

정적 분석 시간 예측

(Online Estimation of Static Analysis Runtime)

이우석

서울대학교 프로그래밍 연구실

2012. 1. 31

9회 소프트웨어 무결점 연구센터 워크샵

동기

- Airac for C

Program	LOC(*.i)	Time
nethack-3.3	211K	4h 30m
emacs-22.1	399K	10h 30m
ghostscript-9.0	1,363K	4h 6m

- Scandal for Android Apps

Application	LOC(*.dex)	Time
Gold Miner	191	1m 21s
Mini Army	480	3m
Xing Metro	253	6h 24m

희망사항

- 분석기에 장착

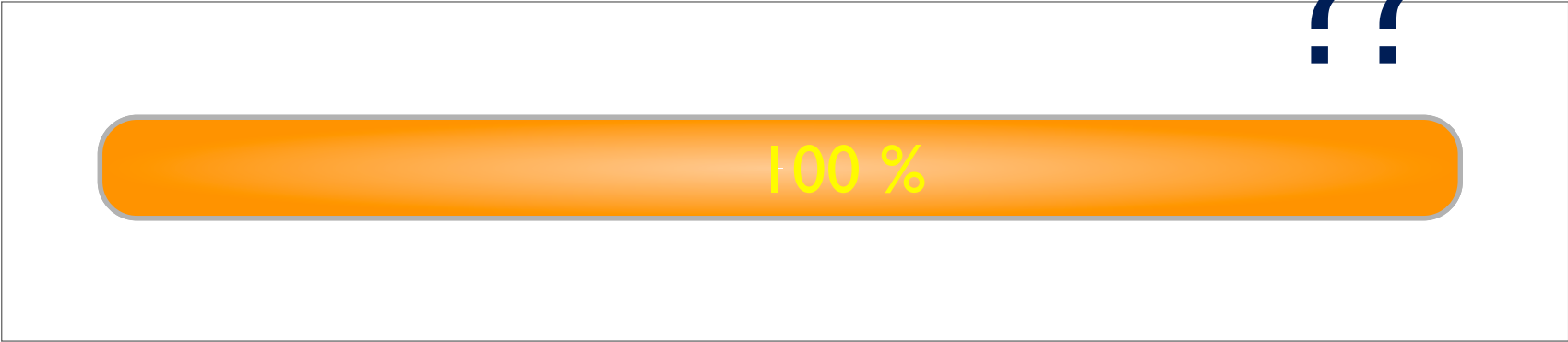


1200 iters / sec - About **2 hours** left.

어려운 점

- 종료 상태를 분석 전에 모름

??



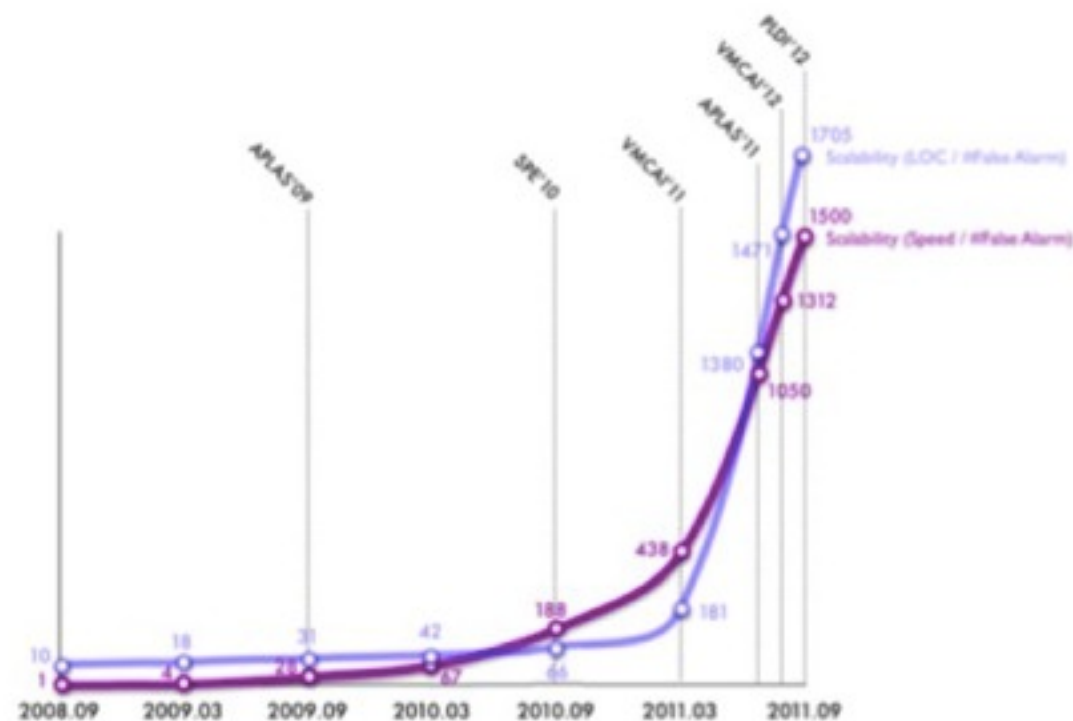
100 %

어려운 점

- 분석기는 계속 진화함



sound-&-global version

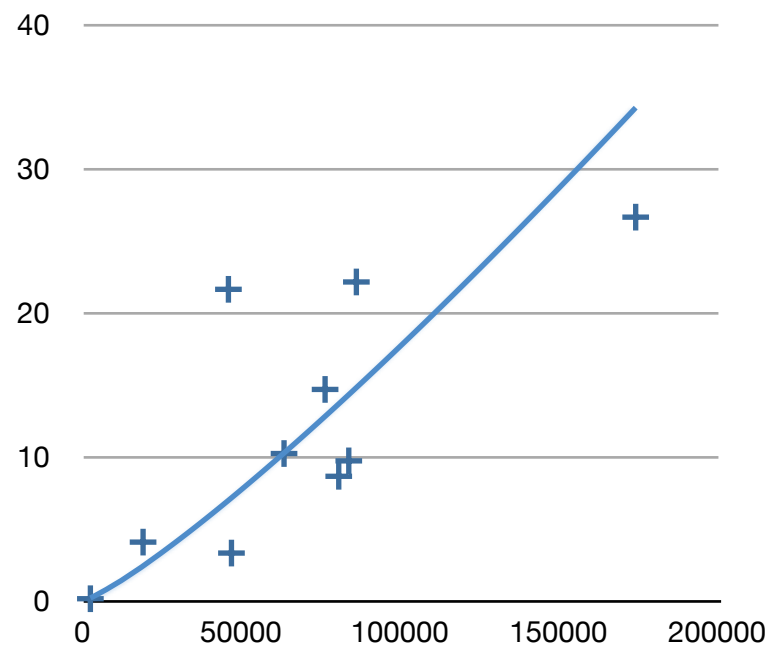


방향

- 알고리즘의 실행시간을 예측하는 방법론[†] - Empirical Hardness Models
 - 1. 성능과 관계된 특성들(features) 정하기
 - 2. 학습용 데이터들에 대해 특성값과 실행시간들 도출
 - 3. (특성값 → 실행시간) 함수 학습 후, 입력되는 문제에 적용

[†] Kevin Leyton-Brown et al., *Empirical hardness models : Methodology and a case study on combinatorial auctions*, JACM 2009

선형 회귀



$$f_w(x_i) = \sum_i w_i x_i$$

가중치 벡터

실행시간 함수

특성 벡터

방향

- 특성들로 실행 중 알 수 있는 값들을 이용하면,
 - 과거적 특성(Historical features) : 지금까지 계산한 것
 - 미래적 특성(Predictive features) : 앞으로 계산할 것
- 실행 중 남은시간 추측(Online Estimation) 가능
- 참고 논문 :
Shai Haim et al., *Online Search Cost Estimation for SAT Solvers*, AAAI 2008

관련연구 : SAT 풀이시간 추정

<i>Feature</i>
Number of <i>variables</i> (<i>var</i>)
Number of <i>clauses</i> (<i>cls</i>)
<i>cls/var</i>
<i>var/cls</i>
Fraction of Binary Clauses
Fraction of Ternary Clauses
Avg. Clause Size
Search Depth (from assignment stack)
Search Depth (in corresponding binary tree) ^a
Backjump Size
Learnt Clause Size
Conflict Clause Size
Fraction of assigned vars before backtracking (<i>abb</i>)
Fraction of assigned vars after backtracking (<i>aab</i>)
<i>aab.mean/abb.mean</i>
<i>abb.mean/aab.mean</i>
$\log(WBE)$

관련연구 : SAT 풀이시간 추정

과거적
특성

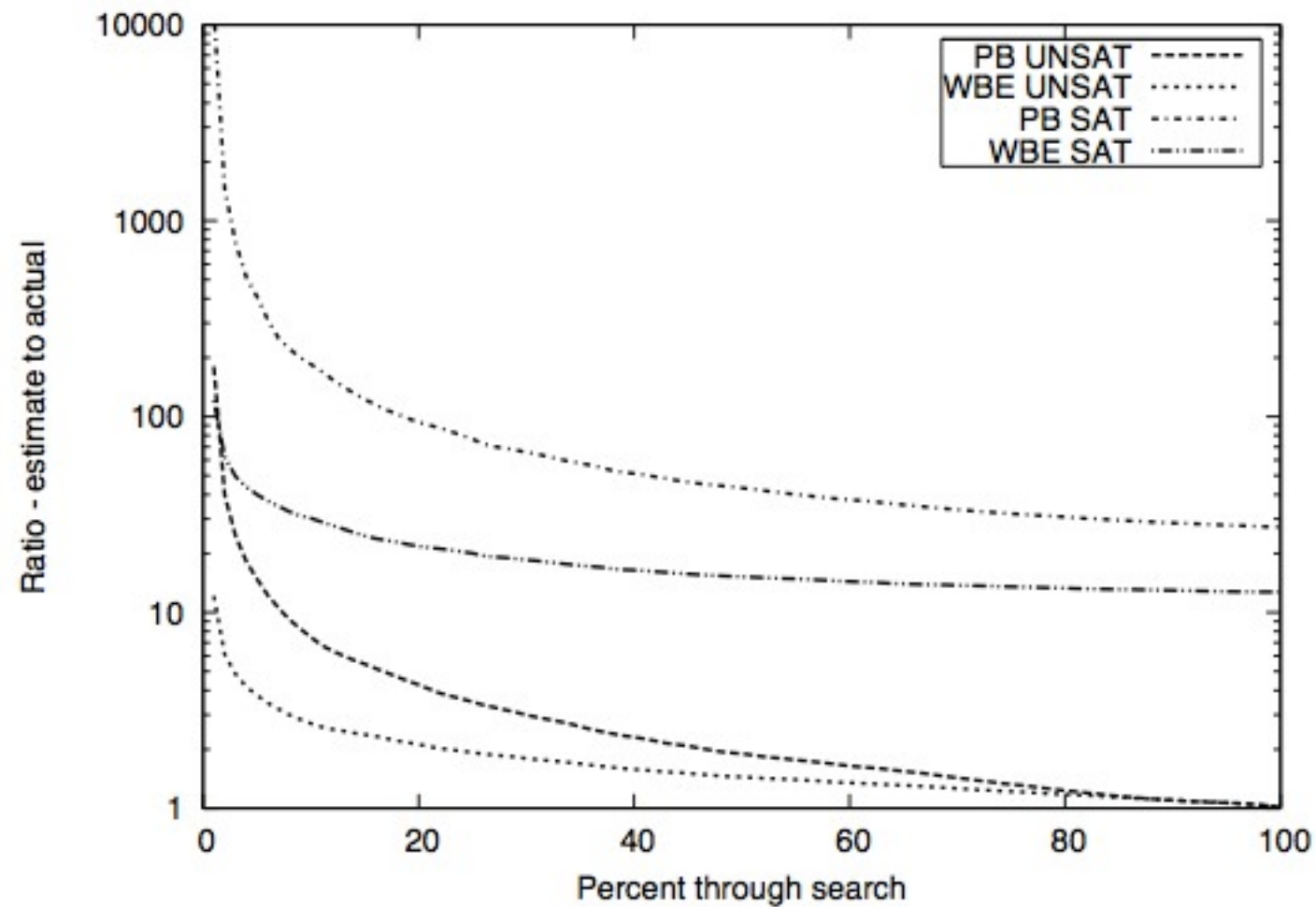
<i>Feature</i>
Number of <i>variables</i> (<i>var</i>)
Number of <i>clauses</i> (<i>cls</i>)
<i>cls/var</i>
<i>var/cls</i>
Fraction of Binary Clauses
Fraction of Ternary Clauses
Avg. Clause Size
Search Depth (from assignment stack)
Search Depth (in corresponding binary tree) ^a
Backjump Size
Learnt Clause Size
Conflict Clause Size
Fraction of assigned vars before backtracking (<i>abb</i>)
Fraction of assigned vars after backtracking (<i>aab</i>)
<i>aab.mean/abb.mean</i>
<i>abb.mean/aab.mean</i>
$\log(WBE)$

관련연구 : SAT 풀이시간 추정

미래적
특성

<i>Feature</i>
Number of <i>variables</i> (<i>var</i>)
Number of <i>clauses</i> (<i>cls</i>)
<i>cls/var</i>
<i>var/cls</i>
Fraction of Binary Clauses
Fraction of Ternary Clauses
Avg. Clause Size
Search Depth (from assignment stack)
Search Depth (in corresponding binary tree) ^a
Backjump Size
Learnt Clause Size
Conflict Clause Size
Fraction of assigned vars before backtracking (<i>abb</i>)
Fraction of assigned vars after backtracking (<i>aab</i>)
<i>aab.mean/abb.mean</i>
<i>abb.mean/aab.mean</i>
<i>log(WBE)</i>

관련연구 : SAT 풀이시간 추정



앞으로 할 일

- 과거적, 미래적 특성 결정
- 그 외 여러가지 특성들 중 중요한 것 결정
(feature selection 을 통해)
- 기계학습 방법 결정
 - 다양한 회귀분석 기법들(Ridge regression, LASSO, neural network ...)