Is Parameter Learning via Weighted Model Integration Tractable?

Zhe Zeng* 1 Paolo Morettin 2 Fangqi Yan 3 Antonio Vergari 1 Andrea Passerini 4 Guy Van den Broeck 1

1 University of California, Los Angeles 2 KU Leuven 3 University of Texas, Austin 4 University of Trento

Weighted Model Integration (WMI)

A framework for hybrid probabilistic inference with algebraic constraints [1].

\[ WMI(\Delta; w, X, B) = \sum_{x \in B} WMI(\Delta \land X, X, B) \]

i.e. integration over the weighted satisfying assignments of a formula over hybrid variables

Algebraic constraints are Satisfiability Modulo Theories (SMT) formulas, as combinations of Boolean literals & linear real arithmetic (LRA) literals in Conjunctive Normal Form (CNF).

Per-literal weights assign weight if the literal is SAT; otherwise assign one.

Together they define a joint weight.

Answer queries

\[ Pr(\theta) = \frac{WMI(\Delta; w, X, B)}{WMI(\Delta, w, X, B)} \]

Contributions

- Provide theoretical insights on tractability of MLE-based parameter learning of WMI.
- Bridging two fields: Hinge-Loss Markov Random Fields (HL-MRFs) and WMI models, by reducing marginal inference of HL-MRFs to WMI inference.

Maximum Likelihood Estimation (MLE)

Optimization objective log-likelihood of dataset

\[ \theta^* = \arg \max_{\theta} L(\theta; D) \]

Convexity?

Prop. The log-likelihood of dataset is concave if the weight functions are log-linear in their parameters.

Tractability of computing gradients?

Partial derivatives of the log-likelihood with respect to a parameter in log-linear weight function:

\[ \frac{\partial}{\partial \theta} \log \mathbb{P}(D|\theta) \]

Tractability?

Thm. The computation of the partial derivative is tractable if

i) WMI model is in the tractable WMI problem class [1];
   ii) the feature functions are in function families that satisfies tractable weight conditions (TWCs),

where A family of weight functions satisfies TWCs iff:

- It is closed under product;
- It admits efficient computation of antiderivatives;
- It is closed under definite integration.

For example, (piecewise) polynomial, log-linear functions ...

Marginal Inference For HL-MRFs via WMI

Probabilistic Soft Logic (PSL) is a statistical relational learning (SRL) framework for modeling probabilistic and relational domains, e.g.,

\[ 0.3 : \text{friend}(A, B) \land \text{votesFor}(B, P) \]

A PSL program induces a Hinge-Loss Markov Random Field (HL-MRF), which defines the distribution over interpretations

\[ Pr(x_1, x_2) = \exp(-0.3 \max(x_1 + x_2 - 1, 0)) \]

Thm. For any HL-MRF, there exists a WMI model with per-literal weights whose WMI density equals to the HL-MRF density.

For example, the WMI model that is equal to the HL-MRF shown above is

\[ \text{Formula} (x_1 + x_2 - 1 \geq 0) \land \text{True} \]

Weight

\[ w_1(x_1, x_2) = \exp(-0.3(x_1 + x_2 - 1)) \]

This allows us to characterize the tractability of marginal inference for HL-MRFs via the analysis for WMI.

References