Learning Accurate Cutset Networks by Exploiting Decomposability

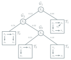
N. Di Mauro, A. Vergari, and F. Esposito

Department of Computer Science, LACAM Laboratory
University of Bari "Aldo Moro", Italy

14th Conference of the Italian Association for Artificial Intelligence

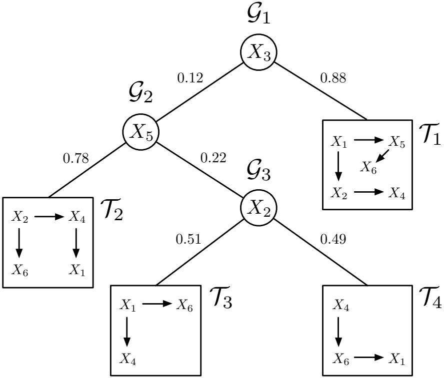
Introduction

Tractable Probabilistic Graphical Models



- Probabilistic Graphical Models
  - powerful formalism to model rich and structured domains
  - capture independences among random variables into a graph
  - computing exact inference in PGMs is a NP-Hard problem
- Tractable Probabilistic Graphical Models
  - provide exact and efficient inference but less expressive
  - tree-structured models, Bayesian and Markov Networks compiled into Arithmetic Circuits, and Sum-Product Networks
- Cutset Networks
  - weighted probabilistic model trees
  - OR-trees having tree-structured models as leaves
  - non-negative weights on inner edges
  - Inner nodes, i.e., conditioning OR nodes, are associated to random variables and outgoing branches represent conditioning on the values for those variables domains.

Cutset Networks



Given $\mathbf{X}$ be a set of discrete variables, a *CNet* is defined as follows:

1. a CLtree, with scope $\mathbf{X}$, is a CNet;
2. given $X_i \in \mathbf{X}$ a variable with $|Val(X_i)| = k$, graphically conditioned in an OR node, a weighted disjunction of k CNets $\mathcal{G}_i$ with same scope $\mathbf{X}_{\setminus i}$ is a CNet, where all weights $w_{i,j}$, $j = 1, \ldots, k$, sum up to one, and $\mathbf{X}_{\setminus i}$ denotes the set $\mathbf{X}$ minus the variable X_i.

Contribution

dCSN

The dCSN algorithm

- avoiding decision tree heuristics
  - choosing the best variable directly maximizing the log-likelihood
- complex structures penalized adopting the BIC

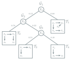
$$\text{score}_{\text{BIC}}(\langle \mathcal{G}, \boldsymbol{\gamma} \rangle) = \log P_{\mathcal{D}}(\langle \mathcal{G}, \boldsymbol{\gamma} \rangle) - \frac{\log M}{2} \text{Dim}(\mathcal{G})$$

- Bagging in order to obtain a mixture of CNets
  - k bootstrapped samples $\mathcal{D}_i$ from the dataset $\mathcal{D}$
    - leading to k CNets $\mathcal{G}_i$
  - resulting bagged CNet $\mathcal{G}$ set to a weighted sum of CNets $\mathcal{G}_i$

$$\hat{P}(\xi : \mathcal{G}) = \sum_{i=1}^{k} \mu_i P(\xi : \mathcal{G}_i),$$

where $\mu_i = \ell_{\mathcal{D}}(\langle \mathcal{G}_i, \boldsymbol{\gamma}_i \rangle) / \sum_{j=1}^{k} \ell_{\mathcal{D}}(\langle \mathcal{G}_j, \boldsymbol{\gamma}_j \rangle)$

Cutset Networks



Proposition 1 (CNet log-likelihood decomposition)

$$\ell_{\mathcal{D}}(\langle \mathcal{G}, \boldsymbol{\gamma} \rangle) = \sum_{\xi \in \mathcal{D}} \sum_{i=1,\ldots,n} \log P(\xi[X_i] | \xi[\mathrm{Pa}_i]) \tag{1}$$

$$\ell_{\mathcal{D}}(\langle \mathcal{G}, \boldsymbol{\gamma} \rangle) = \sum_{j=1,\ldots,k} M_j \log w_{i,j} + \ell_{\mathcal{D}_j}(\langle \mathcal{G}_j, \boldsymbol{\gamma}_{\mathcal{G}_j} \rangle) \tag{2}$$

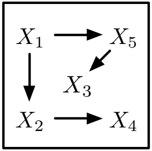
Proposition 2 (BIC decomposition)

$$\ell_{\mathcal{D}_l}(\langle \mathcal{G}_i, \boldsymbol{\gamma}_i \rangle) - \ell_{\mathcal{D}_l}(\langle \mathcal{T}_l, \boldsymbol{\theta}_l \rangle) > \frac{\log M}{2} \tag{3}$$

- instead of recomputing the likelihood on the complete dataset $\mathcal{D}$ we can evaluate only the local improvement
- the decomposition of $\mathcal{T}_l$ is independent from all other $\mathcal{T}_j, j \neq l$ being their local contributions to the global log-likelihood independent
  - it is not significant the order we choose to decompose leaf nodes

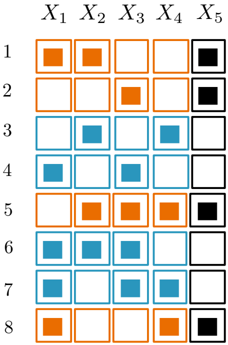
dCSN example I

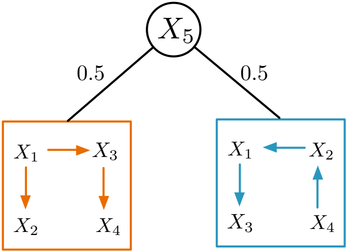
	X_1	X_2	X_3	X_4	X_5
1	■	■			■
2			■		■
3		■		■	
4	■		■		
5		■	■	■	■
6	■	■	■		
7	■		■	■	
8	■			■	■



- starting with a single CLTree for all variables X_1, X_2, X_3, X_4, X_5

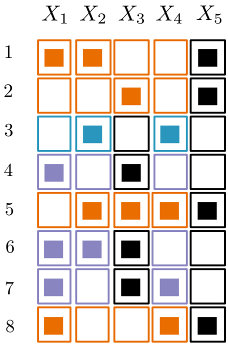
dCSN example II

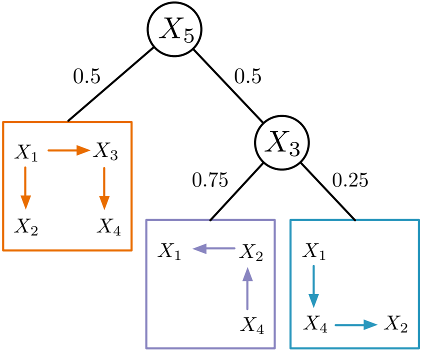




- checking whether there is a decomposition
  - adding OR node on variable X_5 applied on two CLtrees with higher ll

dCSN example III





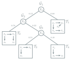
- recursively apply the decomposition process
  - adding OR node on variable X_3 applied on two CLtrees with higher ll

Experiments

Empirical risk for all algorithms

	CNet	CNetP	dCSN	CNet-B	CNetP-B	dCSN-B	MT	MCNet
NLTCS	-6.11	-6.06	**-6.04**	-6.09	-6.02	-6.02	**-6.01**	-6.00
MSNBC	-6.06	**-6.05**	-6.05	-6.06	**-6.04**	-6.04	-6.08	-6.04
Plants	**-13.24**	-13.25	-13.35	-12.30	-12.38	**-12.21**	-12.93	-12.78
Audio	-44.58	**-42.05**	**-42.06**	-42.09	-40.71	**-40.17**	**-40.14**	-39.73
Jester	-61.71	**-55.56**	**-55.30**	-57.76	-53.17	**-52.99**	-53.06	-52.57
Netflix	-65.61	**-58.71**	**-58.57**	-63.08	-57.63	**-56.63**	-56.71	-56.32
Accidents	-30.97	-30.69	**-30.17**	-30.25	-30.28	**-28.99**	-29.69	-29.96
Retail	-11.07	**-10.94**	-11.00	-10.99	-10.88	-10.87	**-10.84**	-10.82
Pumsb-star	-24.65	-24.42	**-23.83**	-24.39	-24.19	**-23.32**	-23.70	-24.18
DNA	-90.48	-87.59	**-87.19**	-90.66	-86.85	**-84.93**	**-85.57**	-85.82
Kosarek	-11.19	**-11.04**	-11.14	-10.97	-10.85	-10.85	**-10.62**	-10.58
MSWeb	-10.07	-10.07	**-9.94**	-9.95	-9.91	-9.86	**-9.82**	-9.79
Book	-37.62	-37.35	**-37.22**	-35.88	-35.62	-35.92	**-34.69**	-33.96
EachMovie	**-59.19**	**-58.37**	**-58.47**	-54.22	-54.02	**-53.91**	-54.51	-51.39
WebKB	-162.85	-162.17	**-161.16**	-156.79	-156.94	**-155.20**	-157.00	-153.22
Reuters-52	**-88.72**	**-88.55**	**-88.60**	-86.22	-86.89	**-85.69**	-86.53	-86.11
BBC	**-262.08**	**-263.08**	**-262.08**	**-252.01**	-257.72	**-251.14**	-259.96	-250.58
Ad	-16.92	-16.92	**-14.81**	-15.94	-16.02	**-13.73**	-16.01	-16.68

Conclusions



- a new approach to learn the structure of CNets model
  - exploiting the decomposable score and maximizing the likelihood
  - formulating a score including the BIC criterion
  - introducing informative priors on smoothing parameters
- mixtures of CNets with bagging as an alternative to EM
- evaluation on standard benchmarks proving the validity of our claims

Future Work

- latent nodes such as in latent tree models
- (gradient) boosting

Code available at `http://www.di.uniba.it/~ndm/dcsn/`