An application of machine learning to RCF decision procedures

Zongyan Huang
zh242@cam.ac.uk

Abstract

Machine Learning has been applied to the problem-dependent selection of the most efficient decision procedure in the theorem prover MetiTarski. The machine-learned selection process did better than any fixed decision procedure (Z3, Mathematica and QEPCAD).

Introduction

- MetiTarski is an automatic theorem prover which can prove logical statements involving special functions (log, sin, cos, etc.,.) During its proof search, MetiTarski generates sequences of subproblems in the decidable theory of RCF.
- The theory of Real closed fields (RCF) concerns boolean combinations of polynomial equations and inequalities over the real numbers. No single RCF decision procedure always gives the fastest runtime
- Machine learning uses statistical methods to extract information from supplied training examples, which it can then apply to new problems
- The Support Vector Machine (SVM) was used to do the classification in this work (Joachims' SVM-Light)

Methodology

For each decision procedure, MetiTarski was run on all problems. Then for each problem, the best decision procedure is the one for which MetiTarski solved the problem in least time.

Three main tasks:

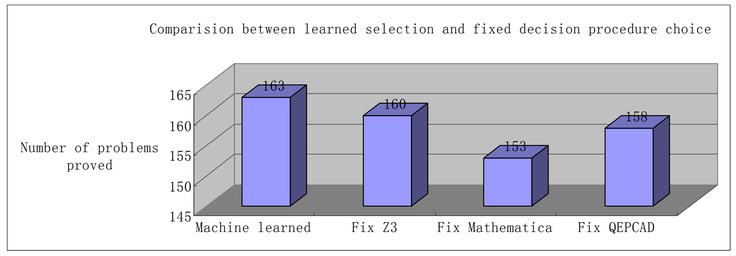
- Identify features of the MetiTarski problems that might be relevant to the correct choice of the RCF decision procedure (which special functions, number of variables, etc.,.)
- Select the best kernel function and parameter values for SVM-Light based on F_1-score maximization
- Combine the models for the decision procedures and compare the margin values produced by three decision procedure classifiers. The classifier with most positive (or least negative) margin was selected.

UNIVERSITY OF CAMBRIDGE
Computer Laboratory

Results

The experiment was done on 825 MetiTarski problems in the TPTP format. The data was randomly split into a learning set (418 problems), a validation set (213 problems) and a test set (194 problems). A time limit of 60 CPU seconds was set for each proof attempt. The total number of problems proved out of 194 test problems was used to measure the efficacy of the machine learned selection process.

Using machine learning to select the best decision procedure yields better results than any individual fixed decision procedure.



Future work

- Extend to the heuristic selection within decision procedures
- Extend the range of features used and apply feature selection
- Provide feedback for development of RCF decision procedures