University of Leicester

Boosting Automated Reasoning by Mining Existing Proofs

Thomas Gransden – Department of Computer Science, University of Leicester – tg75@student.le.ac.uk

Motivation

There are many proofs that rely on human intervention to provide the necessary sequence of proof steps. This is especially true in significant mathematical and industrial proofs. This project investigates an approach that looks to provide greater automation for theorem provers by applying data mining techniques to the large libraries of proofs that are available.

Example

We show an example proof about logarithms from the Isabelle/HOL library that cannot be fully automated using the standard Isabelle proof tools – human intervention is required to guide the proof:

Lemma powr_divide2 : "x powr a / x powr b = x powr (a-b)"

apply (simp add: powr_def)
apply (subst exp_diff [THEN sym])
apply (simp add: left_diff_distrib)
done

Combining Data Mining and Theorem Proving

In Higher Order Theorem Proving, emphasis has been put on providing the user with proof hints based on existing proofs. Some example applications are:

- **PGTips Recommender System** – This system gives statistical hints based on commonly occurring sequences of tactics in proof libraries.
- **ML4PG [2] –** makes use of clustering algorithms and suggests proofs that are similar to the goal that is trying to be proven.

In First Order Theorem Proving, much work has been carried out by Urban et al [4] on the premise selection problem. Broadly, this is the problem of selecting prior knowledge from large theories libraries that is most useful for proving a new conjecture.

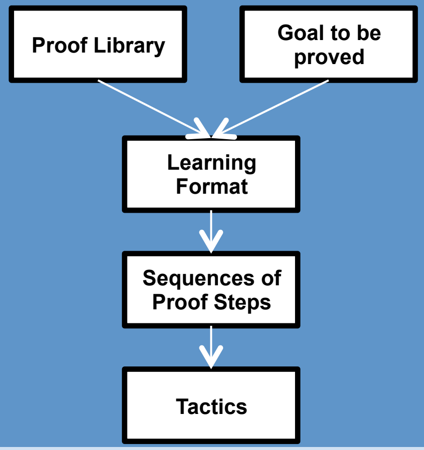
Tactic Mining Approach

This aim of this project is to build upon the existing work and implement a tactic mining tool for theorem provers. The tactics will be formed from sequences of proof steps that are contained in proof libraries.

There has been previous tactic mining work by Hazel Duncan [3]. Duncan's approach searched for commonly occurring sequences of proof steps and then used genetic programming techniques to combine these sequences into tactics. In the evaluation, the tactics proved moderately effective in terms of their applicability.

We provide an outline of our tactic mining strategy:

- Abstract information about the proof library and current goal into an appropriate learning format
- Identify useful proof step sequences from the library that could help to prove the goal.
- Generalise the proof steps into tactics.
- Apply the tactics and prove the goal.
- Update the proof library with the new information.



Research Questions

1) How can we deal with the complexities of Higher Order proofs?

The complex higher order language may have constructs such as variable instantiations and proof directives that make the proof extremely specific. We must find ways of managing this information so that we can generalize sequences of proof steps into more widely applicable tactics.

2) How can proof patterns be found?

As explained in Bundy et al 's work[1], a proof can be thought of as a hierarchy of levels – the tactic level, the goal level and the proof tree level. An important open research question is which techniques can help to find patterns in each level of the proof.

3) How can we interface between the theorem prover and tactic miner?

We must make sure that the contents of the proof libraries are abstracted into a format that can be understood by the tactic miner, and the results from the tactic miner must be interpreted back into the theorem proving environment.

4) Can we incorporate negative information to learn from?

By including failed and discarded proof derivations, we will be able to learn sequences of proof steps that don't work. This would allow us to implement a supervised tactic inference approach.

References

[1] – A Statistical Relational Learning Challenge – Bundy et al. *ICML 2012*

[2] – Machine Learning in Proof General. Komendantskaya and Heras. *UITP 2012*

[3] – The Use Of Data Mining for the Automatic Formation of Tactics. Hazel Duncan. *University of Edinburgh, 2007*

[4] – Learning semantics for Automated Reasoning. Urban et al. *NIPS Workshop on Learning Semantics*, 2011

Work supervised by Dr. Neil Walkinshaw and Prof. Rajeev Raman.