Choosing a Solver for Your Query

An SVM-based Query Classification

Oscar Soria Dustmann, Felix Rath, Philipp Martin, Klaus Wehrle

Observation

Instrs	Time(s)	TSolver(%)	Queries	[avgQT]
249643898	4201.31	95.66	449725	9 ms

Tool: `dirname`, randomly chosen

Observation

Instrs	Time(s)	TSolver(%)	Queries	[avgQT]
249643898	4201.31	95.66	449725	9 ms

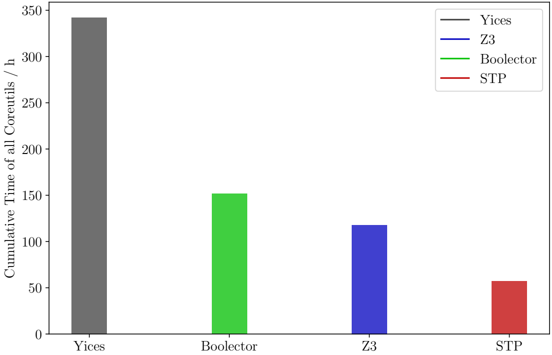
Tool: `dirname`, randomly chosen

Which solver is the best?

Observation

Instrs	Time(s)	TSolver(%)	Queries	[avgQT]
249643898	4201.31	95.66	449725	9 ms

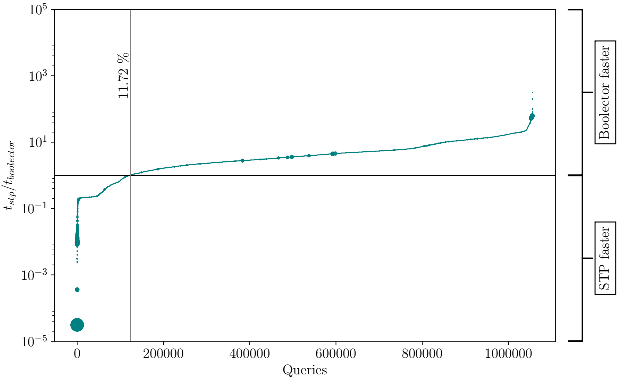
Tool: `dirname`, randomly chosen



Observation

Instrs	Time(s)	TSolver(%)	Queries	[avgQT]
249643898	4201.31	95.66	449725	9 ms

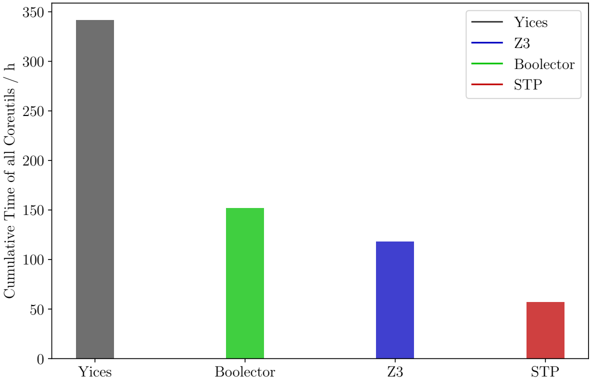
Tool: `dirname`, randomly chosen



Solver Choice

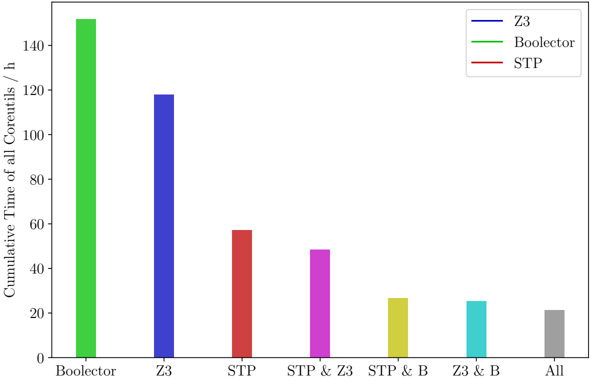
Choose Best Solver for Each Query

Choose Best Solver for Each Query

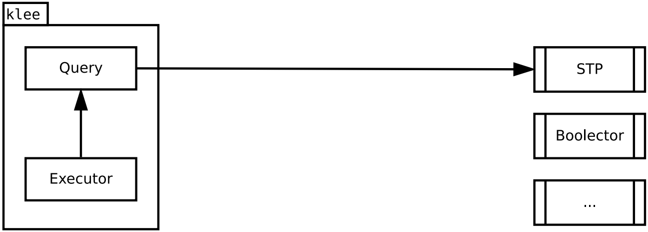


Solver Choice

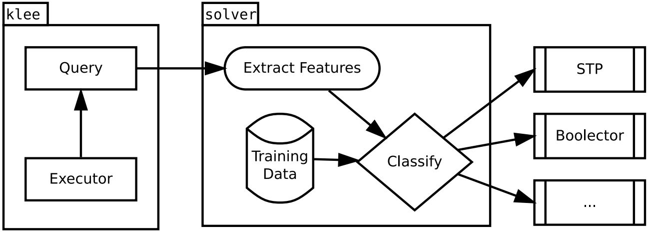
Choose Best Solver for Each Query



Choose Best Solver for Each Query

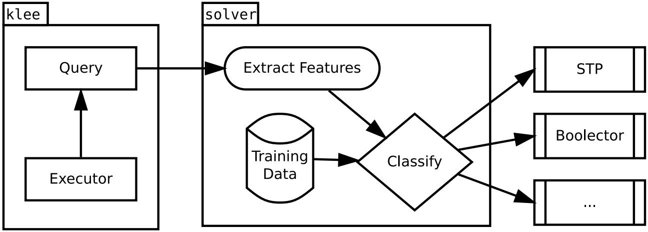


Choose Best Solver for Each Query with Machine Learning



Solver Choice

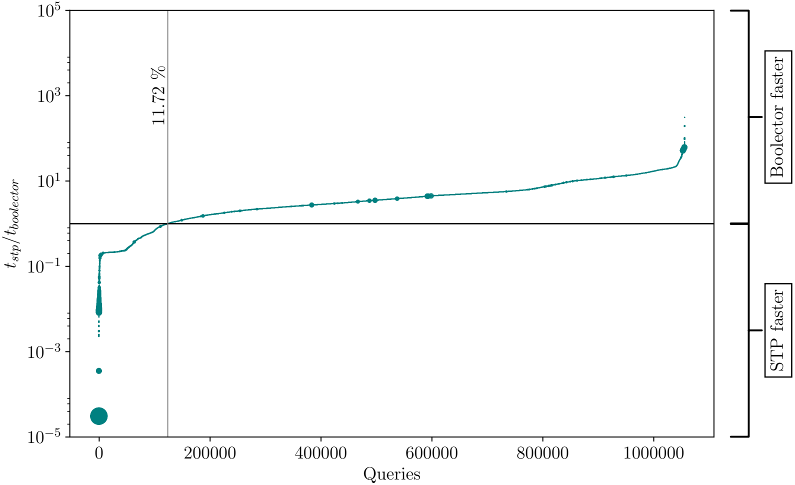
Choose Best Solver for Each Query with Machine Learning



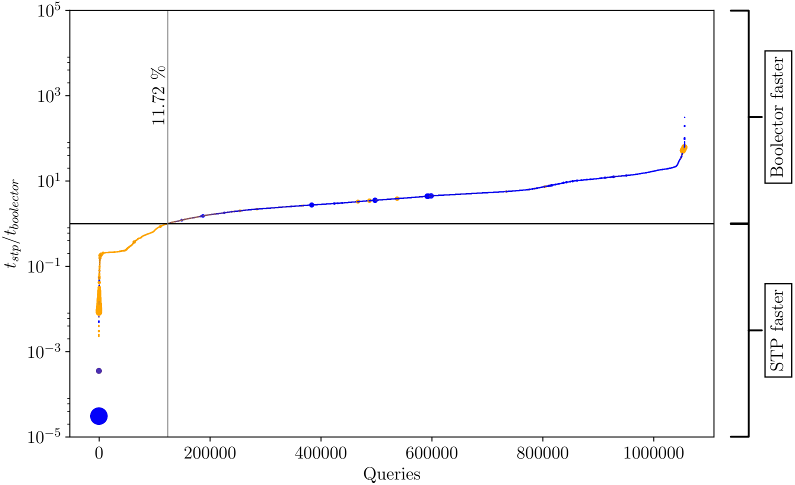
Features

- size, domain and range of arrays
- operators, constants and variables in constraints
- occurrence of the latter in the query overall

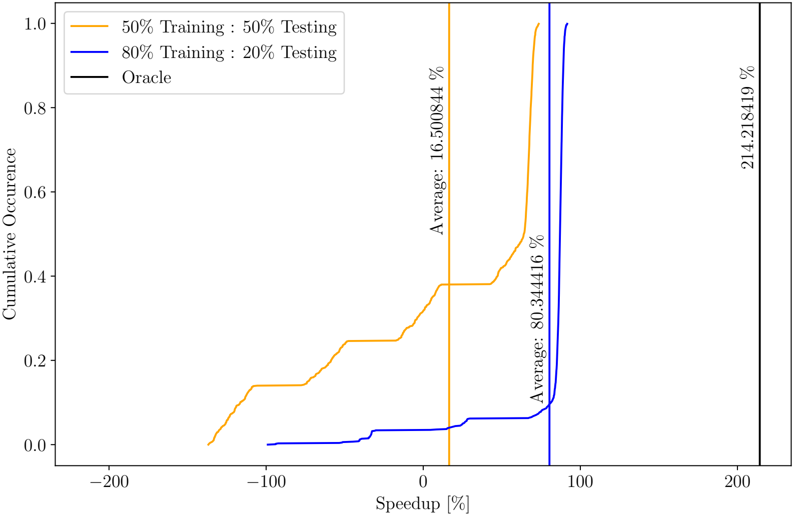
Classification Success Rate (100:100)



Classification Success Rate (100:100)



Influence of the Training Set



Application Area

- Deployment

Application Area

- Deployment
  - Statically trained

Application Area

- Deployment
  - Statically trained
  - On-the-fly learning in each run

Conclusion

- Deployment
  - Statically trained
  - On-the-fly learning in each run
- High potential for speed up

Conclusion

- Deployment
  - Statically trained
  - On-the-fly learning in each run
- High potential for speed up
- Feature extraction: No relevant overhead

Conclusion

- Deployment
  - Statically trained
  - On-the-fly learning in each run
- High potential for speed up
- Feature extraction: No relevant overhead
  - due to caching

Conclusion

- Deployment
  - Statically trained
  - On-the-fly learning in each run
- High potential for speed up
- Feature extraction: No relevant overhead
  - due to caching
  - feature set linearly extractable

Conclusion

- Deployment
  - Statically trained
  - On-the-fly learning in each run
- High potential for speed up
- Feature extraction: No relevant overhead
  - due to caching
  - feature set linearly extractable
  - SVM fast classification

Conclusion

- Deployment
  - Statically trained
  - On-the-fly learning in each run
- High potential for speed up
- Feature extraction: No relevant overhead
  - due to caching
  - feature set linearly extractable
  - SVM fast classification
- Future work:

Conclusion

- Deployment
  - Statically trained
  - On-the-fly learning in each run
- High potential for speed up
- Feature extraction: No relevant overhead
  - due to caching
  - feature set linearly extractable
  - SVM fast classification
- Future work:
  - Understand/fix over-selection of Boolector

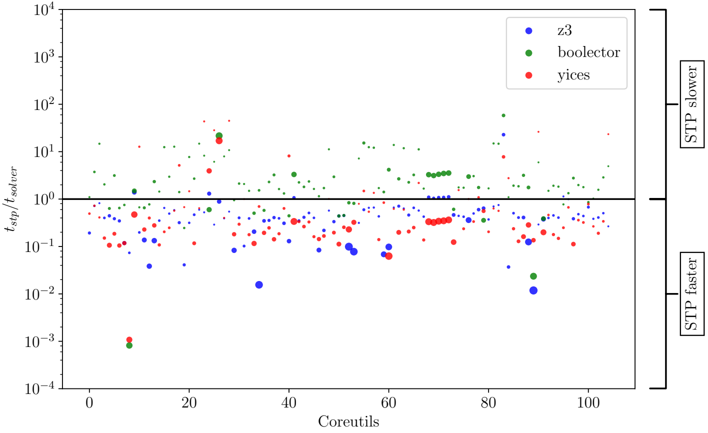
Conclusion

- Deployment
  - Statically trained
  - On-the-fly learning in each run
- High potential for speed up
- Feature extraction: No relevant overhead
  - due to caching
  - feature set linearly extractable
  - SVM fast classification
- Future work:
  - Understand/fix over-selection of Boolector
  - Identify more relevant properties of queries

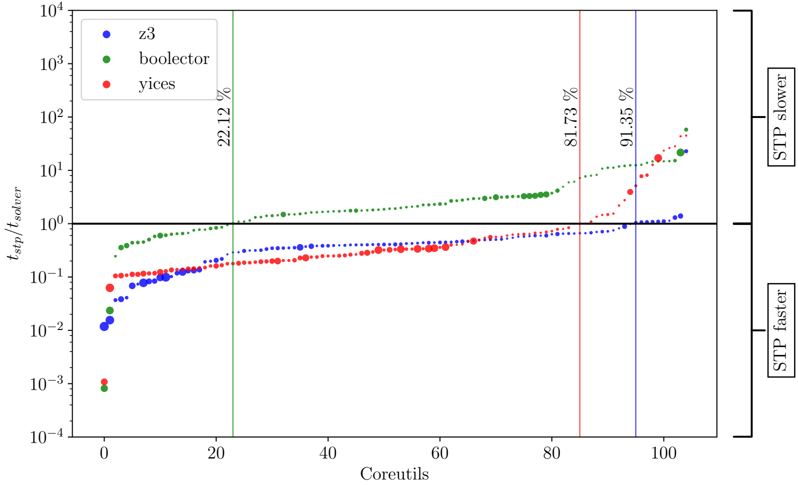
Conclusion

- Deployment
  - Statically trained
  - On-the-fly learning in each run
- High potential for speed up
- Feature extraction: No relevant overhead
  - due to caching
  - feature set linearly extractable
  - SVM fast classification
- Future work:
  - Understand/fix over-selection of Boolector
  - Identify more relevant properties of queries
  - Find reliable / stable training set

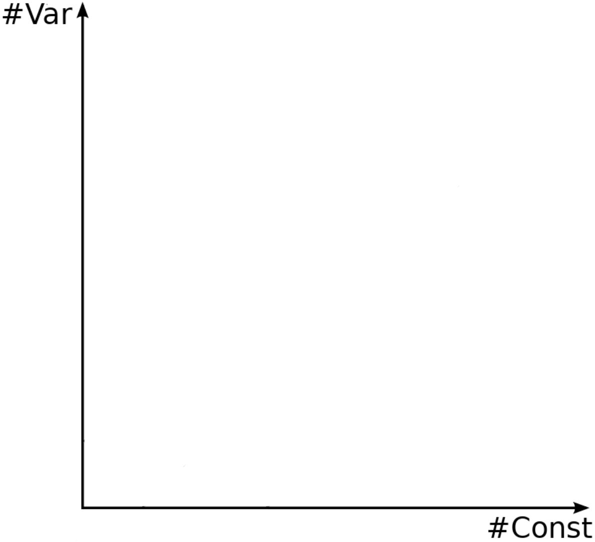
Per-Tool-Comparison



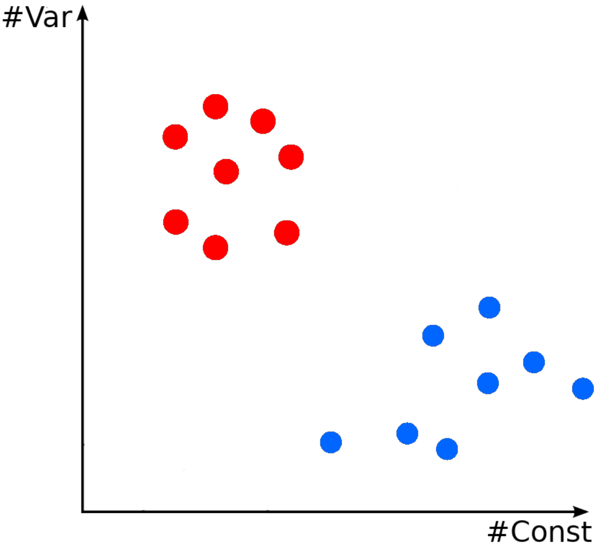
Per-Tool-Comparison



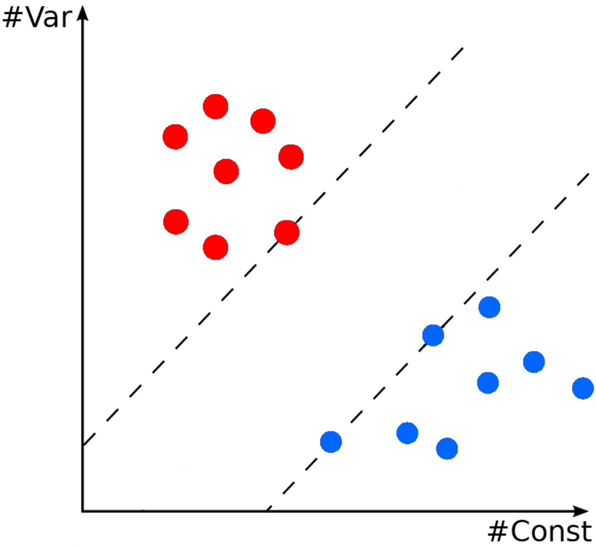
SVM

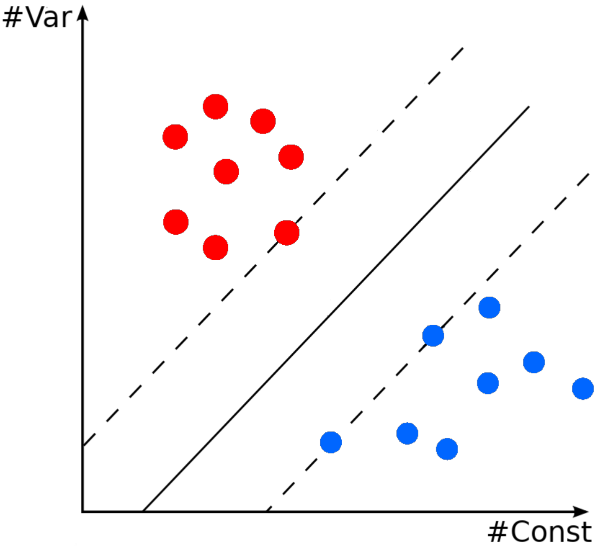


SVM



SVM




#Var
#Const

Feature Set

Array features	Constraint features	Query features
Number of Arrays	Number of constraints	Constant count
Total size	Total constant count	Variable count
Maximum size	Maximum constant count	Operator count
Minimum size	Minimum constant count	Operator counts by type
Average size	Average constant count	
Maximum domain	Total variable count	
Minimum domain	Maximum variable count	
Average domain	Minimum variable count	
Maximum range	Average variable count	
Minimum range	Total operator count	
Average range	Maximum operator count	
	Minimum operator count	
	Average operator count	
	Operator counts by type	