Formal Language Driven Data Analysis Research Group Report

Semyon Grigorev

Saint Petersburg State University

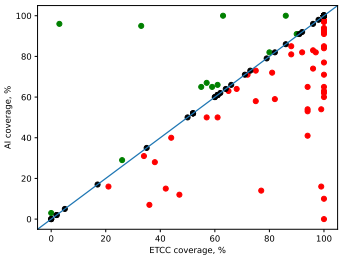
May 15, 2024

$\mathcal{AI}$-Guided Symbolic Execution

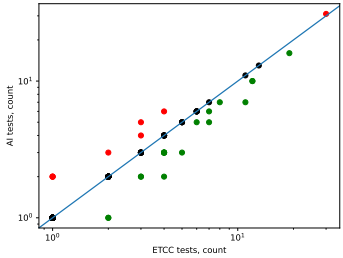
- ✔ Reusable infrastructure for training developed and implemented
  - Wrapper for SVM to convert it to server
  - Python client — $\mathcal{AI}$ agent to training
  - Basic manipulation with neural networks
- ✔ Basic dataset: train/validation/test
- ✔ Basic performance tuning
- ✔ First version of $\mathcal{AI}$ agent which guide SVM
  - Can be used in various engines: we use language-independent features
  - With tuning
- ⚙ Dataset extension
- ⚙ GNN quality improvement
- ⚙ Training infrastructure improvement: performance, flexibility, documentation, ...

$\mathcal{AI}$-Guided Symbolic Execution: Preliminary results

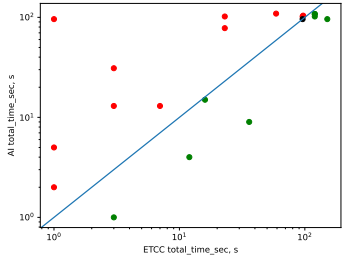
- 190 methods: algorithms, data structures, real-world projects
- AI — our AI-based agent
- ETCC (ExecutionTreeContributedCoverage) — one of the best algorithmic strategies



coverage comparison on the same methods
red: ETCC win, green: AI win, black: equal



tests comparison on the same methods
red: ETCC win, green: AI win, black: equal, logscale



total_time_sec comparison on the same methods
red: ETCC win, green: AI win, black: equal, logscale

- Same coverage: 132, AI less: 46, AI more: 12
- For same coverage: less tests, comparable time