Automatically Generating Features for Learning Program Analysis Heuristics for C-like Languages

Kwonsoo Chae (KU), Hakjoo Oh (KU), Kihong Heo (SNU), Hongseok Yang (Oxford)

1. Problem
- Common limitations of existing data-driven approaches → manual feature engineering:
  Time-consuming  Domain knowledge  Analysis-dependent

What people expect Reality

Our goal: Automatically generate features to automatically represent each query as a feature vector (B^k).

2. Learning Analysis Heuristics
- Example: partially flow-sensitive interval analysis

3. Automatic Feature Generation
- Query prediction & final analysis

4. Results

5. Takeaways
- A framework for automatically generating features for learning program analysis heuristics
- A method that uses a program reducer for generating feature programs, which capture important behaviors of static analysis
- The notion of abstract data flow graphs as generic features in data-driven static analysis
- Experimental evaluations with three different kinds of static analyses.

6. Conclusion
- We have identified and solved the problem of manual feature engineering in data-driven static analysis.