Dr. Konstantin Korovin
Reader
Departement of Computer Science
The University of Manchester
Manchester M13 9PL

Tel +44 (0) 161 3067005
email konstantin.korovin@manchester.ac.uk

2nd June 2025

Review of PhD thesis Filip Bártek

The main focus of this thesis is development of machine learning methods for automated theorem proving in first-order logic. Automated theorem proving has many applications ranging from software and hardware verification to theorem proving in mathematics. In practice, automated theorem provers (ATPs) have many parameters that govern the proof search and are crucial for successfully solving complex problems. Unfortunately, in general there is no effective way of finding parameter values that are suitable for solving a given problem. Until recently, most theorem provers have been relying on ad hoc heuristics defined by developers or expert users.

The work presented in this thesis is aimed to address this problem using methods based on machine learning. Let me summarize the main results which are split over 5 chapters, based in peer reviewed publications.

1. In Chapter 2, the candidate considers the problem of choosing symbol precedence in simplification orderings LPO and KBO. Simplification orderings are used for reducing inferences in the superposition calculus and are essential for efficiency of ATPs. For n symbols there are $n!$ possible precedences, which makes the problem challenging. The candidate introduced the notion of a preference matrix based on pairwise symbol preferences which allowed for efficient model representation and training. The syntactic nature of symbols is abstracted away using symbol features so the model can be applied to problems containing unseen symbols or if the symbol meaning is not consistent across the problems. For the experiments a linear regression model Elastic-Net and Gradient Boosting regressor were trained on the TPTP library. The results are promising, and show performance comparable with human designed heuristics.

2. In Chapter 3, the candidate extended his work from Chapter 2 to an advanced formula representation using graph convolutional neural networks (GNN). This representation is both lossless and invariant under signature renaming. The last point is important for generalizabilty to unseen problems and signatures. A GNN model is trained to produce symbol embeddings which then used to estimate symbol costs, from which precedence cost can be computed. The training is based on comparison of ATP performance on pairs of precedences. The evaluation result show impressive ATP performance gain of 4.8% compared to the base line.

3. Chapter 4 is about GNN-based clause selection. Clause selection is another important parameter in theorem proving. The approach is based on

training GNN that can be used to compute symbol weights and in turn these weights can be used for scoring clauses. The key advantage of this approach is that symbol weights are computed at the beginning of problem solving and the clause scoring based on symbol weights has only negligible overhead during the proof search. This is an important feature as ATPs can produce millions of clauses which need to be dynamically scored and such efficient clause evaluation is essential for ATP performance. The new system increased the performance of a baseline configuration of Vampire by 6.6%, which is more than the increase introduced by AVATAR, a major technique in theorem proving introduced a number of years ago.

4. Chapter 5 is devoted to a system that discovers new complementary strategies and assembles them into schedules. This system based on several stages: random strategy probing, strategy optimization, and schedule construction. Impressively, only after 2 hours of training a schedule based on newly discovered strategies improved on the default strategy by 26% which is further improved after more training.

5. Chapter 6, extends work in Chapter 5 by specializing schedules to classes of problems. The schedule specialisation is done by trained collection of boosted trees which is used to select schedule to solve the given problem based on problem features.

It is clear that results presented in this thesis considerably advance state-of-the-art by introducing novel methods based on machine learning for discovering heuristics for automated theorem proving. Developed methods show impressive practical improvements over already very high performing automated theorem prover Vampire, the leading system in the area. The methods presented in this work are of high standards, with proper and extensive evaluations and unbiased analysis of the results. I believe the work presented in this thesis will have an important impact on developers of automated theorem provers and will be adopted by scientific community in this area.

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 sincerely,

Dr Konstantin Korovin