Learning Credal Sum-Product Networks

AKBC, 2020

Amelie Levray, Vaishak Belle

University of Edinburgh

Learning probabilistic models

Tractable learning

- Thin junction trees
- Approximate inference
- Sum product networks ullet
- Sentential decision diagrams
- Relational SPNs

Missing data & incomplete knowledge

Imprecise probabilities

- Imputation schemes
- Upper & lower measures
- <u>Credal semantics</u>
- Fuzzy semantics
- Possibilistic semantics

Motivation: credal semantics sets of conditional probability measures associated with variables

- to predictions that are **unreliable** and **overconfident**
- not certain (so adjust confidence of prediction)
- contamination strategies for **robustness**
- interval [Ceylan et al. 2016]

Learning single distributions in the presence of incomplete knowledge may led

Added semantic value of knowing that the distribution of a random variable is

Conversely, if pre-trained model known to have used missing data, consider

• In probabilistic KBs, extracted tuples accorded confidence probabilities, but other atoms assumed to be impossible. What if new source has that atom? Can't meaningfully provide posterior probability. Let atoms be accorded an

Credal SPNs Inference polynomial in network size



$$S(x_1, x_2, \overline{x_1}, \overline{x_2}) = .7(.4x_1 + .6\overline{x_1})(.9x_2 + .1\overline{x_2}) + .3(.2x_1 + .8\overline{x_1})(.9x_2 + .1\overline{x_2})$$

A bottom up pass for conditional probabilities



$$\left\{ \begin{array}{l} .1 \leq w_1 \leq .5, .5 \leq w_2 \leq .9, \\ .57 \leq w_5 \leq .84, .16 \leq w_6 \leq .43, \\ (w_3, w_4) \in CH((.3, .7), (.4, .6)), \\ (w_7, w_8) \in CH((.22, .78), (.5, .5)), \\ w_1 + w_2 = 1, w_5 + w_6 = 1 \end{array} \right\}$$

- Weights are sets of extreme points [Maua et al. 2017]
- Our objective: with missing data,

can we learn CSPNs effectively?

Recursive scheme for learning SPNs Two metrics: complete vs possible completions



			V	ar	ia	ы	es	
1								
1		•		\vdash				ſ
1	50							
0	tar							
0	500							
1	2	2						
1								
1								l

G

instance number

2

3

5

6 7

8

. . .

- 2



Empirical results More compact models, better likelihood results

	Log Likelihood score + time (in s) / size (number of nodes)							
Dataset	0%	missing values on 1% of inst			missing values on 5% of inst			
	LearnSPN	avg_ll	min_11	opt_ll	avg_ll	min_ll	opt_11	
NLTCS	-6.114	-6.194	-6.981	-6.111	-6.352	-8.556	-5.308	
	47.981 / 2584	932.832 / 53811			891.506 / 5458			
Plants	-12.988	-15.042 -18.753 -13.052		-17.397 -24.900 -14.581				
	137.057 / 20714	662.244 / 8661			4387.299 / 14281			
Audio	-40.528	-40.587	-42.832	-33.862	-41.774	-45.178	-33.129	
	770.684 / 35491	7307.784 / 72822		1231.921 / 2021				
Jester	-53.444	-53.351	-54.934	-48.412	-54.188	-58.543	-48.092	
Jester	1082.599 / 47769	4143.19 / 87467		13199.7 / 101708				
Netflix	-57.408	-57.453	-59.337	-53.371	-58.817	-62.860	-53.544	
	1145.967 / 39447	5682.157 / 43229		16367.691 / 37674				
	-11.131	-11.095	-14.363	-8.114	-11.129	-20.745	-6.260	
Retail	83.637 / 3790	1631.054 / 6335			5701.983 / 6937			
Dumch stor	-25.548	-28.529	-36.434	-30.197	-35.783	-50.307	-30.787	
Pumso-star	431.414 / 73949	4154.584 / 347880			1606.189 / 2128			
DNA	-85.272	-85.596	-93.196	-77.174	-98.039	-102.135	-81.192	
DNA	66.37 / 22627	220.202 / 18957		146.218 / 5069				
MSWeb	-10.212	-10.284	-19.722	-7.344	-10.493	-34.322	-6.753	
	175.64 / 11000	12035.587 / 9145		26367.275 / 3836				
Book	-36.656	-34.903	-42.475	-21.446	-35.112	-60.904	-17.330	
	304.339 / 95077	5245.763 / 46970		17944.24 / 68137				
EachMowie	-52.836	-53.569	-64.057	-43.548	-57.047	-98.468	-41.812	
Eachiviovie	229.943 / 86066	1732.422 / 16808		4508.381 / 31141				
WebKB	-158.696	-156.738	-164.892	-132.033	-159.509	-186.550	-109.624	
	493.989 / 241777	25	18.345 / 210	75	33	39.987 / 22 0)66	
Reuters-52	-85.995	-88.537	-111.189	-73.106	-92.910	-122.414	-73.259	
	1507.077 / 427220	11128.818 / 22511		18168.615 / 4768				
20 Newsgrp.	-159.701	-153.567	-166.219	-99.073	-155.182	-188.189	-92.349	
	20656.808 / 3047296	28115.16 / 154923		122947.065 / 197478				
BBC	-248.931	-258.220	-265.545	-195.062	-258.727	-291.568	-197.744	
	883.801 / 290423	2623.016 / 62194		1824.548 / 9976				
Ad	-27.298	-31.963	-87.751	-28.097	-46.117	-98.332	-31.267	
	10197.343 / 690272	94	36.839 / 595	582	93	383.391 / 97	72	



- Blue: SPNs
- Red: 1% missing data CSPNs
- Brown: 5% missing data CSPNs
- Paper also studies theoretical properties
- & complexity results