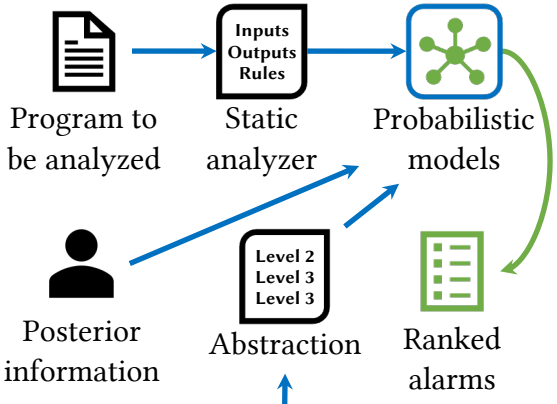
Learning Abstraction Selection for Bayesian Program Analysis

Yifan Zhang, Yuanfeng Shi, Xin Zhang
Peking University, China

Bayesian Program Analysis



Challenges

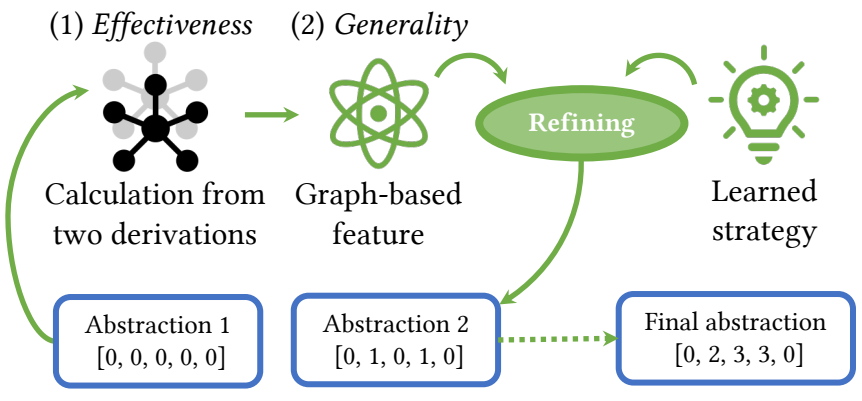
(1) *Effectiveness*: conventional approaches are **ineffective.**

Finer abstraction → **Better generalization ability**

(2) *Generality*: Our method should work well for **a wide range of** Bayesian analyses.

Pointer **Datarace** **Taint**

Learning-Based Approach: BinGraph

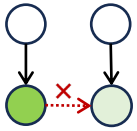


Abstraction Selection

Goal: optimization for generalization ability

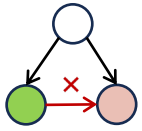
Too precise

Prevent posterior information from propagating to relevant analysis results effectively:
Over-fitting



Too coarse

Cause posterior information to propagate to irrelevant analysis results falsely:
Under-fitting



Evaluation

Better generalization ability than all baselines

	Program	Inversion				Rank-100%-T				Rank-90%-T			
		BinGraph	Base-C	Base-P	Base-R	BinGraph	Base-C	Base-P	Base-R	BinGraph	Base-C	Base-P	Base-R
Datarace analysis	avrora	6,249	3,852	6,944	7,341	761	938	717	744	364	421	392	415
	ftp	388	1,173	432	540	84	169	83	86	77	112	75	77
	sunflow	10,055	17,790	*failed*	14,443	359	460	*failed*	961	254	429	*failed*	315
	raytracer	19	87	18	37	10	32	9	15	10	32	9	15
	luindex	22	32	646	236	13	18	325	120	13	18	325	120
	xalan	*failed*	14,868	*failed*	*failed*	*failed*	326	*failed*	*failed*	*failed*	319	*failed*	*failed*
	Average		31.52%↓	27.81%↓	42.55%↓		37.53%↓	19.39%↓	37.04%↓		36.42%↓	22.34%↓	30.83%↓
Thread-escape analysis	hedc	5,991	14,381	7,045	10,065	372	396	379	393	303	337	316	325
	jspider	20,033	36,838	35,517	40,411	635	662	618	629	451	500	527	540
	montecarlo	496	1,812	583	1,969	78	163	78	154	59	156	61	122
	pool	9,591	20,226	10,450	18,423	371	402	395	411	323	369	361	377
	raytracer	6,101	9,756	6,822	8,353	287	314	319	339	263	287	286	281
	toba-s	53,419	118,509	76,604	104,175	1,278	1,222	1,278	1,256	960	1,052	1,149	1,103
	Average		53.59%↓	20.42%↓	48.22%↓		12.34%↓	2.53%↓	12.84%↓		18.61%↓	9.47%↓	18.10%↓