Relax, compensate and then integrate

Fast approximate probabilistic inference with logical and algebraic constraints

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

September 14th, 2020 - ***DeCoDeML Workshop 2020*** **- ECML-PKDD**

Relax, compensate and then integrate

Fast approximate probabilistic inference with logical and algebraic constraints

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

September 14th, 2020 - ***DeCoDeML Workshop 2020*** **- ECML-PKDD**

Relax, compensate and then integrate

Fast approximate probabilistic inference with logical and algebraic constraints

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

*September 14th, 2020 - **DeCoDeML Workshop 2020*** **- ECML-PKDD**

Relax, compensate and then integrate

Fast approximate probabilistic inference with logical and algebraic constraints

Zhe Zeng*
University of California, Los Angeles

Paolo Morettin*
University of Trento, Italy

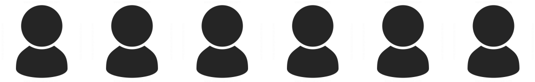
Fanqi Yan*
AMSS, Chinese Academy of Sciences

Antonio Vergari
University of California, Los Angeles

Guy Van den Broeck
University of California, Los Angeles

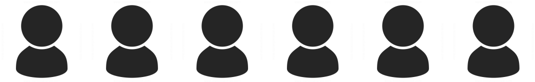
September 14th, 2020 - ***DeCoDeML Workshop 2020*** **- ECML-PKDD**

Skill matching system



Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

Skill matching system



- Each ***player*** has a certain skill

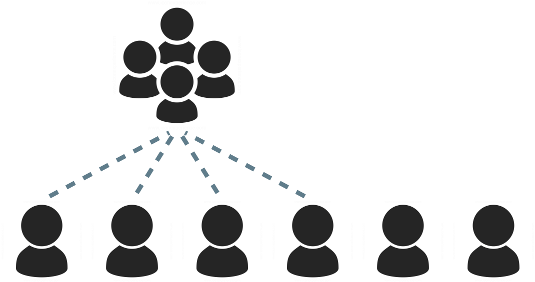
Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

Skill matching system

- Each ***player*** has a certain skill
  $\Longrightarrow$ *continuous variables*

Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

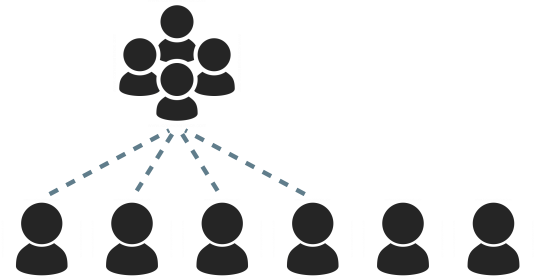
Skill matching system



- 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

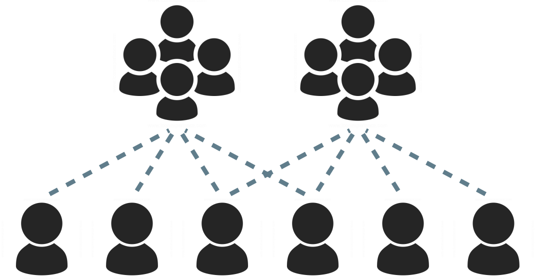


- Each ***player*** has a certain skill
- Players can form ***teams***

  ⟹ *intricate dependencies*

Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

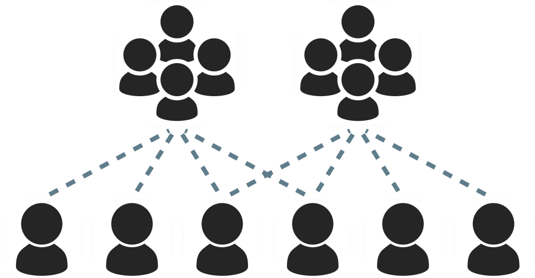
Skill matching system



- 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

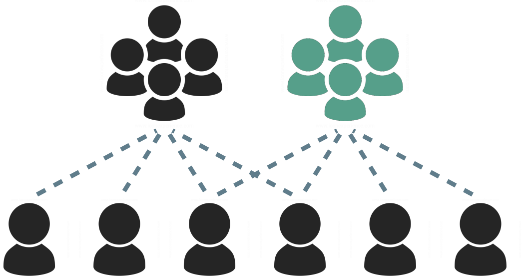


- Each ***player*** has a certain skill
- Players can form ***teams***
- Each team's skill is bounded by its players' skills

  $\Longrightarrow$ *complex constraints!*

Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

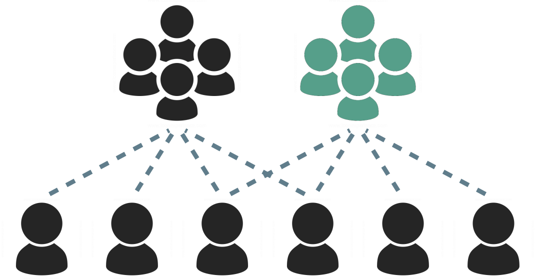
Skill matching system



- 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

Skill matching system

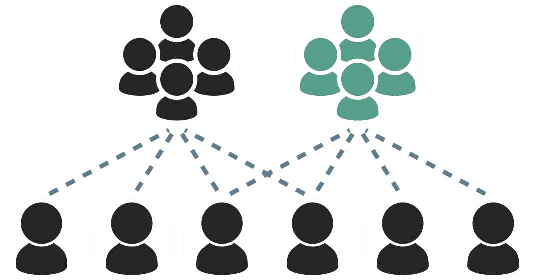


- Each ***player*** has a certain skill
- Players can form ***teams***
- Each team's skill is bounded by its players' skills
- Good teams form a ***squad***

  $\Longrightarrow$ *discrete variables*

Minka et al., "Trueskill 2: An improved bayesian skill rating system", 2018

Skill matching system



"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]*

$\Longrightarrow$ *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 Netowrks (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 Netowrks (HBNs)~~ *[Heckerman et al. 1995; Shenoy et al. 2011]*

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

$\Longrightarrow$ *strong distributional assumptions*

Continuous + **discrete** + **constraints** = ?

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

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

~~Hybrid Bayesian Netowrks (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]*

Continuous + discrete + constraints = ?

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

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

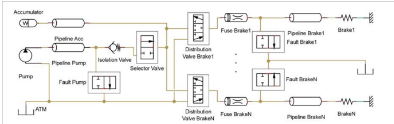
~~Hybrid Bayesian Netowrks (HBNs)~~ *[Heckerman et al. 1995; Shenoy et al. 2011]*

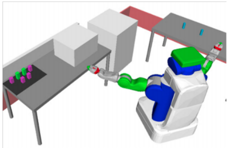
~~Mixed Probabilistic Graphical Models (MPGMs)~~ *[Yang et al. 2014]*

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

$\Longrightarrow$ *cannot deal with complex algebraic constraints*

Continuous + **discrete** + **constraints** = **SMT**



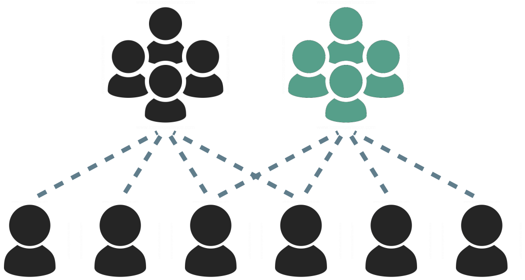


Satisfiability Modulo Theories of the linear arithmetic over the reals (SMT($\mathcal{LRA}$)) delivers all these ingredients by design!

Widely used as a representation language for ***robotics***, ***verification*** and ***planning*** *[Barrett et al. 2010]*

Barrett et al., "Satisfiability modulo theories", 2018

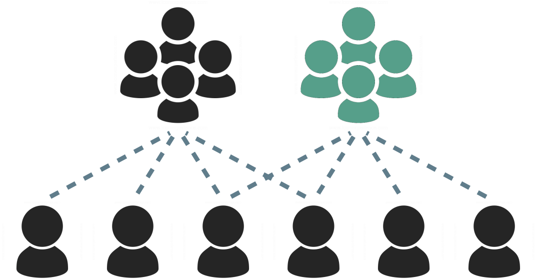
Continuous + discrete + constraints = SMT



- Each ***player*** has a certain skill

Barrett et al., "Satisfiability modulo theories", 2018

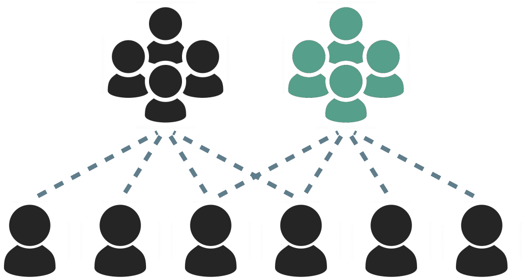
Continuous + discrete + constraints = SMT



- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$

Barrett et al., "Satisfiability modulo theories", 2018

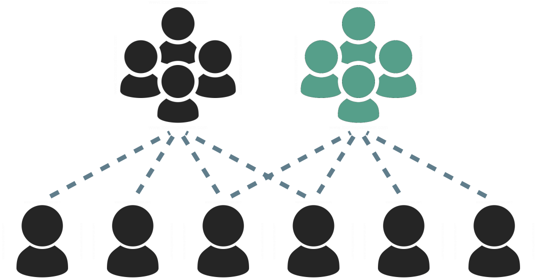
Continuous + **discrete** + **constraints** = **SMT**



- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$
- Each team's skill is bounded by its players' skills

Barrett et al., "Satisfiability modulo theories", 2018

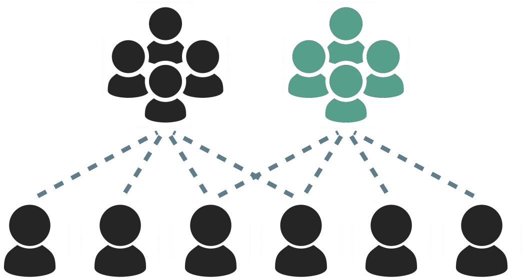
Continuous + discrete + constraints = SMT



- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$
- $\mid X_{T_j} - X_{P_i} \mid < 1$
  for $j = 1, \ldots, M, i = 1, \ldots, |T_j|$

Barrett et al., "Satisfiability modulo theories", 2018

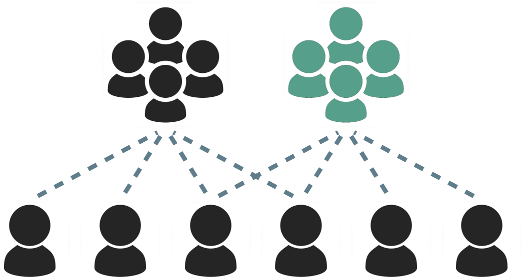
Continuous + **discrete** + **constraints** = **SMT**



- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$
- $\mid X_{T_j} - X_{P_i} \mid < 1$
  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**



- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$
- $\mid X_{T_j} - X_{P_i} \mid < 1$
  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

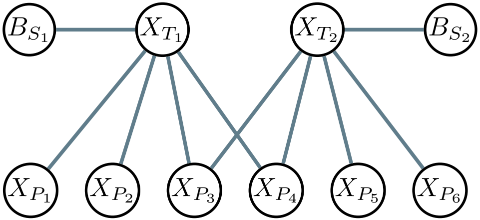
Continuous + **discrete** + **constraints** = **SMT**

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

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

Barrett et al., "Satisfiability modulo theories", 2018

Continuous + **discrete** + **constraints** = **SMT**



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

Barrett et al., "Satisfiability modulo theories", 2018

SMT + weights

$$\bigwedge_i 0 \leq X_{P_i} \leq 10$$

$$\bigwedge_j \bigwedge_{i \in T_j} \mid X_{T_j} - X_{P_i} \mid < 1$$

$$\bigwedge_j (B_{S_j} \Rightarrow X_{T_j} > 2)$$

+

$$\begin{cases} w(X_{P_i}), \\ \quad \text{if } 0 \leq X_{P_i} \leq 10 \\ \\ w(X_{T_j}, X_{P_i}), \\ \quad \text{if } \mid X_{T_j} - X_{P_i} \mid < 1 \\ \\ w(B_{S_j}, X_{T_j}), \\ \quad \text{if } B_{S_j} \Rightarrow X_{T_j} > 2 \end{cases}$$

SMT formula Δ

weight functions $\mathcal{W}$

Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

SMT + weights = Weighted Model Integration

$$\bigwedge_i 0 \leq X_{P_i} \leq 10$$

$$\bigwedge_j \bigwedge_{i \in T_j} \mid X_{T_j} - X_{P_i} \mid < 1$$

$$\bigwedge_j (B_{S_j} \Rightarrow X_{T_j} > 2)$$

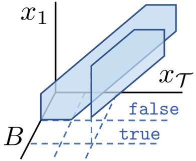
complex support

+

$$\begin{cases} w(X_{P_i}), & \text{if } 0 \leq X_{P_i} \leq 10 \\ w(X_{T_j}, X_{P_i}), & \text{if } \mid X_{T_j} - X_{P_i} \mid < 1 \\ w(B_{S_j}, X_{T_j}), & \text{if } B_{S_j} \Rightarrow X_{T_j} > 2 \end{cases}$$

densities

=



(unnormalized)

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

Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

SMT + densities = Weighted Model Integration

Given an SMT($\mathcal{LRA}$) formula Δ over continuous vars $\mathbf{X}$ and discrete ones $\mathbf{B}$, and weight function $\mathcal{W}$, the ***weighted model integral*** (WMI) is

$$\mathsf{WMI}(\Delta, \mathcal{W}; \mathbf{X}, \mathbf{B}) \triangleq \sum_{\boldsymbol{b} \in \mathbb{B}^{|\mathbf{B}|}} \int_{(\mathbf{x}, \boldsymbol{b}) \models \Delta} w(\mathbf{x}, \boldsymbol{b}) \, d\mathbf{x}.$$

i.e., computing the ***partition function*** of the unnormalized distribution Pr_{Δ}

$\Longrightarrow$ *i.e., integrating the weighted volumes of the feasible regions of Δ!*

Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

WMI

Advanced probabilistic reasoning

"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

WMI

Advanced probabilistic reasoning

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

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

Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

WMI

Advanced probabilistic reasoning

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

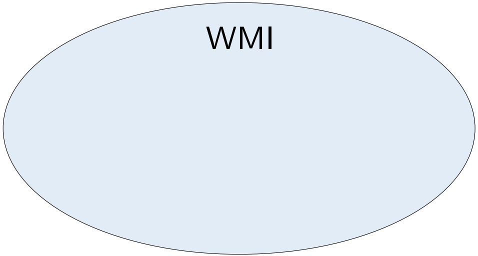
$$\Phi_T : (X_{T_1} > X_{T_2}) \implies T_1 \text{ outperforms } T_2$$

$$\text{Pr}_\Delta(\Phi_T \mid \Phi_S) = \frac{\text{WMI}(\Delta \wedge \Phi_T \wedge \Phi_S, \mathcal{W})}{\text{WMI}(\Delta \wedge \Phi_S, \mathcal{W})} = \frac{4,206}{7,225} \approx 58.22\%$$

$\Rightarrow$ *conditional probabilities as a ratio of two weighted model integrals*

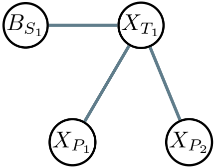
Belle et al., "Probabilistic inference in hybrid domains by weighted model integration", 2015

Tractability of WMI



#P-hard in general

treeWMI



+

$$\begin{cases} w(X_{P_i}) = X_{P_i}^2 \\ w(X_{T_j}, X_{P_i}) = X_{T_j} + X_{P_i} \\ w(B_{S_j}) = 100 \end{cases}$$

=

treeWMI

[Zeng et al. 2020]

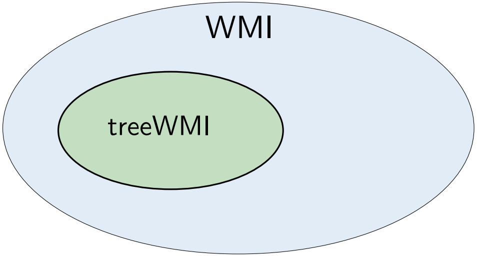
tree-shaped primal graph

tractable weight conditions

polytime WMI inference

Zeng et al., "Scaling up Hybrid Probabilistic Inference with Logical and Arithmetic Constraints via Message Passing", 2020

Tractability of WMI

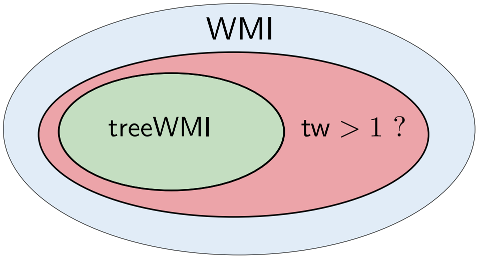


#P-hard in general

largest tractable class known so far

Zeng et al., "Scaling up Hybrid Probabilistic Inference with Logical and Arithmetic Constraints via Message Passing", 2020

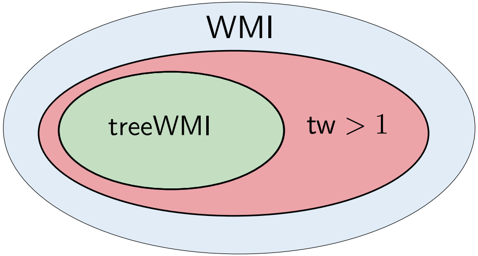
Tractability of WMI



- ***#P-hard*** in general
- largest tractable class known so far
- can we expand it?

Zeng et al., "Scaling up Hybrid Probabilistic Inference with Logical and Arithmetic Constraints via Message Passing", 2020

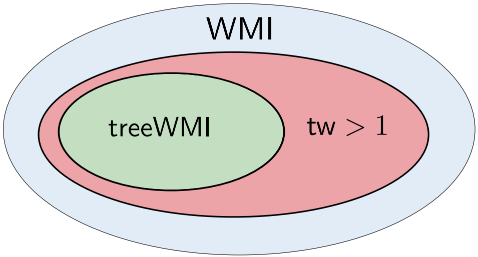
Tractability of WMI



- ***#P-hard*** in general
- largest tractable class known so far
- can we expand it? ***No!***

Zeng et al., "Scaling up Hybrid Probabilistic Inference with Logical and Arithmetic Constraints via Message Passing", 2020

Tractability of WMI



- ***#P-hard*** in general
- largest tractable class known so far
- can we expand it? ***No!***

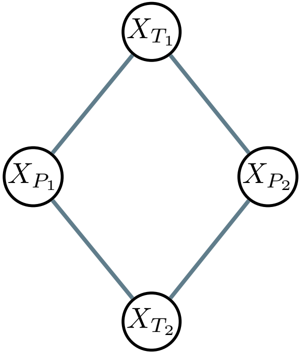
Thm. *Let* $\mathsf{WMI}(\mathrm{tw})$ *be the class of WMI problems with bounded diameter and treewidth* tw*.* $\mathsf{WMI}(\mathrm{tw})$ *is a tractable WMI class* ***iff*** *treewidth* $\mathrm{tw} = 1$*.*

ReCoIn

Approximate WMI Inference

ReCoIn

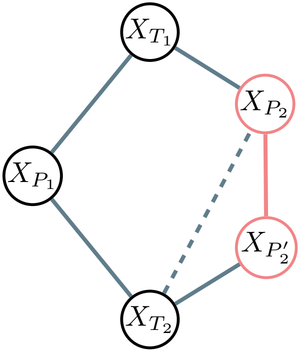
Approximate WMI Inference



- Given a WMI problem with ***loopy primal graph***

ReCoIn

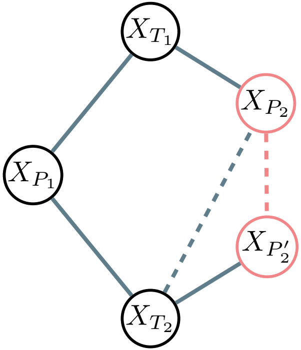
Approximate WMI Inference



- Given a WMI problem with ***loopy primal graph***
- ***Re***lax it by adding ***copies*** of literals

ReCoIn

Approximate WMI Inference



- Given a WMI problem with ***loopy primal graph***
- ***Re***lax it by adding ***copies*** of literals, then removing equality constraints

$\Longrightarrow$ *removing dependencies, breaking loops*

ReCoIn

Approximate WMI Inference

$\{(\ell, w_\ell(x_{P_2}))\}_\ell$

$\{(\ell, w'_\ell(x_{P_2'}))\}_\ell$

- Given a WMI problem with ***loopy primal graph***
- ***Re***lax it by adding ***copies*** of literals, then removing equality constraints
- ***Co***mpensate for the removed dependencies, by introducing certain literals and weights

ReCoIn

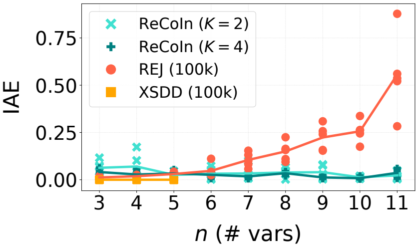
Approximate WMI Inference

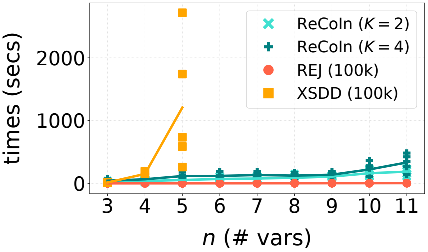
$$w_\ell \leftarrow f(\mathsf{Pr}_\Delta(\ell); w')$$

$$w'_\ell \leftarrow f(\mathsf{Pr}_\Delta(\ell); w)$$

- Given a WMI problem with ***loopy primal graph***
- ***Re***lax it by adding ***copies*** of literals, then removing equality constraints
- ***Co***mpensate for the removed dependencies, by introducing certain literals and weights
- optimize compensating weights iteratively by solving a series of exact ***In***tegration problems

Experiments





$\Longrightarrow$ ReCoIn ***better scales*** *to larger WMI problems than other approximate WMI solvers while still* ***delivering accurate approximations***

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!

Conclusions

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

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

Efficient approximations are not only useful, but ***needed***

Conclusions

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

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

Efficient approximations are not only useful, but ***needed***

$\Longrightarrow$ *ReCoIn delivers fast approximate inference*

Conclusions

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

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

Efficient approximations are not only useful, but ***needed***

$\implies$ *ReCoIn delivers fast approximate inference*

Next

Application to program verification, probabilistic (logic) programming, ...

Conclusions

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

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

Efficient approximations are not only useful, but ***needed***

$\Longrightarrow$ *ReCoIn delivers fast approximate inference*

Next

Application to program verification, probabilistic (logic) programming, ...

Questions?

References I

- Heckerman, David and Dan Geiger (1995). “Learning Bayesian networks: a unification for discrete and Gaussian domains”. In: *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., pp. 274–284.
- Barrett, Clark et al. (2010). “The SMT-LIB initiative and the rise of SMT (HVC 2010 award talk)”. In: *Proceedings of the 6th international conference on Hardware and software: verification and testing*. Springer-Verlag, pp. 3–3.
- Shenoy, Prakash P and James C West (2011). “Inference in hybrid Bayesian networks using mixtures of polynomials”. In: *International Journal of Approximate Reasoning* 52.5, pp. 641–657.
- Kingma, Diederik P and Max Welling (2013). “Auto-encoding variational bayes”. In: *arXiv preprint arXiv:1312.6114*.
- Goodfellow, Ian et al. (2014). “Generative adversarial nets”. In: *Advances in neural information processing systems*, pp. 2672–2680.
- Yang, Eunho et al. (2014). “Mixed graphical models via exponential families”. In: *Artificial Intelligence and Statistics*, pp. 1042–1050.
- Belle, Vaishak, Andrea Passerini, and Guy Van den Broeck (2015). “Probabilistic inference in hybrid domains by weighted model integration”. In: *Proceedings of 24th International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2770–2776.
- Barrett, Clark and Cesare Tinelli (2018). “Satisfiability modulo theories”. In: *Handbook of Model Checking*. Springer, pp. 305–343.
- Minka, Tom, Ryan Cleven, and Yordan Zaykov (2018). “Trueskill 2: An improved bayesian skill rating system”. In:
- Molina, Alejandro et al. (2018). “Mixed sum-product networks: A deep architecture for hybrid domains”. In: *Thirty-second AAAI conference on artificial intelligence*.
- Vergari, Antonio et al. (2019). “Automatic Bayesian density analysis”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 33, pp. 5207–5215.

References II

Zeng, Zhe et al. (2020). “Scaling up Hybrid Probabilistic Inference with Logical and Arithmetic Constraints via Message Passing”. In: *Proceedings of the 37th International Conference on Machine Learning (ICML)*. url: http://starai.cs.ucla.edu/papers/ZengICML20.pdf.