Abductive Reasoning in Logical Credal Networks

Radu Marinescu, Junkyu Lee, Debarun Bhattacharjya, Fabio Cozman, Alexander Gray

NeurIPS-2024

Motivation and Contribution

- Logical Credal Networks or LCNs
  - Many (if not all) real-world applications require:
    - Efficient handling of uncertainty
    - Compact representations of a wide variety of knowledge
  - Logical Credal Networks – a novel probabilistic logic:
    - Allows marginal and conditional probability bounds on logic formulas
    - Markov condition: additional independence assumptions between atoms
    - Exact and approximate marginal inference
- Abductive Reasoning in LCNs
  - MAP and Marginal MAP inference in LCNs (explanations)
  - Exact and approximate MAP/MMAP inference algorithms
  - Promising experimental results on synthetic and realistic benchmarks

Logical Credal Networks

[Marinescu et al, NeurIPS 2022]

- A set of ***probability-labeled sentences*** of the following form:

$$l_q \leq P(q) \leq u_q$$

$$l_{q|r} \leq P(q|r) \leq u_{q|r}$$

  – where q and r can be arbitrary propositional or first-order logic* formulas, l_q and u_q (resp., $l_{q|r}$ and $u_{q|r}$) are lower and upper probability bounds
  – a label $\tau \in \{yes, no\}$ indicates independence between the atoms in q (details to follow)

- Represents set of probability distributions over all interpretations satisfying **LCN's constraints**

Friends of friends are likely friends. If two people are friends, they likely either both smoke or neither does. Smoking likely causes cancer.

$$\begin{aligned}
0.5 &\leq P\left(\mathit{Friends}(x,z) \mid \mathit{Friends}(x,y) \wedge \mathit{Friends}(y,z)\right) \leq 1, \\
0 &\leq P\left(\neg(\mathit{Smokes}(x) \oplus \mathit{Smokes}(y)) \mid \mathit{Friends}(x,y)\right) \leq 0.2, \\
0.03 &\leq P\left(\mathit{Cancer}(x) \mid \mathit{Smokes}(x)\right) \leq 0.04, \\
0 &\leq P\left(\mathit{Cancer}(x) \mid \neg\mathit{Smokes}(x)\right) \leq 0.01,
\end{aligned}$$

*universally quantified FOL with finite domains of values

Example

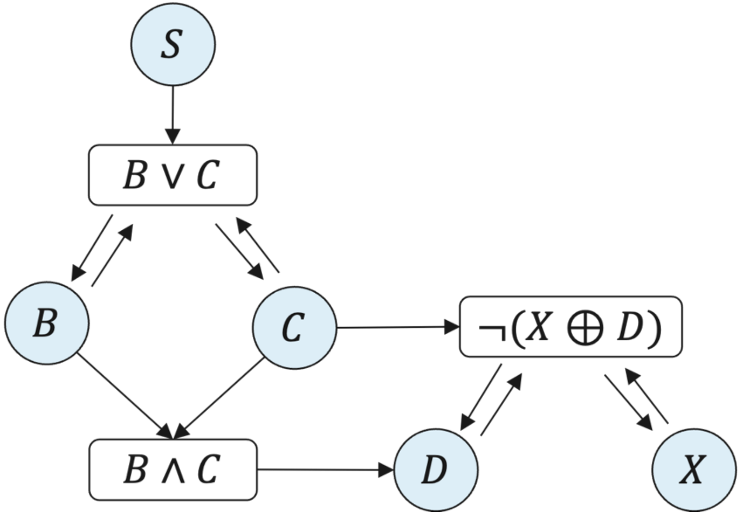
$$0.05 \leq P(B) \leq 0.1$$

$$0.3 \leq P(S) \leq 0.4$$

$$0.1 \leq P(B \vee C | S) \leq 0.2$$

$$0.6 \leq P(D | B \wedge C) \leq 0.7$$

$$0.7 \leq P(\neg(X \oplus D) | C) \leq 0.8$$



- Bronchitis (B) is more likely than Smoking (S); Smoking may cause Cancer (C) or Bronchitis; Dyspnea (D) or shortness of breath is a common symptom for Cancer and Bronchitis; in case of Cancer we have either a positive X-Ray result (X) and Dyspnea, or a negative X-Ray and no Dyspnea. The figure above shows the primal graph where the formula and proposition nodes are displayed as rectangles and shaded circles, respectively.

Local Markov Condition

- Let $\mathcal{L}$ be an LCN, and $\mathcal{M}$ be a model* of $\mathcal{L}$. Given M, every atom x is conditional independent of its non-descendant non-parent (**ndnp**) atoms given its parents in the primal graph G of $\mathcal{L}$.
- We are now ready to make quantitative commitments (i.e., explicit independence assumptions)
- Let x be an atom, $S_x = \{s_1, \dots, s_k\}$ and $T_x = \{t_1, \dots, t_l\}$ be its **parents** and **ndnp's** sets in G
  - we assert $P(x|S_x, T_x) = P(x|S_x)$ or equivalently $P(x, S_x, T_x) \cdot P(S_x) = P(x, S_x) \cdot P(S_x, T_x)$

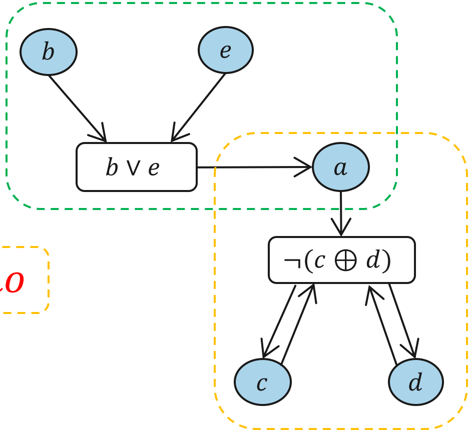
$0.1 \leq P(b) \leq 0.2$

$0.05 \leq P(e) \leq 0.1$

$0.8 \leq P(a|b \vee e) \leq 0.9$

$0.7 \leq P(\neg(c \oplus d)|a) \leq 0.8$ $\tau = no$

$0.01 \leq P(a) \leq 0.08$



b and e are independent
c is CI of $\{b, e\}$ given $\{a, d\}$
d is CI of $\{b, e\}$ given $\{a, c\}$

$P(b|e) = P(b)$

$P(c|b, e, a, d) = P(c|a, d)$

$P(d|b, e, a, c) = P(d|a, c)$

Marginal Inference in LCNs

- Given a query formula ρ, compute posterior lower and upper bounds on $P(\rho)$
  - Let x be an atom, $S_x = \{s_1, \dots, s_k\}$ and $T_x = \{t_1, \dots, t_l\}$ be its **parents** and **ndnp's** sets in G we assert $P(x|S_x, T_x) = P(x|S_x)$ or equivalently $P(x, S_x, T_x) \cdot P(S_x) = P(x, S_x) \cdot P(S_x, T_x)$

n atoms, $N = 2^n$ interpretations (worlds)
$\vec{p} = (p_1, \dots, p_N)$ probability vector

$\vec{A}_q = (a_1, \dots, a_N)$ indicator vector,
$a_i = 1$ if q is true in i^{th} interpretation
$a_i = 0$, otherwise

$\odot$ is the dot-product of two vectors

$$\sum_{i=1}^{N} p_i = 1$$

$$p_i \geq 0, \forall i = 1, \dots, N$$

$l_q \leq P(q) \leq u_q \longrightarrow l_q \leq \vec{A}_q \cdot \vec{p} \leq u_q$

$l_{q|r} \leq P(q|r) \leq u_{q|r} \longrightarrow l_{q|r} \cdot \vec{A}_r \odot \vec{p} \leq \vec{A}_{q \wedge r} \odot \vec{p} \leq u_{q|r} \cdot \vec{A}_r \odot \vec{p}$

$P(x, S_x, T_x) \cdot P(S_x) - P(x, S_x) \cdot P(S_x, T_x) = 0 \longrightarrow (\vec{A}_\alpha \odot \vec{p}) \cdot (\vec{A}_\beta \odot \vec{p}) - (\vec{A}_\gamma \odot \vec{p}) \cdot (\vec{A}_\delta \cdot \vec{p}) = 0$

minimize/maximize $\vec{A}_\rho \odot \vec{p}$

$\alpha = x \wedge s_1 \wedge \dots \wedge s_k \wedge t_1 \wedge \dots \wedge t_l$ $\quad \gamma = x \wedge s_1 \wedge \dots \wedge s_k$
$\beta = s_1 \wedge \dots \wedge s_k$ $\quad \delta = s_1 \wedge \dots \wedge s_k \wedge t_1 \wedge \dots \wedge t_l$

MAP and Marginal MAP Inference

[Marinescu et al, NeurIPS 2024]

- Up until now, focus was on computing a probability interval $P(q) \in [\underline{P}(q), \overline{P}(q)]$ for a given query formula q
- MAP/MMAP inference calls for finding the most probable (complete or partial) explanation of observed evidence in an LCN

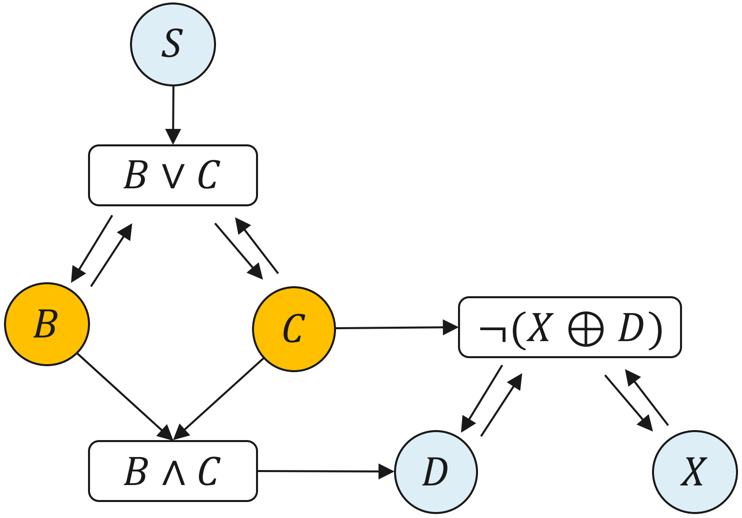
$$0.05 \leq P(B) \leq 0.1$$

$$0.3 \leq P(S) \leq 0.4$$

$$0.1 \leq P(B \vee C | S) \leq 0.2$$

$$0.6 \leq P(D | B \wedge C) \leq 0.7$$

$$0.7 \leq P(\neg(X \oplus D) | C) \leq 0.8$$



Evidence: $\neg B, C$ ➡ MAP: S, D, X

MAP and MMAP Tasks in LCNs

- Define maximin and maximax MAP/MMAP tasks as follows:

Definition 3 (maximin). *Given an LCN $\mathcal{L}$ with n propositions, evidence* $\mathbf{e}$*, and MAP propositions* $\mathbf{Y}$*, the* maximin MAP *(or* maximin MMAP *if $m < n - k$) task is finding a truth assignment* $\mathbf{y}^*$ *to* $\mathbf{Y}$ *having maximum* lower probability, *given evidence* $\mathbf{e}$*, namely:*

$$\mathbf{y}^* = \operatorname*{argmax}_{\mathbf{y} \in \Omega(\mathbf{Y})} \underline{P}(\psi_{\mathbf{y} \wedge \mathbf{e}}) \quad (9)$$

where $\Omega(\mathbf{Y})$ is the set of all truth assignments to the MAP propositions, and $\psi_{\mathbf{y} \wedge \mathbf{e}} = y_1 \wedge \dots \wedge y_m \wedge e_1 \wedge \dots \wedge e_k$ is the conjunction of the literals in $\mathbf{y}$ *and* $\mathbf{e}$*, respectively.*

Definition 4 (maximax). *Given an LCN $\mathcal{L}$ with n propositions, evidence* $\mathbf{e}$*, and MAP propositions* $\mathbf{Y}$*, the* maximax MAP *(or* maximax MMAP *if $m < n - k$) task is finding a truth assignment* $\mathbf{y}^*$ *to* $\mathbf{Y}$ *having maximum* upper probability, *given evidence* $\mathbf{e}$*, namely:*

$$\mathbf{y}^* = \operatorname*{argmax}_{\mathbf{y} \in \Omega(\mathbf{Y})} \overline{P}(\psi_{\mathbf{y} \wedge \mathbf{e}}) \quad (10)$$

where $\Omega(\mathbf{Y})$ is the set of all truth assignments to the MAP propositions, and $\psi_{\mathbf{y} \wedge \mathbf{e}} = y_1 \wedge \dots \wedge y_m \wedge e_1 \wedge \dots \wedge e_k$ is the conjunction of the literals in $\mathbf{y}$ *and* $\mathbf{e}$*, respectively.*

Exact MAP/MMAP Algorithms

$$\mathbf{y}^* = \underset{\mathbf{y}\in\Omega(\mathbf{Y})}{\operatorname{argmax}} \underline{P}(\psi_{\mathbf{y}\wedge\mathbf{e}})$$

- Depth-First Search (DFS)
  - Depth-first search traversal of the search space defined by the MAP propositions Y
  - At each leaf node y, evaluate exactly the lower probability of the query formula $\psi_{y\wedge e}$ i.e., $y_1 \wedge \cdots \wedge y_k \wedge e_1 \wedge \cdots \wedge e_m$
  - Return the optimal configuration $\boldsymbol{y}^*$ **with maximum lower probability.**

- Limited Discrepancy Search (LDS)
  - Depth-first search traversal with max discrepancy $\delta \geq 1$ [Harvey and Ginsberg, 1995]
  - As before, evaluate exactly the lower probability of the query formula $\psi_{y\wedge e}$
  - Return the best configuration $\hat{y}$ found so far
    - If max discrepancy $\delta = |Y|$ then LDS is optimal, namely returns $\boldsymbol{y}^*$ **with maximum lower probability.**

- Simulated Annealing (SA)
  - Stochastic local search-based traversal of the MAP search space
  - For each configuration y evaluate exactly the lower probability of the query formula $\psi_{y\wedge e}$
  - SA converges to the optimal solution if the temperature decays slowly enough

Approximate MAP/MMAP Algorithms

$$\mathbf{y}^* = \underset{\mathbf{y} \in \Omega(\mathbf{Y})}{\operatorname{argmax}} \underline{P}(\psi_{\mathbf{y} \wedge \mathbf{e}})$$

- Approximate MAP (AMAP)
  - Message-passing scheme
  - Extends ARIEL scheme to approximate lower/upper probability of a MAP configuration
  - Using an augmented LCN (more details in [Marinescu et al, NeurIPS 2024])
- Approximate Limited Discrepancy Search (ALDS)
  - Traverse the MAP search space using LDS(δ)
  - Estimate the lower probability of each MAP configuration using AMAP
  - Keep track of the configuration with the largest estimate
- Approximate Simulated Annealing (ASA)
  - Traverse the MAP search space using SA
  - Estimate the lower probability of each MAP configuration using AMAP
  - Keep track of the configuration $\hat{y}$ with the largest estimate
- No guarantees regarding optimality nor we can guarantee lower/upper bounds

Augmented LCN

$0.05 \leq P(B) \leq 0.1$

$0.3 \leq P(S) \leq 0.4$

$0.1 \leq P(B \vee C|S) \leq 0.2$

$0.6 \leq P(D|B \wedge C) \leq 0.7$

$0.7 \leq P(\neg(X \oplus D)|C) \leq 0.8$

$P(W_1|\neg B) = 1$

$P(W_1|B) = 0$

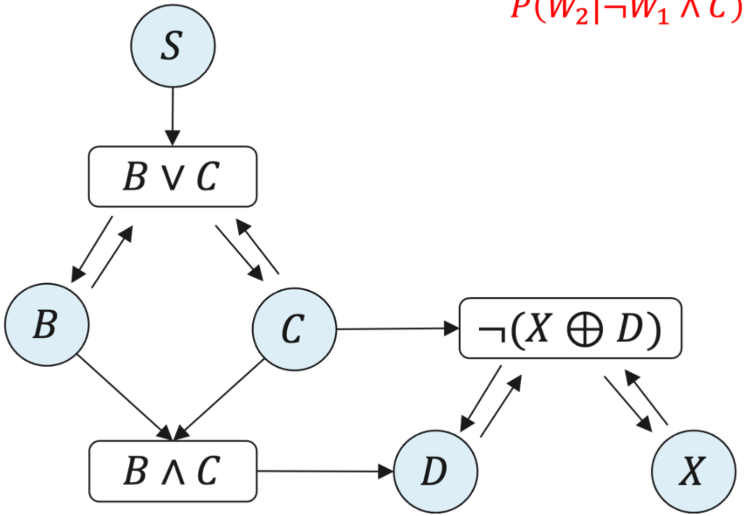
$P(W_2|W_1 \wedge C) = 1$

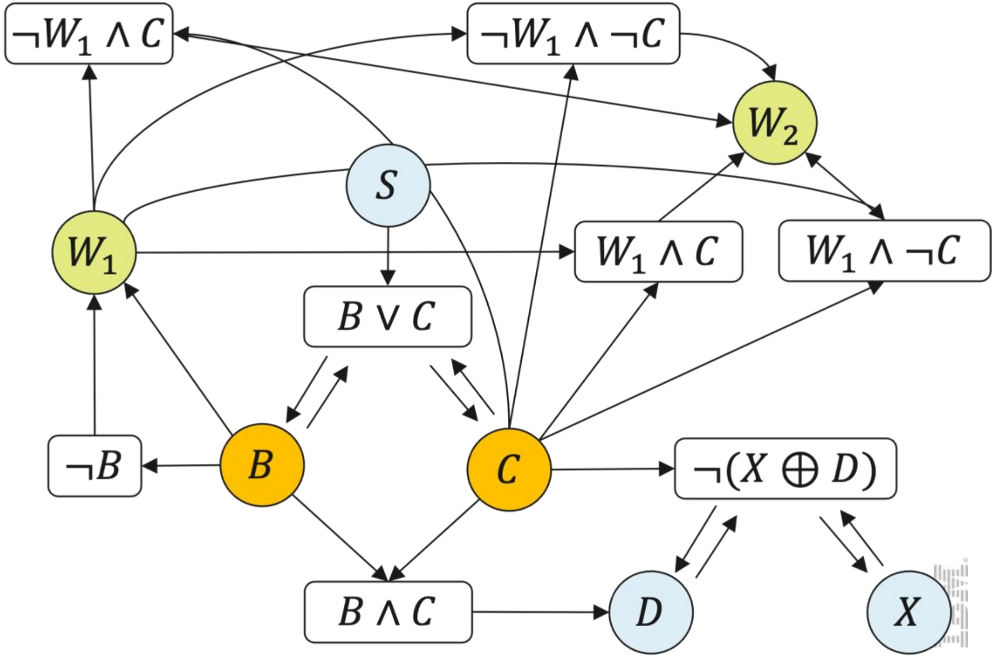
$P(W_2|W_1 \wedge \neg C) = 0$

$P(W_2|\neg W_1 \wedge \neg C) = 0$

$P(W_2|\neg W_1 \wedge C) = 0$

Evidence: $\neg B, C$





MAP/MMAP Inference in LCNs

Exact	Approximate
Depth-First Search	Iterative Message Passing
Limited Discrepancy Search	Limited Discrepancy Search
Simulated Annealing	Simulated Annealing
Exact evaluation of the MAP configurations	Approximate evaluation of the MAP configurations

Outline

- Motivation and Contribution
- Logical Credal Networks
- Marginal and MAP/MMAP Inference
- Experimental Results
- Conclusion

Code* available at: http://github.com/IBM/LCN

Results – MAP Inference

[Marinescu et al., NeurIPS 2024]

- Benchmark problems:
  - Random LCNs
  - LCNs derived from real-world Bayesian networks

Real-world LCNs

LCN	exact MAP eval				approx MAP eval	
	DFS	LDS(3)	SA	AMAP	ALDS(3)	ASA
Toy	2.20	3.18	1.85	0.85	134.83	141.17
Earth	9.19	7.67	2.75	1.28	150.99	162.35
Cancer	16.34	14.09	8.52	2.64	157.92	159.66
Asia	811.82	800.18	312.10	4.07	187.44	201.76
Credit	-	6719.30	2976.55	5.09	204.77	222.52
Engine	4786.12	4502.34	2033.77	6.57	212.61	235.70
Suicide	-	-	-	5.99	220.31	203.68
Tank	-	-	-	8.04	263.65	281.73
Alarm	-	-	-	4.28	216.19	186.67
Hepatitis	-	-	-	8.22	260.38	250.45

Conclusion

- New probabilistic logic that expresses probability bounds for propositional and first-order logic formulas with few restrictions
- Local Markov condition (like in probabilistic graphical models) allows making additional independence assumptions
  - Restricts the space of probability distributions to enable meaningful representations of uncertainty
- Exact inference (marginal, MAP) to answer queries for the new formalism
  - Involves the solution of a non-linear constraint program
- Approximate inference (marginal, MAP) scales to larger problems
  - Involves message-passing along the edges of a factor graph
- Empirical evaluations on random problems and more realistic applications shows promising results, particularly in aggregating multiple sources of knowledge

