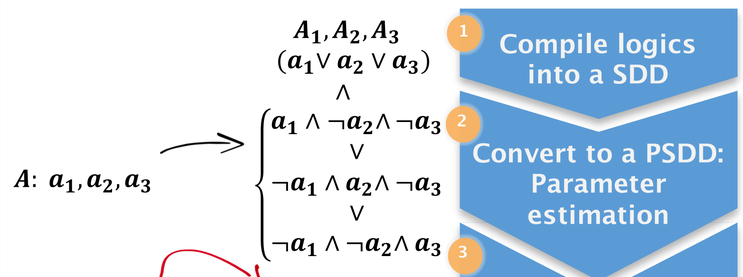
State-of-the-art in 6 datasets

Merging Without Performance Loss (Future Work)



Retain the ability to fit logically constrained distributions

Discrete Multi-valued Data



Datasets	No Constraint	PSDD	LearnPSDD
Adult	−18.41	−14.14	−12.86
CovType	−14.39	−8.81	−7.32

Acknowledgements: we thank Arthur Choi and Yujia Shen for helpful discussions. This research was conducted while JB was a visiting student at UCLA. JB is supported by IWT (SB/141744). This work is partially supported by NSF grants #IIS-1657613, #IIS-1633857 and DARPA XAI grant #N66001-17-2-4032.