Partition Function Estimation: A Quantitative Study

Durgesh Agrawal[1], Yash Pote[2], and Kuldeep S. Meel[2]

[1] Indian Institute of Technology, Kanpur
[2] School of Computing, National University of Singapore

IJCAI 2021

Partition Function

- $\mathbf{X}$: Set of discrete random variables
- Let $\mathbf{S} \subseteq 2^{\mathbf{X}}$. For each $\mathbf{x} \in \mathbf{S}$, $f_{\mathbf{x}}$: factor, i.e. a non-negative real valued function over $\mathbf{x}$.
- $\sigma(\mathbf{x})$: An assignment of $\mathbf{x}$.
- Probability distribution: $\mathbb{P}(\sigma(\mathbf{X})) = \frac{1}{Z} \prod_{\mathbf{x} \in \mathbf{S}} f_{\mathbf{x}}(\sigma(\mathbf{x}))$

Objective: Compute the partition function $Z := \sum_{\sigma(\mathbf{X})} \prod_{\mathbf{x} \in \mathbf{S}} f_{\mathbf{x}}(\sigma(\mathbf{x}))$

Motivation

Partition functions occur prominently in the study of Probabilistic Graphical Models, which have applications in

- Image segmentation
- Image recognition
- Computational Protein Design

  ... and many more

Classification - I

Solvers can be classified into four broad groups based on their core technical idea:

1. Message passing
   - Belief Propagation(BP), Fractional BP, Generalized BP, Edge Deletion BP, Double Loop Generalized BP, Conditional BP, Tree Expectation Propagation, Join Tree.
2. Variable elimination
   - Bucket Elimination, Weighted Mini Bucket Elimination
3. Model counting
   - Ace, GANAK, miniC2D, WeightCount, Weighted Integral by Sums and Hashing
4. Sampling
   - SampleSearch, Dynamic Importance Sampling, FocusedFlatSAT

Classification - II

Solvers in practice differ in a few ways:

- Exact
  - Model Counting, JTree
- Approximate
  - Approximate Model Counting, BP on polytrees, Sampling.
- Guarantee-less
  - Loopy BP, Weighted Mini Bucket Elimination

Question: How can practitioners choose the best solving technique?

Experimental Design Challenges - I

Challenge: Ground truth is unknown for a majority of problems.
Solution: Use a conjectured value of Z when ground truth is unavailable.

- Create a mini-dataset for which the ground truth is known.
- Select *reliable* methods that return either an accurate answer or no output at all on the mini-dataset.
- For problems in the actual dataset, conjectured partition function is the median of answers returned by *reliable* methods.

Experimental Design Challenges - II

Challenge: Some solvers converge without terminating within time limit.
Solution: Collect suboptimal results within time limit when solver does not converge.

- Give two timeouts to a solver - soft (9500s) and hard (9500s+500s).
- When soft timeout is reached, allow the algorithm to exit gracefully and return an output based on incomplete execution.
- If hard timeout is reached and no answer is returned, terminate the solver.

Experimental Design Challenges - III

Challenge: This is a functional problem rather than a decision problem.
Solution: Evaluate methods on a metric which is a function of accuracy and runtime.

- Define TAP Score - an extension of the PAR-2 scoring system.
- Penalise a solver for inaccurate and delayed output.
- Lower TAP Score $\implies$ Better performance

Experimental Results - Benchmarks Solved

- Exact solver Ace solves maximum number of problems.
- For large variable cardinality problems, Belief Propagation variants solve more problems than Ace.
- Other exact algorithms solve significantly fewer problems.

Experimental Results - Accuracy

- Belief Propagation variants have varying accuracies over the dataset.
- Ace returns exact answers for the maximum number of problems.
- At least one method solves every problem with relative error $< 2^{0.01}$.

Experimental Results - Runtime Variation

- Bucket Elimination and SampleSearch return answer immediately or not at all.
- At least one method solves every problem in < 20 seconds.
- At least one method solves 99.7% problems in the dataset with relative error $< 2^{0.01}$ in < 500 seconds.

Experimental Results - TAP Score

- Belief Propagation variants have low overall TAP Score and consistent performance across all classes despite lack of formal guarantees.
  - Work best on an assorted dataset.
- Ace performs better than all other exact methods.
  - Ideal choice for obtaining exact results.
- Virtual Best Solver's TAP Score is $\sim$6.5X better than the best solver.

Future Directions

- Research focused on the development of portfolio solvers.
- Development of approximate techniques with higher scalability.

The End