Probabilistic Inference with Algebraic Constraints

Theoretical Limits and Practical Approximations

Zhe Zeng*
University of California, Los Angeles

Paolo Morettin*
University of Trento, Italy

Fanqi Yan*
University of Texas at Austin

Antonio Vergari
University of California, Los Angeles

Guy Van den Broeck
University of California, Los Angeles

*December - **NeurIPS 2020** Spotlight*

Probabilistic Inference with Algebraic Constraints

Theoretical Limits and Practical Approximations

Zhe Zeng*
University of California, Los Angeles

Paolo Morettin*
University of Trento, Italy

Fanqi Yan*
University of Texas at Austin

Antonio Vergari
University of California, Los Angeles

Guy Van den Broeck
University of California, Los Angeles

*December - **NeurIPS 2020** Spotlight*

Probabilistic Inference with Algebraic Constraints

Theoretical Limits and Practical Approximations

Zhe Zeng*
University of California, Los Angeles

Paolo Morettin*
University of Trento, Italy

Fanqi Yan*
University of Texas at Austin

Antonio Vergari
University of California, Los Angeles

Guy Van den Broeck
University of California, Los Angeles

*December - **NeurIPS 2020** Spotlight*

Probabilistic Inference with Algebraic Constraints

Theoretical Limits and Practical Approximations

Zhe Zeng*
University of California, Los Angeles

Paolo Morettin*
University of Trento, Italy

Fanqi Yan*
University of Texas at Austin

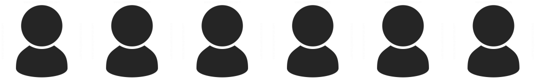
Antonio Vergari
University of California, Los Angeles

Guy Van den Broeck
University of California, Los Angeles

*December - **NeurIPS 2020** Spotlight*

Motivation

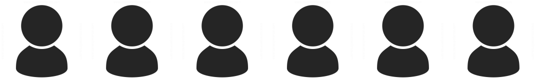
Example: Skill matching system



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

Motivation

Example: Skill matching system

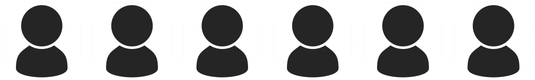


- Each ***player*** has a certain skill
  $\Longrightarrow$ *continuous variables*

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

Motivation

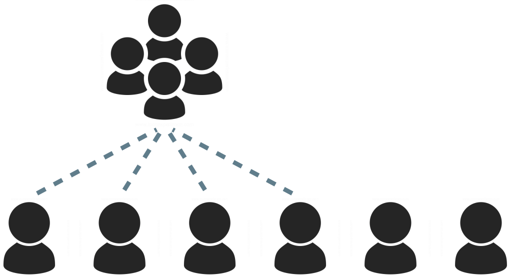
Example: Skill matching system



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

Motivation

Example: Skill matching system



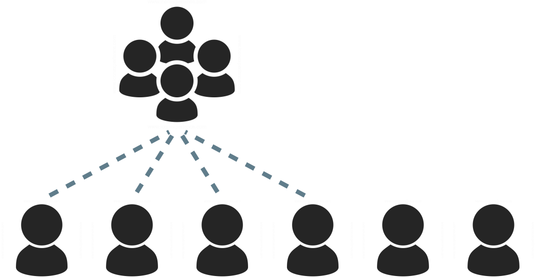
- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$

- Players can form ***teams***
  $\Longrightarrow$ *complex constraints*

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

Motivation

Example: Skill matching system

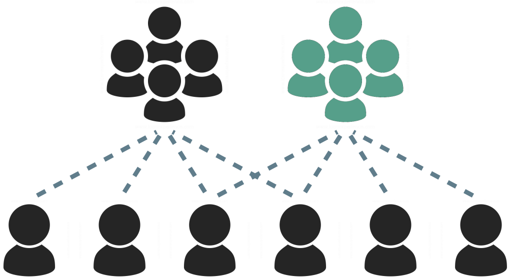


- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$
- $| X_{T_j} - X_{P_i} | < 1$
  for $j = 1, \ldots, M, i = 1, \ldots, |T_j|$

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

Motivation

Example: Skill matching system

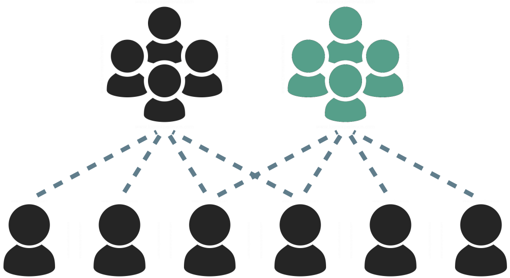


- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$
- $\mid X_{T_j} - X_{P_i} \mid < 1$
  for $j = 1, \ldots, M, i = 1, \ldots, |T_j|$
- Good teams form a ***squad***
  $\Longrightarrow$ *discrete variables*

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

Motivation

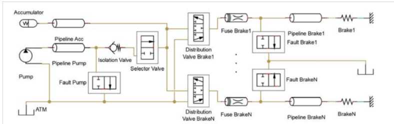
Example: Skill matching system

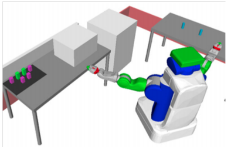


- $0 \leq X_{P_i} \leq 10$
  for $i = 1, \ldots, N$
- $\mid X_{T_j} - X_{P_i} \mid < 1$
  for $j = 1, \ldots, M, i = 1, \ldots, |T_j|$
- $B_{S_j} \Rightarrow X_{T_j} > 2$
  for $j = 1, \ldots, M, i = 1$

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

Continuous + discrete + constraints = SMT





Satisfiability Modulo Theories of linear arithmetic over the reals (SMT($\mathcal{LRA}$)) delivers all the 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** = ?

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]*

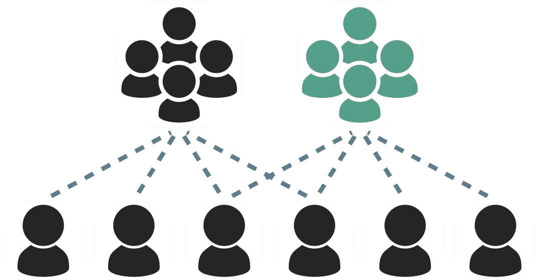
~~Variational Autoencoders (VAEs)~~ *[Kingma et al. 2013]*

~~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]*

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

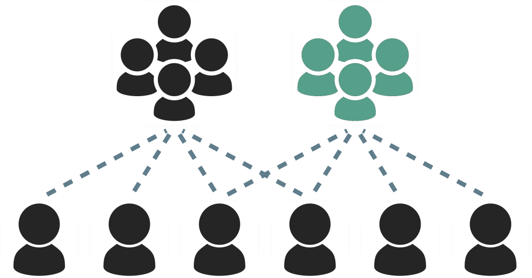
Continuous + **discrete** + **constraints** = **SMT**

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

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

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

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



"What is the probability of team T_1 outperforming team T_2, if T_1 is a squad but T_2 is not?"

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

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

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

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

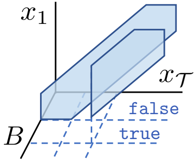
complex support

+

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

densities

=



(***unnormalized***)

$\text{Pr}_{\Delta}(\mathbf{X}, \mathbf{B})$

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

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

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

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

$\Longrightarrow$ *integrating the **densities** of the **feasible regions** of Δ!*

i.e., computing the ***partition function*** of the unnormalized distribution $\Pr_{\Delta}$

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

"What is the probability of team T_1 outperforming 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

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

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

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

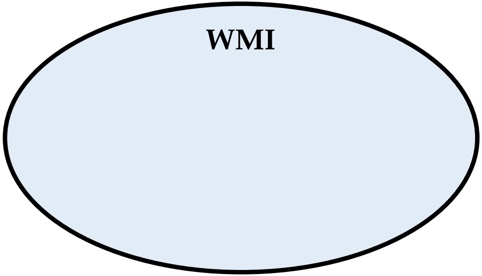
$\Rightarrow$ *conditional probabilities as a ratio of two weighted model integrals*

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

Tractability of WMI

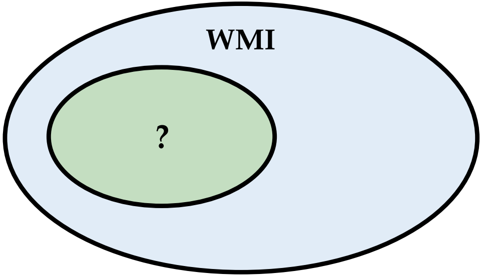
Why is building inference algorithms for hybrid domains difficult?

Tractability of WMI



#P-hard in general

Tractability of WMI



#P-hard in general

what would be tractable?

Primal Graph

Discrete Graphical Models

$$\bigwedge_{i=1,2} (X_i \Rightarrow X_{i+1})$$

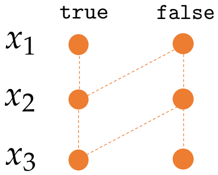
WMI models

$$\bigwedge_{i=1,2} \{(X_i - 0.1 \leq X_{i+1} \leq X_i + 0.1) \lor (X_i + 0.9 \leq X_{i+1} \leq X_i + 1.1)\}$$

Primal Graph

Tree Primal Graph

Discrete Graphical Models



tractability ✔

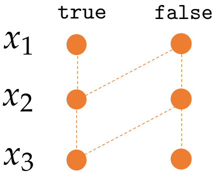
WMI models

?

Primal Graph

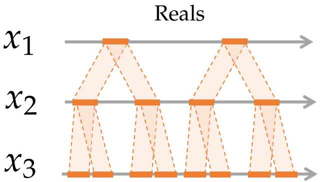
Tree Primal Graph

Discrete Graphical Models



tractability ✔

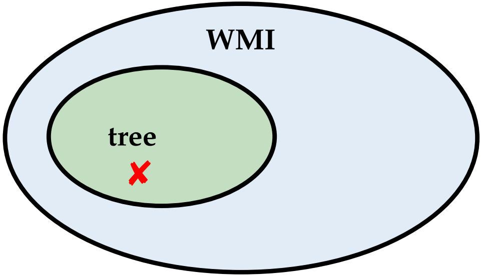
WMI models



tractability ✘

Primal Graph

Tractability of WMI

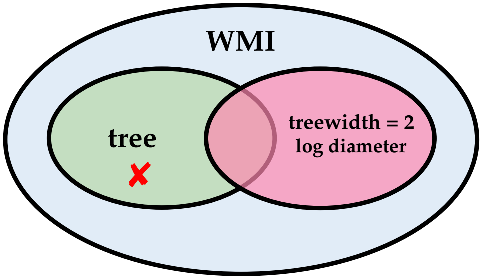


■ ***#P-hard*** in general

■ tree WMI problem class ✘

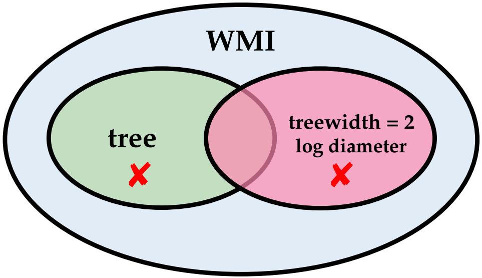
WMI Inference on tree-shaped primal graphs with unbounded-diameter is #P-hard!

Tractability of WMI



- ***#P-hard*** in general
- tree WMI problem class ✗
- logarithmic diameter and treewidth two?

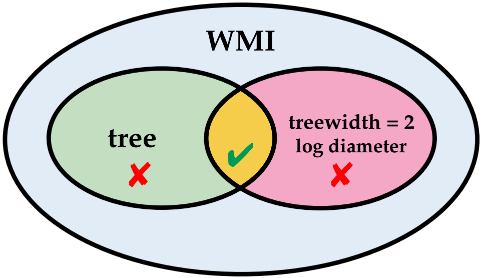
Tractability of WMI



- ***#P-hard*** in general
- tree WMI problem class ✗
- logarithmic diameter and treewidth two ✗

WMI inference on primal graphs with bounded-diameter but treewidth two is #P-hard!

Tractability of WMI



- ***#P-hard*** in general
- tree WMI problem class ✗
- logarithmic diameter and treewidth two ✗
- intersection ✔ *[Zeng et al. 2020]*

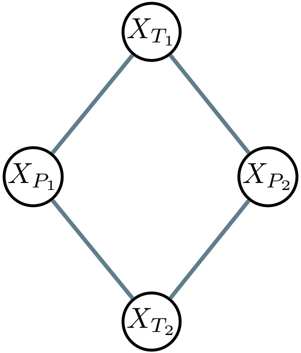
...but how can we perform inference on general WMI problems?

ReCoIn

Approximate WMI Inference

ReCoIn

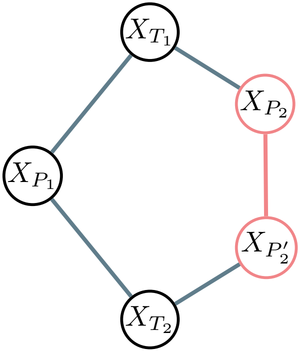
Approximate WMI Inference



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

ReCoIn

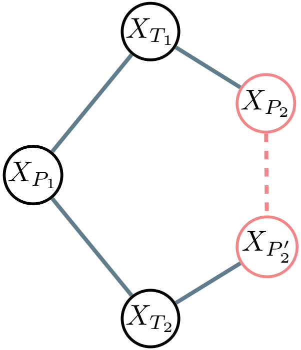
Approximate WMI Inference



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

ReCoIn

Approximate WMI Inference

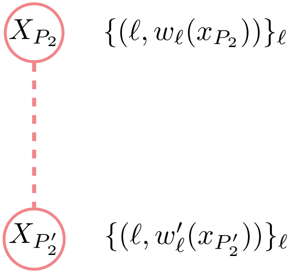


- Given a WMI problem with ***loopy primal graph***
- ***Re***lax it by adding ***copies*** of literals, then removing equality constraints

  $\Longrightarrow$ *removing dependencies, breaking loops*

ReCoIn

Approximate WMI Inference



- 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

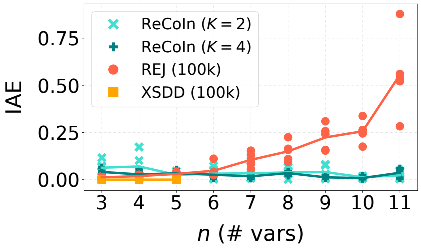
Approximate WMI Inference

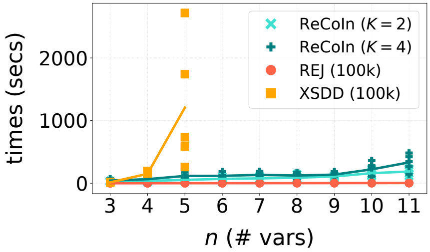
$$w_\ell \leftarrow f(\Pr_\Delta(\ell); w')$$

$$w'_\ell \leftarrow f(\Pr_\Delta(\ell); w)$$

- 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 ***scales better*** *to larger WMI problems 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, ...

Code

`github.com/UCLA-StarAI/recoin`

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.