SNU 4541.664A Program Analysis
Note 7

Prof. Kwangkeun Yi

요약 해석(*abstract interpretation*)

- 요약 해석: 아이디어
- 요약 해석 틀: 개관
- 요약 해석 틀: 내용
- 요약 해석 틀: Theorem들

요약해석(*abstract interpretation*)

- 프로그램 분석 = 프로그램의 요약 실행
- 분석할 프로그램의 의미(실행)의 요약본 계산

일상에서의 요약해석

$128 \times 22 + (1920 \times -10) + 4$는 어떤 수 일까요?

- 0.1초 후: "정수입니다."
- 2초 후: "짝수입니다."
- 3초 후: "$-10,0000$과 $1,0000$ 사이의 수입니다."
- 5초 후: "음수입니다."
- 1시간 후: "$-1,6380$입니다."

요약 해석(*abstract interpretation*)의 파워

프로그램 분석기 디자인의 눈을 뜨게한 가장 강력하고 간단한 틀(framework).

- "틀": 넣으면 좋은게 나온다, 재사용
- "가장 강력한": 모든 분석이 이 틀안에서 이해됨 [CC95b,CC93,CC95,Co97b].
- "간단": 틀 사용법이 간단
- "눈을 뜨게한": 어떤 분석이건 결국은 xx를 요약한 것

요약의 필요성

분석할 프로그램의 요약된 실행 = 그 소스언어의 요약된 의미 구조(*semantics*)

- 요약이 필요한 이유?
  - 요약없이 실행해 보면(simulation)서 모두를 포섭할 수 없다
  - 요약없이는 분석이 끝이 없다
- 요약은 생략이 아니다
  - 실제: $\{2, 4, 6, 8, \cdots\}$
  - "짝수"(요약) vs "4의 배수"(대충)

요약 해석으로 분석하기

는 다음을 하는 것

1. 프로그램의 실제 실행의 정의: 어떻게 무엇으로
2. 프로그램의 요약 실행의 정의: 어떻게 무엇으로
3. 올바른 요약 실행인지 확인: 어떻게 무엇으로
4. 요약 실행을 계산하는 방법: 어떻게 무엇으로

요약해석의 주요 논문[CC77,CC79,CC92a,CC92b]의 정리

요약 해석 틀

실제 실행	$[\![C]\!] = fix F \in D$
요약 실행	$[\![\hat{C}]\!] = \lim_{i \in \mathbb{N}} \hat{F}^i(\bot_{\hat{D}}) \in \hat{D}$
올바름	$[\![C]\!] \approx [\![\hat{C}]\!]$
구현	$[\![\hat{C}]\!]$ 의 자동계산

틀이 요구하는 것:

- D와 $\hat{D}$사이의 어떤 관계
- $F \in D \to D$와 $\hat{F} \in \hat{D} \to \hat{D}$의 어떤 관계

틀이 보장하는 것:

- 올바름: $[\![C]\!] \approx [\![\hat{C}]\!]$
- 구현법: $[\![\hat{C}]\!]$ 자동 계산하는 법
- 자유로움: 맘대로 이 안에서

요약 해석 디자인: step 1

프로그램의 실제 실행을 정의

- 의미공간(*semantic domain*) CPO D 를 정의
- 실제 실행은 연속 함수 $F \in D \to D$ 의 최소 고정점(*least fixed point*) $lfpF$ 으로 정의

$$lfpF = \bigsqcup_{i \in \mathbb{N}} F^i(\bot_D)$$

계획: $lfpF$를 포섭하는 요약된 물건 구하기

요약 해석 디자인: step 2

프로그램의 요약된 실행을 정의

- 요약된 의미공간(*abstract domain*) CPO $\hat{D}$을 정의
  - D와 $\hat{D}$은 갈로아 연결(*Galois connection*)
- 요약된 실행함수 $\hat{F} \in \hat{D} \to \hat{D}$ 를 정의
  - $\hat{F}$는 단조 함수(*monotonic function*)거나
  - $\hat{F}$는 팽창 함수(*extensive function*)

계획: $lfp F$를 포섭하는 요약된 물건을 $\hat{F}$가지고 구하기

요구1: 갈로아 연결(*Galois connection*)

D와 $\hat{D}$은 갈로아 연결(*Galois connection*)

$$D \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} \hat{D}$$

되어 있어야

- 갈로아 연결의 정의:

$$\forall x \in D, \hat{x} \in \hat{D} : \alpha(x) \sqsubseteq \hat{x} \iff x \sqsubseteq \gamma(\hat{x}).$$

- 갈로아 연결의 의미:
  - $\hat{D}$에서 큰 원소일수록 보다 많은 것을 의미
  - α는 실제를 요약하고(*abstraction function*)
  - γ는 요약한 원소가 뜻하는 실제를 정의(*concretization function*).

계획: 요약 분석은 $\bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot})$의 윗뚜껑(*upper bound*)을 계산하기

요구2: $\hat{F}$의 성질

- 요약된 실행함수 $\hat{F}$는 단조(*monotonic*) 함수거나:

$$\forall x, y \in \hat{D} : x \sqsubseteq y \Rightarrow \hat{F}(x) \sqsubseteq \hat{F}(y)$$

팽창(*extensive*) 함수이어야:

$$\forall x \in \hat{D} : x \sqsubseteq \hat{F}(x).$$

계획: 요약 분석은 $\bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot})$의 윗뚜껑(*upper bound*)을 계산하기

요구3: F와 $\hat{F}$의 관계

- 실제 실행함수 F와 요약된 실행함수 $\hat{F}$ 사이는

$$\alpha \circ F \sqsubseteq \hat{F} \circ \alpha, \quad \text{다시 말해,} \quad F \circ \gamma \sqsubseteq \gamma \circ \hat{F}$$

  이거나
- 실제 실행함수 F와 요약된 실행함수 $\hat{F}$ 사이는

$$\alpha(f) \sqsubseteq \hat{f} \text{ 이면 } \alpha(F\ f) \sqsubseteq \hat{F}\ \hat{f}$$

  이어야

계획: 요약 분석은 $\bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot})$의 윗뚜껑(*upper bound*)을 계산하기

결과: 안전한 요약 분석

요약 분석 $= \bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot})$ 의 윗뚜껑(*upper bound*)을 유한시간내에 계산하기

- 그러한 윗뚜껑(*upper bound*) $\hat{\mathcal{A}}$는 항상

$$\alpha(lfpF) \sqsubseteq \hat{\mathcal{A}}, \quad \text{다시 말해}$$
$$lfpF \sqsubseteq \gamma\hat{\mathcal{A}}$$

  를 만족: Theorem[fixpoint-transfer, fixpoint-transfer2]
- 즉, 분석결과 $\hat{\mathcal{A}}$가 실제실행 $lfpF$을 "포섭한다."

$\bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot})$ 의 윗뚜껑 계산법

- 요약된 의미공간(*abstract semantic domain*) $\hat{D}$의 높이가 유한하다면, 곧바로

$$\bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot})$$

  를 계산
- 요약된 의미공간(*abstract semantic domain*) $\hat{D}$의 높이가 무한하다면, 다음을 만족하는

$$\bigsqcup_{i \in \mathbb{N}} (\hat{F}^i(\hat{\bot})) \sqsubseteq \lim_{i \in \mathbb{N}} (\hat{X}_i)$$

  유한한 체인 $\{\hat{X}_i\}_i$를 계산

유한 체인 $\{\hat{X}_i\}_i$ 찾기

$$\bigsqcup_{i \in \mathbb{N}} (\hat{F}^i(\hat{\bot})) \sqsubseteq \lim_{i \in \mathbb{N}} (\hat{X}_i)$$

인 유한 체인 $\{\hat{X}_i\}_i$?

- $\hat{F}$가 단조(*monotonic*) 함수이면, $\hat{F}$에 축지법(*widening operator*) $\triangledown$를 적용한 체인:

$$\begin{array}{rcl} \hat{X}_0 & = & \hat{\bot} \\ \hat{X}_{i+1} & = & \begin{cases} \hat{X}_i & \hat{F}(\hat{X}_i) \sqsubseteq \hat{X}_i \text{ 이면} \\ \hat{X}_i \triangledown \hat{F}(\hat{X}_i) & \text{아니면} \end{cases} \end{array}$$

축지법 $\triangledown$의 조건

조건

- $\forall a, b \in \hat{D} : (a \sqsubseteq a \triangledown b) \wedge (b \sqsubseteq a \triangledown b)$
- $\forall$증가하는 체인$\{a_i\}_i$: 체인$x_0 = a_0, x_{i+1} = x_i \triangledown a_{i+1}$는 유한

이면

- $\{\hat{X}_i\}_i$ 은 유한 체인
- 그 끝($\hat{F}(\hat{X}) \sqsubseteq \hat{X}$인 $\hat{X}$ (why?))은

$$\bigsqcup_{i \in \mathbb{N}} (\hat{F}^i(\hat{\bot})) \sqsubseteq \lim_{i \in \mathbb{N}} (\hat{X}_i)$$

을 만족: Theorem[widen's safety]

축지법 결과 다듬기

$\hat{F}$가 단조(*monotonic*) 함수라면,

- 축지법을 써서 계산된 $\hat{\mathcal{A}} \stackrel{\text{let}}{=} \lim_{i \in \mathbb{N}}(\hat{X}_i)$를
- 좁히기(*narrowing operator*) $\triangle$을 써서 정교하게 다듬을 수 있다.
- 다음의 체인 $\{\hat{Y}_i\}_i$을 계산

$$\begin{aligned} \hat{Y}_0 &= \hat{\mathcal{A}} \\ \hat{Y}_{i+1} &= \hat{Y}_i \triangle \hat{F}(\hat{Y}_i) \end{aligned}$$

좁히기 $\triangle$의 조건

조건

- $\forall a, b \in \hat{D} : a \sqsupseteq b \Rightarrow a \sqsupseteq (a \triangle b) \sqsupseteq b$
- $\forall$감소하는 체인$\{a_i\}_i$: 체인 $y_0 = a_0, y_{i+1} = y_i \triangle a_{i+1}$는 유한

이면

- $\{\hat{Y}_i\}_i$은 유한 체인
- 그 끝은

$$\bigsqcup_{i \in \mathbb{N}} (\hat{F}^i(\hat{\bot})) \sqsubseteq \lim_{i \in \mathbb{N}} (\hat{Y}_i)$$

을 만족: Theorem[narrow's safety]

Fixpoint Transfer Theorems

왜 위와 같이만 하면 올바른 분석이 되는가?

Theorem (fixpoint transfer)

D와 $\hat{D}$는 각각 CPO이고 갈로아 연결이 되어있다. 함수 $F : D \to D$는 연속함수이고 $\hat{F} : \hat{D} \to \hat{D}$는 단조함수이거나 팽창함수이다. $\alpha \circ F \sqsubseteq \hat{F} \circ \alpha$ 이다. 그러면,

$$\alpha(lfp F) \sqsubseteq \bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot}).$$

Theorem (fixpoint transfer2)

CPO D와 $\hat{D}$는 갈로아 연결 $D \underset{\alpha}{\overset{\gamma}{\leftrightarrows}} \hat{D}$ 되어있다. $F : D \to D$이고 $\hat{F} : \hat{D} \to \hat{D}$ 이다. $\alpha\, f \sqsubseteq \hat{f}$ 이면 $\alpha(F\, f) \sqsubseteq \hat{F}\, \hat{f}$ 이다. 그러면,

$$\alpha(lfp F) \sqsubseteq \bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot}).$$

Widening/Narrowing Theorems

왜 위와 같이만 하면 올바른 분석이 되는가?

Theorem (widen's safety)

$\hat{D}$는 CPO 이고, $\hat{F} : \hat{D} \to \hat{D}$는 단조(monotonic) 함수이고, $\triangledown : \hat{D} \times \hat{D} \to \hat{D}$가 축지법 조건을 만족하면, 축지법으로 정의되는 체인 $\{\hat{X}_i\}_i$ 은 유한하고 그 끝은 $\lim_{i \in \mathbb{N}} \hat{X}_i \sqsupseteq \bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot})$이다.

Theorem (narrow's safety)

$\hat{D}$는 CPO 이고, $\hat{F} : \hat{D} \to \hat{D}$는 단조(monotonic) 함수 이고, $\triangle : \hat{D} \times \hat{D} \to \hat{D}$는 좁히기 조건을 만족하고 $\hat{F}(\hat{A}) \sqsubseteq \hat{A}$ 이면, 좁히기로 정의되는 체인 $\{\hat{Y}_i\}_i$ 은 유한하고 그 끝도 $\lim_{i \in \mathbb{N}} \hat{Y}_i \sqsupseteq \bigsqcup_{i \in \mathbb{N}} \hat{F}^i(\hat{\bot})$이다.