Why are Hard Problems Easy to Solve in Practice?

Md Solimul Chowdhury

November 8, 2022

The problem of boolean Satisfiability is NP-Complete and intractable, in general. However, systematic search based boolean solvers, which tackle this hard problem, are surprisingly efficient in solving very large formulas from application domains. This behavior defies the theoretical understanding that the problem has exponential time-complexity in the worst case. This phenomenon has led to the belief that unlike the tractable algorithms, input formula size is not a good indicator of the hardness of boolean reasoning with systematic solvers. Explaining the performance behavior of system solvers has been studied under the lens of parameterized complexity, which concerns with determining the complexity of reasoning algorithms as a function of parameters of the input problems. Many parameters such as backdoor, backbone, tree-width, community structures, and mergability has been proposed to explain the scaling behavior of the systematic solvers. However, none of these parameters provide explanations which are both theoretically and empirically satisfying. Explaining the performance behavior of the systematic algorithms has been noted as one of the central problems in the area [1, 4].

Recently, it has been shown that two learned lemma related measures *learning rate of lemmas* and *average quality of the lemmas* are correlated with solvers performance [5]. In a spinoff of my PhD work [3], under a finer-grained experimental setting than [5], I discovered that the *proportion of highest quality learned lemmas* has very high correlation with the solvers performance, where the two measures proposed in [5] are unable to provide a consistent explanation. This result indicates that at a finer level, a solver's ability to learn highest quality lemma is a more accurate performance indicator of that solver. In another spinoff of my PhD work [2], I showed that performance of solvers is weakly correlated with *distributions of learned lemmas* over the course of the search.

Both of these results has brought out some preliminary insights on the relationship between performance behavior of systematic solvers and patterns of lemma learning, which extends the previous understanding. In the future, I want to pursue a deeper exploration on explaining performance behaviors of solvers in terms of learned lemmas. I intend to explore causal links between quality of generated lemmas and solver's performance towards developing a theoretical framework, and answer questions such as, exactly how learning of succession of highest-quality lemmas helps a solver to solve a given problem efficiently? Additionally, I want to revisit the scaling behavior of systematic solvers with respect to varying parameters for problems from application domains, and develop better understanding how these parameters influence solver behaviors.

References

[1] Albert Atserias, Armin Biere, Samuel Buss, Antonina Kolokolova, Jakob Nordstrom, and Karem Sakallah. Theoretical foundations of applied SAT solving. In *Workshop on Theoretical Foundations of Applied SAT Solving*, pages 1–10. Banff International Research Station for Mathematical Innovation and Discovery, 2014.

[2] Md Solimul Chowdhury. *Empirical Insights Driven CDCL SAT Algorithms.* PhD thesis, University of Alberta, 2021.

[3] Md Solimul Chowdhury, Martin Müller, and Jia-Huai You. Exploiting glue clauses to design effective CDCL branching heuristics. In *Proceedings of CP-2019*, pages 126–143.

[4] Vijay Ganesh and Moshe Y. Vardi. On the unreasonable effectiveness of SAT solvers. In *Beyond the Worst-Case Analysis of Algorithms*, pages 547–566. 2020.

[5] Jia Hui Liang, Hari Govind V. K., Pascal Poupart, Krzysztof Czarnecki, and Vijay Ganesh. An empirical study of branching heuristics through the lens of global learning rate. In *SAT-2017*, pages 119–135.