Efficient SMT-based Weighted Model Integration for AI verification

Giuseppe Spallitta¹, Gabriele Masina¹, Paolo Morettin², Andrea Passerini¹, Roberto Sebastiani¹
¹University of Trento ²KU Leuven
Code and benchmarks available at: https://github.com/unitn-sml/wmi-pa

Why AI verification is becoming relevant

- Developing neural networks (NN) to help humans in their everyday jobs is becoming easier and easier...
- But are we sure that they learned what we expect?

It is essential to formally verify properties on neural network and probabilistic models before releasing them on market:
- Fairness
- Safety
- Robustness to noise

Weighted Model Integration (WMI)

Let \( \mathbf{x} \equiv \{x_1, \ldots, x_N\} \in \mathbb{R}^N \) and \( \mathbf{A} \equiv \{A_1, \ldots, A_M\} \in \mathbb{B}^M \), \( \varphi(\mathbf{x}, \mathbf{A}) \) denotes an SMT\((\mathcal{L}A_4)\) formula, while \( w(\mathbf{x}, \mathbf{A}) \) denotes a non-negative weight function s.t. \( \mathbb{R}^N \times \mathbb{B}^M \rightarrow \mathbb{R}^+ \).

The Weighted Model Integral of \( w(\mathbf{x}, \mathbf{A}) \) over \( \varphi(\mathbf{x}, \mathbf{A}) \) is:

\[
\text{WMI}(\varphi, w | \mathbf{x}, \mathbf{A}) \equiv \sum_{\mathbf{a} \in \mathbb{B}^M} \text{WMI}_{\mathbf{a}}(\varphi | \mathbf{x}, \mathbf{a}) w(\mathbf{x}, \mathbf{a})
\]

\[
\text{WMI}_{\mathbf{a}}(\varphi, w | \mathbf{x}) \equiv \int_{\varphi(\mathbf{x})} w(\mathbf{x}) \, d\mathbf{x}
\]

Hybrid probabilistic inference through WMI

\[
\Pr(R_1 \lor R_2 | M) = \frac{\text{WMI}(\Delta_M \land (R_1 \lor R_2), w)}{\text{WMI}(\Delta_M, w)}
\]

Main issues: can we compute it efficiently?
- Integrals are hard to compute...
- The more complex the models, the higher the number of integrals we have to deal with!

Key idea

We propose SA-WMI-PA, a novel WMI approach based on partial enumeration which is aware of the conditional structure of \( w \):

- Compile the weight function into a valid formula that drives the enumeration of assignments in WMIPA.
  \( \implies \) Takes the best out of KC and WMIPA approaches, the current state-of-the-art frameworks.

Structure-aware WMI-PA

Generate a formula from \( w, sk(w) \), s.t.:
- Its atoms are all and only conditions in \( w \);
- \( \varphi \) is equivalent to \( \varphi \land sk(w) \);
- Any partial truth value assignment \( \mu \) to the conditions of \( w \) ensures \( w_{\mu} \) is arithmetically computable.

Main advantages:
- The SMT solver is not forced to enumerate all conditions on \( w \) if they do not impact the final value of \( w_{\mu} \);
- We can perform partial enumeration on the Boolean atoms.
  \( \implies \) Fewer integrals to compute

Experimental results

Synthetic problems

Probabilistic inference on DTS