PhD Thesis Abstract

Title
Beyond the Structure of SAT Formulas

Author
Jesús Giráldez-Cru

Abstract
Nowadays, many *real-world* problems are encoded into SAT instances and efficiently solved by modern SAT solvers. These solvers, usually known as *Conflict-Driven Clause Learning* (*CDCL*) SAT solvers, include a variety of sophisticated techniques, such as clause learning, lazy data structures, conflict-based adaptive branching heuristics, or random restarts, among others. However, the reasons of their efficiency in solving real-world, or *industrial*, SAT instances are still unknown. The common wisdom in the SAT community is that these technique exploit some *hidden* structure of real-world problems.

In this thesis, we characterize some important features of the underlying structure of industrial SAT instances. Namely, they are the *community structure* and the *self-similar structure*. We observe that most industrial SAT formulas, viewed as graphs, have these two properties. This means that (i) in a graph with a clear community structure, i.e. having high *modularity*, we can find a partition of its nodes into communities such that most edges connect nodes of the same community; and (ii) in a graph with a self-similar pattern, i.e. being *fractal*, its shape is kept after *re-scalings*, i.e., grouping sets of nodes into a single node. We also analyze how these structures are affected by the effects of CDCL techniques during the search.

Using the previous structural studies, we propose three applications. First, we face the problem of generating pseudo-industrial random SAT instances using the notion of *modularity*. Our model generates instances similar to (classical) random SAT formulas when the modularity is low, but when this value is high, our model is also adequate to model realistic pseudo-industrial problems. Second, we propose a method based on the community structure of the instance to detect relevant learnt clauses. Our technique augments the original instance with this set of relevant clauses, and this results into an overall improvement of the efficiency of several state-of-the-art CDCL SAT solvers. Finally, we analyze the classification of industrial SAT instances into families using the previously analyzed structure features, and we compare them to other classifiers commonly used in portfolio SAT approaches.

In summary, this dissertation extends the understandings of the structure of SAT instances, with the aim of better explaining the success of CDCL techniques and possibly improve them, and propose a number of applications based on this analysis of the underlying structure of SAT formulas.

School: Artificial Intelligence Research Institute, Spanish National Research Council (IIIA-CSIC), and Universitat Autònoma de Barcelona (UAB)

Supervisors: Jordi Levy

Jesús Giráldez-Cru
giraldez@kth.se

KTH Royal Institute of Technology, Lindstedtsv 3, SE-100 44, Stockholm, SWEDEN

Graduated: May 24, 2016

Link to full text: http://www.tesisenred.net/handle/10803/386422

Also published in:
The Community Structure of SAT Formulas (SAT 2012)
The Fractal Dimension of SAT Formulas (IJCAR 2014)
A Modularity-based Random SAT Instances Generator (IJCAI 2015)
Using Community Structure to Detect Relevant Learnt Clauses (SAT 2015)
On the Classification of Industrial SAT Families (CCIA 2015)
Generating SAT Instances with Community Structure (Art.Int. 2016)

Notes:
Grade: Excellent Cum Laude.