WST 2004

RISE OF THE TOOLS

REPORT ON THE TERMINATION COMPETITION

CLAUDE MARCHE, ALBERT RUBIO

Brief History

At WST'2003, Albert Rubio initiated:

- An exhibition/competition of termination tools
- The *Termination Problems Data Base*, made of contributions from participants (TRSs and LPs)
- A discussion for a common syntax for the TPDB

After WST'2003:

- The common syntax has been fixed
- A common syntax for SRSs also fixed
- TPDB in the common syntax: `http://www.lri.fr/~marche/wst2004-competition`

Rules of the WST’2004 competition

- Each participant provides its tool sources or binaries to be run on a PC (Pentium 4, 2.8 GHz, 512 Mo RAM, Debian GNU/Linux)
  - input: a termination problem file in TPDB syntax
  - output: 3 possible answers: YES, NO, anything else
- 3 categories corresponding to input syntax: TRS, SRS and LP
- Each tool run automatically on all examples of the TPDB.
  - CPU time limit set to one 1 minute
  - Results (answers and time) were available live on the Web.
  - YES and NO answers counted separately. Time does not count
  - Several scoring for TRS: standard, modulo theory, innermost strategy, etc.

Results

Preamble: there are several reasons why the evaluations are not accurate:

- some syntax errors when reading problem files
- TPDB not made for evaluation of tools:
  - origin of problems: old ones in the literature + submitted by participants
  - there are redundancies
  - not necessarily representative of "usual" termination problems
  - not enough "hard" problems

Results: SRS category

- Number of Pbs: 104
- TORPA: 88 (84 %) (6 NO)
- AProVE: 87 (83 %) (5 NO)
- MatchBox: 49 (47 %) (7 NO)
- TTT: 33 (31 %) (2 NO)
- CiME: 24 (23 %) (0 NO)

Results: TRS category, standard rewriting

- Number of Pbs: 514
- AProVE: 410 (79 %) (22 NO)
- TTT: 397 (77 %) (12 NO)
- CiME: 297 (57 %) (1 NO)
- Matchbox: 96 (18 %) (21 NO)
- No other participants. . .

Results: TRS category, modulo theory

- Theory = C or AC
- Number of Pbs: 60
- AProVE: 54 (90 %)
- CiME: 46 (76 %)
- No other participants. . .

Results: TRS category, innermost strategy

- Number of Pbs: 58
- AProVE: 55 (94%) (2 NO)
- TTT: 43 (74 %)
- Cariboo: 35 (60 %)
- CiME: 13 (22%)
- Matchbox: 7 (12 %)

Results: TRS category, other sub-categories

Sub-categories with only 1 competitor:

- Context-sensitive strategies : 41 pbs, AProVE : 32 (78 %)
- Conditional rewriting: 6 pbs, AProVE: 5 (83 %)
- Outermost strategy: 6 pbs, Cariboo: 6 (100 %)

Results: LP category

- Number of Pbs: 190
- AProVE: 120 (63 %) (2 NO)
- TALP: 111 (58 %)
- No other participants. . .

Conclusions

Today, AProVE is the termination tool of choice

- handles each category/variant (except outermost strategy. . .)
- implements most of the up-to-date techniques

- . . . but there are pbs not solved by AProVE and solved by others
- The competition boosted the development of tools.
- Efforts to make them fully automatic
  - Fair comparison between tools.
  - Suitable for a naive user.
  - Suitable for being used as back-end tools
    (Remark: tools should have both a specific user interface and a command line version reading files in common syntax)

Scientific issues

- Analysing pbs not solved by any tools:
  - Scaling up issues
  - Class of difficult termination problems:
    - Problems coming from translation (encoding strategies, sorts, etc.)
    - Mix with numerical computations (LP category)
    - Others ?
- Deeper analysis of non-termination
- Generating proof trace

Next competition

- A long time before the next workshop ?
- With many more tools. . .
- More categories: higher order TRS, . . . , Typed LP, Functional-logic programming
- With an improved TPDB
- Only on a selected part of the TPDB, with more time on each pb.

Volunteers seeked for improving the TPDB !

EXHIBITION OF THE TOOLS

TOMORROW, 16:00, ROOM SFO 1