Instance Features for Non-CNF Solver Portfolios

Mladen Nikolić Aleksandar Zeljić
Milan Todorović Predrag Janičić

5^{th} Workshop on Formal and Automated Theorem Proving and Aplications

Contents

Motivation — non-CNF solvers

- Non-CNF solvers — a new approach to solving SAT
- Exploit structure of the formulae
- Support various logical connectives ($\neg, \wedge, \vee, \Rightarrow, \Leftrightarrow$)
- Use compact structures to represent formulae
- NoClause, NFLsat, kw_aig, etc.

Motivation — portfolio SAT solvers

- Portfolio approach to CNF SAT is very successful
- Instance features are an important aspect of portfolio approach
- Instance features can be independent of particular portfolio algorithm

Related work

- SAT portfolios: SATzilla, ArgoSmArT variants, ISAC
- SAT as a classification task
- MAX-SAT portfolio
- QBF portfolio
- Tuning CSP systems

Instance features

- We propose a general set of non-CNF instance features
- All features can be computed in linear time with respect to input size
- We assume that input is a DAG
- Features include characteristics of both the DAG and the formula tree

Problem size features

- *Number of variables*
- *Number of logical connectives used in the formula tree* l_f
- *Number of logical connectives used in DAG* l_d
- *Size of the formula tree* s_t
- *Size of formula DAG* s_d
- *Compression factor and its reciprocal* s_t/s_d, s_d/s_t
- *Number of variable nodes in formula tree* v_f
- *Ratio of nodes representing logical connectives and nodes representing variables and its reciprocal (in formula tree)* l_f/v_f, v_f/l_f

Connectives based features

- *Distribution of frequencies (C_1) of connectives in formula tree*: distribution, entropy
- *Distribution of frequencies (C_2) of connectives in formula DAG*: distribution + entropy
- *Pearson correlation coefficient* $\rho(C_1, C_2)$
- *Ratio of effectively positive occurrences of each variable* $p/(p+n)$: mean, variation coefficient, and entropy
- *Ordered ratio of effectively positive occurrences of each variable* $\min\{p,n\}/(p+n)$: mean, variation coefficient, and entropy

Graph based features

- *Formula tree branch (root-leaf path) length*: mean, min, max, variation coefficient
- *DAG nodes outbound degree (connectives arity)*: mean, min, max, variation coefficient, entropy
- *DAG nodes inbound degree*: mean, min, max, variation coefficient, entropy
- *Ratio of DAG depth and number of variables and its reciprocal*
- *Ratio of DAG edges number and DAG nodes number and its reciprocal*

Evaluation of ArgoSmArT NCNF

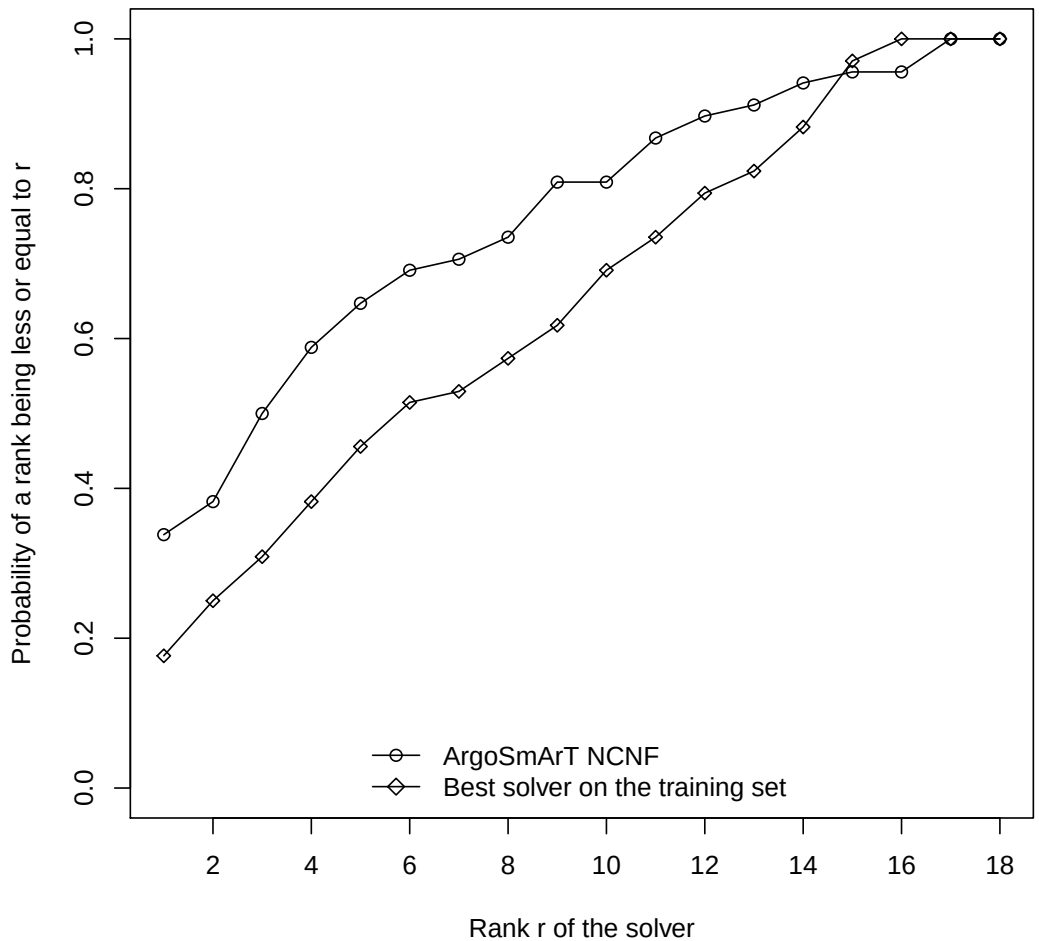
- Evaluation of a simple k nearest neighbors-based algorithm portfolio
- ArgoSmArT NCNF uses 3 solvers — Nflsat, Lingeling, and kw_aig
- Training set consisted of 746 instances
- Cutoff time was 600 seconds
- Test set consisted of SAT Race 2010 instances
- Instances were all in AIG format, but we can easily support other formats

Preliminary results

	N. solved	Time	PAR 10
Best on t.s.	62	42486	350286
ArgoSmArT NCNF	65	39675	322858
VBS	68	34885	294055
Gain	50%	37%	49%

- The minimal, mean and maximal feature computation times are 0.008, 0.89, and 21.9
- Largest instance had more than 4 million conjunction nodes

Preliminary results



Future work

- Feature selection might improve our results
- Running time prediction using proposed features
- Finding other applications for the proposed features