

Stuttgart, May 26th, 2025

PhD Thesis Filip Bártek

To whom it may concern:

please find below my review of the PhD thesis *Machine Learning for Saturation-Based Theorem Proving* submitted by Filip Bártek.

General description

The thesis deals with the broad topic of improving the performance of automated theorem provers. Theorem provers are programs that are able to show that certain conjectures can be deduced from a given set of assumptions. They have a wide field of applications, from software verification to legal reasoning. In particular, they support progress in mathematics and all domains depending on complex mathematical theories. This topic is hence highly relevant both in the narrow sense, but also in the wider sense.

The particular approach taken by Filip Bártek is the use of machine learning techniques (in the broader sense) to improve the performance. This thesis primarily comprises 5 different papers, with the first three representing one line of research (improving individual strategies of a theorem prover), the 4th and 5th a slightly different one (improving the performance by optimizing a portfolio of different strategies).

The thesis starts with the problem of finding a good precedence for the term orderings that form a critical component of the search strategy of modern theorem provers. It's based on manually selected symbol features, and addresses the problem in a name-agnostic way. I particularly like the way the problem is broken down to a pairwise symbol comparison.

The next chapter elaborates on the previous idea, but it replaces the manual feature extraction with a graph neural network that encodes the proof problem in a graph and uses the resulting embeddings for the learning process, yielding much better results.

This GNN approach is then taken up to lean clause evaluations. Again, the problem is reduced to learning a relative preferences between clauses in pairs, one of which was useful in previous proofs, the other not. The results are among the best I have seen for neural-

network based search heuristics for theorem provers.

The last two major chapters deal with portfolios of strategies. The first describes the automatic discovery of diverse new strategies by a combination of random probing and strategy optimization, and the efficient construction of schedules that solve the maximum number of proof problems within a given time limit. Chapter 6, the last research chapter, finally deals with the classification of problems, and picking different strategy schedules based on problem class.

The thesis concludes with a summary and outline of possible future work.

Evaluation

As stated above, the work is relevant both in the narrow sense, but also in the wider sense. The five papers that make up the core of this thesis represent careful experimental work and significant implementation effort. The results are always interesting, and usually lead to a significant improvement in the theorem prover under consideration. Several of the ideas presented are new not only in the published literature, but also among oral folklore of ATP developers. This shows a significant amount of creativity. I'm also happy with the depth of the research. The results are not low-hanging fruit, but rather the well-deserved reward for careful work and great persistence.

The thesis has been carefully prepared, and I found no significant problems with the content or the presentation. The mastery of the literature is not perfect, but quite good. The documentation of code and data is exemplary.

Overall, I enjoyed reading the thesis, and learned several useful new ideas.

The author of the thesis proved to have an ability to perform research and to achieve scientific results. I do recommend the thesis for presentation with the aim of receiving a Ph.D. degree.

Yours faithfully,



Stephan Schulz