Journal of Symbolic Computation 69 (2015) 109–128

Contents lists available at ScienceDirect

Journal of Symbolic Computation

www.elsevier.com/locate/jsc

Learning-assisted theorem proving with millions of lemmas ☆

Cezary Kaliszyk[a], Josef Urban[b]

[a] University of Innsbruck, Austria
[b] Radboud University, Nijmegen, Netherlands

ARTICLE INFO

Article history:
Received 8 February 2014
Accepted 25 June 2014
Available online 24 October 2014

Keywords:
Flyspeck
Lemma mining
Machine learning
Artificial intelligence

ABSTRACT

Large formal mathematical libraries consist of millions of atomic inference steps that give rise to a corresponding number of proved statements (lemmas). Analogously to the informal mathematical practice, only a tiny fraction of such statements is named and re-used in later proofs by formal mathematicians. In this work, we suggest and implement criteria defining the estimated usefulness of the HOL Light lemmas for proving further theorems. We use these criteria to mine the large inference graph of the lemmas in the HOL Light and Flyspeck libraries, adding up to millions of the best lemmas to the pool of statements that can be re-used in later proofs. We show that in combination with learning-based relevance filtering, such methods significantly strengthen automated theorem proving of new conjectures over large formal mathematical libraries such as Flyspeck.

1. Introduction: automated reasoning over large mathematical libraries

In the last decade, large formal mathematical corpora such as the Mizar Mathematical Library (Grabowski et al., 2010) (MML), Isabelle/HOL (Wenzel et al., 2008) and HOL Light (Harrison, 1996a)/Flyspeck (Hales, 2006) have been translated to formats that allow easy experiments with ex-

☆ This research was partly supported by FWF grant P26201, and by NWO grant *Knowledge-based Automated Reasoning* 612.001.208.

E-mail addresses: cezary.kaliszyk@uibk.ac.at (C. Kaliszyk), josef.urban@gmail.com (J. Urban).
URLs: http://cl-informatik.uibk.ac.at/~cek/ (C. Kaliszyk), http://cs.ru.nl/~urban/ (J. Urban).

http://dx.doi.org/10.1016/j.jsc.2014.09.032
0747-7171/

ternal automated theorem provers (ATPs) and AI systems (Urban, 2004; Meng and Paulson, 2008; Kaliszyk and Urban, 2014).

The problem that has immediately emerged is to efficiently perform automated reasoning over such large formal mathematical knowledge bases, providing as much support for authoring computer-understandable mathematics as possible. Reasoning with and over such large ITP (interactive theorem proving) libraries is however not just a new problem, but also a new opportunity, because the libraries already contain a lot of advanced knowledge in the form of concepts, theorems, proofs, and whole theory developments. Such large pre-existing knowledge allows mathematicians to state more advanced conjectures, and experiment on them with the power of existing symbolic reasoning methods. The large amount of mathematical and problem-solving knowledge contained in the libraries can be also subjected to all kinds of knowledge-extraction methods, which can later complement more exhaustive theorem-proving methods by providing domain-specific guidance. Developing the strongest possible symbolic reasoning methods that combine such knowledge extraction and re-use with correct deductive search is an exciting new area of artificial intelligence and symbolic computation.

Several symbolic AI/ATP methods for reasoning in the context of a large number of related theorems and proofs have been suggested and tried already, including: (i) methods (often external to the core ATP algorithms) that select relevant premises (facts) from the thousands of theorems available in such corpora (Meng and Paulson, 2009; Hoder and Voronkov, 2011; Kühlwein et al., 2012), (ii) methods for internal guidance of ATP systems when reasoning in the large-theory setting (Urban et al., 2011), (iii) methods that automatically evolve more and more efficient ATP strategies for the clusters of related problems from such corpora (Urban, 2014), and (iv) methods that learn which of such specialized strategies to use for a new problem (Kühlwein et al., 2013b).

In this work, we start to complement the first set of methods – ATP-external premise selection – with *lemma mining* from the large corpora. The main idea of this approach is to enrich the pool of human-defined main (top-level) theorems in the large libraries with the most useful/interesting lemmas extracted from the proofs in these libraries. Such lemmas are then eligible together with (or instead of) the main library theorems as the premises that are given to the ATPs to attack new conjectures formulated over the large libraries.

This high-level idea is straightforward, but there are a number of possible approaches involving a number of issues to be solved, starting with a reasonable definition of a *useful/interesting lemma*, and with making such definitions efficient over corpora that contain millions to billions of candidate lemmas. These issues are discussed in Sections 4 and 5, after motivating and explaining the overall approach for using lemmas in large theories in Section 2 and giving an overview of the recent related work in Section 3.

As in any AI discipline dealing with large amount of data, research in the large-theory field is driven by rigorous experimental evaluations of the proposed methods over the existing corpora. For the first experiments with lemma mining we use the HOL Light system, together with its core library and the Flyspeck library. The various evaluation scenarios are defined and discussed in Section 6, and the implemented methods are evaluated in Section 7. Section 8 discusses the various future directions and concludes.[1]

2. Using lemmas for theorem proving in large theories

The main task in the automated reasoning in large theories (ARLT) domain is to prove new conjectures with the knowledge of a large body of previously proved theorems and their proofs. This setting reasonably corresponds to how large ITP libraries are constructed, and hopefully also emulates how human mathematicians work more faithfully than the classical scenario of a single hard problem consisting of isolated axioms and a conjecture (Urban and Vyskočil, 2013). The pool of previously proved theorems ranges from thousands in large-theory ATP benchmarks such as MPTP2078 (Alama et al., 2014), to tens of thousands when working with the whole ITP libraries.[2]

[1] This paper is an extended version of (Kaliszyk and Urban, 2013b).

[2] 23,323 theorems are in the HOL Light/Flyspeck library (SVN revision 3437), about 20,000 are in the Isabelle/HOL library, and about 50,000 theorems are in the Mizar library.