Scaling up Hybrid Probabilistic Inference with Logical and Arithmetic Constraints via Message Passing

Zhe Zeng*
University of California, Los Angeles

Paolo Morettin*
University of Trento, Italy

Fanqi Yan*
AMSS, Chinese Academy of Sciences

Antonio Vergari
University of California, Los Angeles

Guy Van den Broeck
University of California, Los Angeles
Scaling up Hybrid Probabilistic Inference with Logical and Arithmetic Constraints via Message Passing

Zhe Zeng*
University of California, Los Angeles

Paolo Morettin*
University of Trento, Italy

Fanqi Yan*
AMSS, Chinese Academy of Sciences

Antonio Vergari
University of California, Los Angeles

Guy Van den Broeck
University of California, Los Angeles
Scaling up Hybrid Probabilistic Inference with Logical and Arithmetic Constraints via Message Passing

Zhe Zeng*
University of California, Los Angeles

Paolo Morettin*
University of Trento, Italy

Fanqi Yan*
AMSS, Chinese Academy of Sciences

Antonio Vergari
University of California, Los Angeles

Guy Van den Broeck
University of California, Los Angeles
Scaling up Hybrid Probabilistic Inference with Logical and Arithmetic Constraints via Message Passing

Zhe Zeng*
University of California, Los Angeles

Paolo Morettin*
University of Trento, Italy

Fanqi Yan*
AMSS, Chinese Academy of Sciences

Antonio Vergari
University of California, Los Angeles

Guy Van den Broeck
University of California, Los Angeles
Skill matching system

Minka et al., “Trueskill 2: An improved bayesian skill rating system”, 2018
Skill matching system

Each player has a certain skill

Minka et al., “Trueskill 2: An improved bayesian skill rating system”, 2018
Skill matching system

Each player has a certain skill

\[ \text{continuous variables} \]

Minka et al., “Trueskill 2: An improved bayesian skill rating system”, 2018
Skill matching system

Each player has a certain skill
Players can form teams

Minka et al., “Trueskill 2: An improved bayesian skill rating system”, 2018
Skill matching system

- Each *player* has a certain skill
- Players can form **teams**

⇒ *intricate dependencies*

---

Minka et al., “Trueskill 2: An improved bayesian skill rating system”, 2018
Skill matching system

- Each player has a certain skill
- Players can form teams
- Each team’s skill is bounded by its players’ skills

Minka et al., “Trueskill 2: An improved bayesian skill rating system”, 2018
Skill matching system

Each player has a certain skill
Players can form teams
Each team’s skill is bounded by its players’ skills ⇒ complex constraints!

Minka et al., “Trueskill 2: An improved bayesian skill rating system”, 2018
Each player has a certain skill
Players can form teams
Each team’s skill is bounded by its players’ skills
Good teams form a squad

Minka et al., “Trueskill 2: An improved bayesian skill rating system”, 2018
Skill matching system

- Each player has a certain skill
- Players can form teams
- Each team’s skill is bounded by its players’ skills
- Good teams form a squad

Minka et al., “Trueskill 2: An improved bayesian skill rating system”, 2018
“What is the probability of team $T_1$ to outperform team $T_2$, if $T_1$ is a squad but $T_2$ is not?”

Minka et al., “Trueskill 2: An improved bayesian skill rating system”, 2018
Continuous + discrete + constraints = ?
Continuous + discrete + constraints = ?

Generative adversarial networks (GANs) [Goodfellow et al. 2014]
Variational Autoencoders (VAEs) [Kingma et al. 2013]
Continuous + discrete + constraints = ?

Generative adversarial networks (GANs) [Goodfellow et al. 2014]
Variational Autoencoders (VAEs) [Kingma et al. 2013]

⇒ limited inference capabilities, no constraints
Continuous + discrete + constraints = ?

Generative adversarial networks (GANs) [Goodfellow et al. 2014]

Variational Autoencoders (VAEs) [Kingma et al. 2013]

Hybrid Bayesian Networks (HBNs) [Heckerman et al. 1995; Shenoy et al. 2011]

Mixed Probabilistic Graphical Models (MPGMs) [Yang et al. 2014]
Continuous + discrete + constraints = ?

Generative adversarial networks (GANs) [Goodfellow et al. 2014]

Variational Autoencoders (VAEs) [Kingma et al. 2013]

Hybrid Bayesian Networks (HBNs) [Heckerman et al. 1995; Shenoy et al. 2011]

Mixed Probabilistic Graphical Models (MPGMs) [Yang et al. 2014]

⇒ strong distributional assumptions
Continuous + discrete + constraints = ?

Generative adversarial networks (GANs) \cite{Goodfellow14}

Variational Autoencoders (VAEs) \cite{Kingma13}

Hybrid Bayesian Networks (HBNs) \cite{Heckerman95, Shenoy11}

Mixed Probabilistic Graphical Models (MPGMs) \cite{Yang14}

Tractable Probabilistic Circuits (PCs) \cite{Molina18, Vergari19}
Continuous + discrete + constraints = ?

Generative adversarial networks (GANs) [Goodfellow et al. 2014]

Variational Autoencoders (VAEs) [Kingma et al. 2013]

Hybrid Bayesian Networks (HBNs) [Heckerman et al. 1995; Shenoy et al. 2011]

Mixed Probabilistic Graphical Models (MPGMs) [Yang et al. 2014]

Tractable Probabilistic Circuits (PCs) [Molina et al. 2018; Vergari et al. 2019]

⇒ cannot deal with complex constraints
Continuous + discrete + constraints = SMT

Satisfiability Modulo Theories of the linear arithmetic over the reals (SMT($\mathcal{LRA}$)) delivers all these ingredients by design!

Widely used as a representation language for robotics, verification and planning [Barrett et al. 2010]
Continuous + discrete + constraints = SMT

Each player has a certain skill

Barrett et al., “Satisfiability modulo theories”, 2018
Continuous + discrete + constraints = SMT

\[ 0 \leq X_{P_i} \leq 10 \]
for \( i = 1, \ldots, N \)

Barrett et al., “Satisfiability modulo theories”, 2018
Continuous + discrete + constraints = SMT

0 ≤ X_{P_i} ≤ 10
for i = 1, . . . , N

Each team’s skill is bounded by its players’ skills

Barrett et al., “Satisfiability modulo theories”, 2018
Continuous + discrete + constraints = SMT

\[ 0 \leq X_{P_i} \leq 10 \]
for \( i = 1, \ldots, N \)

\[ |X_{T_j} - X_{P_i}| < 1 \]
for \( j = 1, \ldots, M, i = 1, \ldots, |T_j| \)

Barrett et al., “Satisﬁability modulo theories”, 2018
Continuous + discrete + constraints = SMT

\[ 0 \leq X_{P_i} \leq 10 \]
\[ \text{for } i = 1, \ldots, N \]

\[ |X_{T_j} - X_{P_i}| < 1 \]
\[ \text{for } j = 1, \ldots, M, i = 1, \ldots, |T_j| \]

Good teams form a squad

Barrett et al., “Satisfiability modulo theories”, 2018
Continuous + discrete + constraints = SMT

0 ≤ X_{P_i} ≤ 10
for i = 1, \ldots, N

| X_{T_j} - X_{P_i} | < 1
for j = 1, \ldots, M, i = 1, \ldots, |T_j|

B_{S_j} \Rightarrow X_{T_j} > 2
for j = 1, \ldots, M, i = 1

Barrett et al., “Satisfiability modulo theories”, 2018
Continuous + discrete + constraints = SMT

\[ \Delta = \bigwedge_i 0 \leq X_{P_i} \leq 10 \bigwedge_j \bigwedge_{i \in T_j} |X_{T_j} - X_{P_i}| < 1 \bigwedge_j (B_{S_j} \Rightarrow X_{T_j} > 2) \]

a single CNF SMT(\(\mathcal{LRA}\)) formula \(\Delta\)...

Barrett et al., “Satisfiability modulo theories”, 2018
Continuous + discrete + constraints = SMT

a single CNF SMT(\mathcal{LRA}) formula \Delta...and its primal graph

---

Barrett et al., “Satisfiability modulo theories”, 2018
\[ \bigwedge_{i} 0 \leq X_{P_i} \leq 10 \]
\[ \bigwedge_{j} \bigwedge_{i \in T_j} |X_{T_j} - X_{P_i}| < 1 \]
\[ \bigwedge_{j} (B_{S_j} \Rightarrow X_{T_j} > 2) \]

\[ \Delta \]

\[ + \]

\[ w(X_{P_i}), \quad \text{if } 0 \leq X_{P_i} \leq 10 \]
\[ w(X_{T_j}, X_{P_i}), \quad \text{if } |X_{T_j} - X_{P_i}| < 1 \]
\[ w(B_{S_j}, X_{T_j}), \quad \text{if } B_{S_j} \Rightarrow X_{T_j} > 2 \]

\[ \forall \]

Belle et al., “Probabilistic inference in hybrid domains by weighted model integration”, 2015
SMT + weights = Weighted Model Integration

\[ \bigwedge_i 0 \leq X_{P_i} \leq 10 \]
\[ \bigwedge_j \bigwedge_{i \in T_j} |X_{T_j} - X_{P_i}| < 1 \]
\[ \bigwedge_j (B_{S_j} \Rightarrow X_{T_j} > 2) \]

complex support

\begin{align*}
& \text{if } 0 \leq X_{P_i} \leq 10 \\
& \text{if } |X_{T_j} - X_{P_i}| < 1 \\
& \text{if } B_{S_j} \Rightarrow X_{T_j} > 2 \\
\end{align*}

\( w(X_{P_i}) \)
\( w(X_{T_j}, X_{P_i}) \)
\( w(B_{S_j}, X_{T_j}) \)

\( \text{densities} \)

\( \text{(unnormalized)} \)

\( \Pr_{\Delta}(\mathbf{X}, \mathbf{B}) \)

Belle et al., “Probabilistic inference in hybrid domains by weighted model integration”, 2015
Given an SMT(\(\mathcal{LRA}\)) formula \(\Delta\) over continuous vars \(X\) and discrete ones \(B\), and weight function \(W\), the \textit{weighted model integral} (WMI) is

\[
\text{WMI}(\Delta, \mathcal{W}; X, B) \triangleq \sum_{b \in \mathcal{B}} \int_{(x,b) \models \Delta} w(x, b) \, dx.
\]

i.e., computing the \textit{partition function} of the unnormalized distribution \(\Pr_\Delta\)

\[
\Rightarrow \quad \text{i.e., integrating the weighted volumes of the feasible regions of } \Delta!
\]

Belle et al., “Probabilistic inference in hybrid domains by weighted model integration”, 2015
“What is the probability of team $T_1$ to outperform team $T_2$, if $T_1$ is a squad but $T_2$ is not?”

Belle et al., “Probabilistic inference in hybrid domains by weighted model integration”, 2015
Advanced probabilistic reasoning

\[ \Phi_S : (B_{S_1} = 1 \land B_{S_2} = 0) \implies T_1 \text{ is a squad}, \ T_2 \text{ is not} \]
\[ \Phi_T : (X_{T_1} > X_{T_2}) \implies T_1 \text{ outperforms } T_2 \]

Belle et al., “Probabilistic inference in hybrid domains by weighted model integration”, 2015
Advanced probabilistic reasoning

\[ \Phi_S : (B_{S_1} = 1 \land B_{S_2} = 0) \implies T_1 \text{ is a squad, } T_2 \text{ is not} \]

\[ \Phi_T : (X_{T_1} > X_{T_2}) \implies T_1 \text{ outperforms } T_2 \]

\[
\Pr_D(\Phi_T \mid \Phi_S) = \frac{\text{WMI}(\Delta \land \Phi_T \land \Phi_S, \mathcal{W})}{\text{WMI}(\Delta \land \Phi_S, \mathcal{W})} = \frac{4,206}{7,225} \approx 58.22\%
\]

\[ \implies \text{conditional probabilities as a ratio of two weighted model integrals} \]

Belle et al., “Probabilistic inference in hybrid domains by weighted model integration”, 2015
Tractable WMI

WMI

#P-hard in general
```
\begin{align*}
B_{S_1} & \quad X_{T_1} \\
X_{P_1} & \quad X_{P_2}
\end{align*}
```

```
+ \begin{cases}
    w(X_{P_i}) = X_{P_i} \\
    w(X_{T_j}, X_{P_i}) = X_{T_j} X_{P_i} \\
    w(B_{S_j}, X_{T_j}) = X_{T_j}^2
\end{cases}
```

\[ \text{treeMI} \]

- tree-shaped primal graph
- constrained monomials \( \bigwedge \)
- polytime WMI inference

---

Zeng et al., “Efficient Search-Based Weighted Model Integration”, 2019
Tractable WMI

WMI

treeMI

#P-hard in general

largest tractable class known so far

Zeng et al., “Efficient Search-Based Weighted Model Integration”, 2019
Tractable WMI

- \#P-hard in general
- largest tractable class known so far
- still \#P-hard!

Zeng et al., “Efficient Search-Based Weighted Model Integration”, 2019
Tractable WMI

- #P-hard in general
- largest tractable class known so far
- still #P-hard!
- can we do better?

Zeng et al., “Efficient Search-Based Weighted Model Integration”, 2019
We frame tractable WMI inference at scale as a *message passing* scheme...

...on primal graphs...
We frame tractable WMI inference at scale as a *message passing* scheme...

...on primal graphs turned into *factor graphs*
We frame tractable WMI inference at scale as a *message passing* scheme...

...on primal graphs turned into *factor graphs*

- comprising an *upward*
We frame tractable WMI inference at scale as a message passing scheme...

...on primal graphs turned into factor graphs

comprising an upward and a downward pass
We frame tractable WMI inference at scale as a \textit{message passing} scheme...

...on primal graphs turned into \textit{factor graphs}

- comprising an \textit{upward} and a \textit{downward} pass
- exchanging messages from \textit{node to factors}

\[
m_{x_i \rightarrow f_S}(x_i) = \prod_{f_{S'} \in \text{neigh}(x_i) \setminus f_S} m_{f_{S'} \rightarrow x_i}(x_i)
\]
We frame tractable WMI inference at scale as a message passing scheme...

...on primal graphs turned into factor graphs

- comprising an upward and a downward pass
- exchanging messages from node to factors
- and from factors to nodes

\[
m_{f_{ij} \rightarrow x_i}(x_i) = \int f_{ij}(x_i, x_j) \cdot m_{x_j \rightarrow f_{ij}}(x_j) \, dx_j
\]
Tractable Weight Conditions

Which parametric family $\Omega$ for weights to guarantee tractable WMI inference?
Tractable Weight Conditions

Which parametric family $\Omega$ for weights to guarantee tractable WMI inference?

\[
m_{f_{ij} \rightarrow x_i}(x_i) = \int \prod_{\Gamma \in \Delta_S} [x_S \models \Gamma] \prod_{\ell \in \mathcal{L}_\Gamma} w_\ell(x_S)_{[x_S \models \ell]} \cdot m_{x_j \rightarrow f_{ij}}(x_j) \, dx_j
\]

\[
m_{x_i \rightarrow f_S}(x_i) = \prod_{f_{S'} \in \text{neigh}(x_i) \setminus f_S} m_{f_{S'} \rightarrow x_i}(x_i)
\]
Tractable Weight Conditions

Which parametric family $\Omega$ for weights to \textit{guarantee tractable WMI inference}?

$$m_{f_{ij} \rightarrow x_i}(x_i) = \int \prod_{\Gamma \in \Delta_S} \prod_{\ell \in \mathcal{L}_\Gamma} w_{\ell}(x_S) \cdot m_{x_j \rightarrow f_{ij}}(x_j) \, dx_j$$

$$m_{x_i \rightarrow f_S}(x_i) = \prod_{f_{S'} \in \text{neigh}(x_i) \setminus f_S} m_{f_{S'} \rightarrow x_i}(x_i)$$

Weights $\mathcal{W} \in \Omega$ should be \textit{closed under product}...
Tractable Weight Conditions

Which parametric family $\Omega$ for weights to guarantee tractable WMI inference?

\[
m_{f_{ij} \rightarrow x_i}(x_i) = \int \prod_{\Gamma \in \Delta_S} \left[ x_S \models \Gamma \right] \prod_{\ell \in \mathcal{L}_\Gamma} w_\ell(x_S)^{[x_S] = \ell} \cdot m_{x_j \rightarrow f_{ij}}(x_j) \, dx_j
\]

\[
m_{x_i \rightarrow f_S}(x_i) = \prod_{f_{S'} \in \text{neigh}(x_i) \setminus f_S} m_{f_{S'} \rightarrow x_i}(x_i)
\]

Weights $\mathcal{W} \in \Omega$ should be closed under product, closed under integration, and tractable for symbolic integration.
Tractable Weight Conditions

Which parametric family $\Omega$ for weights to guarantee tractable WMI inference?

\[
\begin{align*}
\mathbf{m}_{f_{ij}\rightarrow x_i}(x_i) &= \int \prod_{\Gamma \in \Delta_S} \chi_S \models \Gamma \prod_{\ell \in \mathcal{L}_\Gamma} w_\ell(x_S) \chi_S \models \ell \cdot \mathbf{m}_{x_j \rightarrow f_{ij}}(x_j) \, dx_j \\
\mathbf{m}_{x_i \rightarrow f_S}(x_i) &= \prod_{f_{S'} \in \text{neigh}(x_i) \setminus f_S} \mathbf{m}_{f_{S'} \rightarrow x_i}(x_i)
\end{align*}
\]

Weights $\mathcal{W} \in \Omega$ should be closed under product, closed under integration, and tractable for symbolic integration

$\implies$ e.g., arbitrary polynomials, exponentiated linear polynomials, etc.
An SMT formulation induces a *piecewise weight representation* strikingly different from message passing for classical PGMs!

\[
\begin{align*}
\mathbf{m}_{f_{ij} \rightarrow x_i}(x_i) &= \int \prod_{\Gamma \in \Delta_S} \left[ \mathbf{x}_S = \Gamma \right] \prod_{\ell \in \mathcal{L}_\Gamma} \omega_\ell(\mathbf{x}_S)^{[\mathbf{x}_S|\ell]} \cdot \mathbf{m}_{x_j \rightarrow f_{ij}}(x_j) \, dx_j \\
\mathbf{m}_{x_i \rightarrow f_S}(x_i) &= \prod_{f_{S'} \in \text{neigh}(x_i) \setminus f_S} \mathbf{m}_{f_{S'} \rightarrow x_i}(x_i)
\end{align*}
\]
An SMT formulation induces a **piecewise weight representation** strikingly different from message passing for classical PGMs!
An SMT formulation induces a **piecewise weight representation**

⇒ strikingly different from message passing for classical PGMs!

The number of all pieces in MP-WMI is $\mathcal{O}(4nc)^{2d+2}$, where $d$ is the graph diameter

⇒ the primal graph should have a **bounded diameter**!
Tractable WMI

- #P-hard in general
- the largest tractable class known before
- still #P-hard
- new largest class!

Zeng et al., “Efficient Search-Based Weighted Model Integration”, 2019
Scaling-up inference

Large set of synthetic benchmarks up to $N = 100$ vars, 5 trials, different primal graphs

**STAR**
- treewidth: 1
- diameter: 2

MP-WMI takes a *fraction of the time* of other exact WMI solvers like PA [Morettin et al. 2017] and F-XSDD [Zuidberg Dos Martires et al. 2019]
Scaling-up inference

Large set of synthetic benchmarks up to $N = 100$ vars, 5 trials, different primal graphs

**SNOW**
- treewidth: 1
- diameter: $\log(N)$

MP-WMI takes a *fraction of the time* of other exact WMI solvers like PA [Morettin et al. 2017] and F-XSDD [Zuidberg Dos Martires et al. 2019]
Scaling-up inference

Large set of synthetic benchmarks up to $N = 100$ vars, 5 trials, different primal graphs

MP-WMI takes a fraction of the time of other exact WMI solvers like PA [Morettin et al. 2017] and F-XSDD [Zuidberg Dos Martires et al. 2019]
Query amortization

A single message exchange allows to amortize univariate and bivariate queries ⇒ also all marginals and all moments!

MP-WMI answers 100 WMI queries faster than competitors solving 10 [Zeng et al. 2019]
Conclusions

Real-world data is noisy...
Conclusions

Real-world data is *noisy, complex*...
Conclusions

Real-world data is *noisy, complex* and *mixed continuous-discrete*...
Conclusions

Real-world data is noisy, complex and mixed continuous-discrete...

*The WMI framework* is very appealing for probabilistic inference in the real-world!
Real-world data is noisy, complex and mixed continuous-discrete...

*The WMI framework* is very appealing for probabilistic inference in the real-world!

MP-WMI delivers fast inference and defines the *largest class of tractable WMI models*
Conclusions

Real-world data is noisy, complex and mixed continuous-discrete...
The WMI framework is very appealing for probabilistic inference in the real-world!
MP-WMI delivers fast inference and defines the largest class of tractable WMI models

Next

However, MP-WMI requires tree-shaped bounded diameter primal graphs

⇒ we can build approximate inference schemes on it!
Conclusions

Real-world data is noisy, complex and mixed continuous-discrete...

The WMI framework is very appealing for probabilistic inference in the real-world!

MP-WMI delivers fast inference and defines the largest class of tractable WMI models.

Next

However, MP-WMI requires tree-shaped bounded diameter primal graphs

⇒ we can build approximate inference schemes on it!

Code

github.com/UCLA-StarAI/mpwmi
References


Minka, Tom, Ryan Cleven, and Yordan Zaykov (2018). “Trueskill 2: An improved bayesian skill rating system”. In:

