解釈可能性論理の形式化についての報告

@ TPP2025

野口 真柊

2025/12/03

神戸大学システム情報学研究科

- このスライド: https://sno2wman.github.io/slides-for-tpp2025/main.pdf:
- 形式化された Lean のコード: https://github.com/FormalizedFormalLogic/Foundation
- より技術的な内容は https://formalizedformallogic.github.io/Foundation/book/Monthly-Reports/Monthly-Report-2025___10 も参考にしてみてください.

1. Formalized Formal Logic

https://github.com/FormalizedFormalLogic/Foundation

目標: 数理論理学の諸事実を Lean で形式化する． 例えば以下が形式化済み・進行中.

- 命題論理
  - 自動証明
  - 中間論理
- 様相命題論理
  - Kripke 意味論
  - 近傍意味論
- 一階述語論理・算術・集合論
  - 完全性定理
  - カット除去定理
  - Gödel の不完全性定理
  - ZFC の無矛盾性
- 証明可能性論理
  - Solovay の算術的完全性定理
  - 分類定理
- 解釈可能性論理 (今回の報告)

2. 解釈可能性論理と今回の成果

2 つの数学的理論（公理系）の間に様々な関係がある.

例: 理論の間の関係の例

- Peano 算術 PA は帰納法を弱めた算術 $\mathrm{I}\Sigma_1$ の真の拡大である.
- 標準的な公理的集合論 ZFC は PA を内部で実行できる.
  - 言語がそもそも異なる $\mathcal{L}_{\mathrm{Set}} := \{\in, =\}, \mathcal{L}_{\mathrm{Arith}} := \{0, 1, +, \times, =\}$ が適切に翻訳出来る.
- Lean は ZFC や PA を内部で実行できる.
- 真のクラスが扱える集合論 NBG は ZFC の保守的拡大である.
  i.e. NBG で証明可能な集合に関する事実は ZFC でも証明可能.

この現象を 2 項関係 $\rhd$ を用いて ZFC $\rhd$ PA などと表すことにする. 文脈によって $\rhd$ が何を意図しているかは異なる. 上の例では **ZFC は PA を解釈可能**とする.

1 項様相 $\Box$ と 2 項様相 $\rhd$ を扱う様相論理を解釈可能性論理と呼ぶ.

- $\mathbf{GL} :=$ K $+$ L $+$ (Nec)
- $\mathbf{IL}^{-} := \mathbf{GL}$ $+$ J3 $+$ J6 $+$ (R1) $+$ (R2)
- $\mathbf{IL}^{-}$(J2) $:= \mathbf{IL}^{-}$ $+$ J2
- $\mathbf{IL} := \mathbf{GL}$ $+$ J1 $+$ J2 $+$ J3 $+$ J4 $+$ J5

命題 1: $\mathbf{IL} = \mathbf{IL}^{-}$(J1, J2, J5)

- K : $\Box(p \to q) \to (\Box p \to \Box q)$
- L : $\Box(\Box p \to p) \to \Box p$
- J1 : $\Box(p \to q) \to (p \rhd q)$
- J2 : $(p \rhd q) \to (q \rhd r) \to (p \rhd r)$
- J2^{+} : $(p \rhd (q \lor r)) \to (q \rhd r) \to (p \rhd r)$
- J3 : $(p \rhd q) \to (q \rhd r) \to ((p \lor q) \rhd r)$
- J4 : $(p \rhd q) \to (\Diamond p \to \Diamond q)$
- J4^{+} : $\Box(p \to q) \to (r \rhd p \to r \rhd q)$
- J5 : $\Diamond p \to p$
- J6 : $\Box p \leftrightarrow (\neg p \rhd \bot)$
- (Nec) : $\varphi \mid \Box\varphi$
- (R1) : $\varphi \to \psi \mid \chi \rhd \varphi \to \chi \rhd \psi$
- (R2) : $\varphi \to \psi \mid \psi \rhd \chi \to \varphi \rhd \chi$

GL の Kripke フレーム[1] $\langle W, R\rangle$ に W の元で添字付けられた関係の族を加えたフレーム $\langle W, R, \{S_w\}_{w\in W}\rangle$ を Veltman フレーム[2]と呼ぶ． (b) は満たすとする．

モデルは通常通り付値関数 $V : W \times \mathrm{PropVar} \to \mathbf{2}$ を加える．

$\rhd$ の意味は次で与える．

$$x \vDash A \rhd B :\Longleftrightarrow \forall y.[xRy, y \vDash A \Rightarrow \exists z.[yS_x z \,\&\, z \vDash B]]$$

命題 2: 任意のフレーム F で J3, J6, (R1), (R2) は成り立つ．

命題 3:
- $F \vDash$ (j1) $\Longleftrightarrow F \vDash$ J1
- $F \vDash$ (j2) $\Longleftrightarrow F \vDash$ J2 $\Longleftrightarrow F \vDash$ J2$^+$
- $F \vDash$ (j4) $\Longleftrightarrow F \vDash$ J4 $\Longleftrightarrow F \vDash$ J4$^+$
- $F \vDash$ (j5) $\Longleftrightarrow F \vDash$ J5

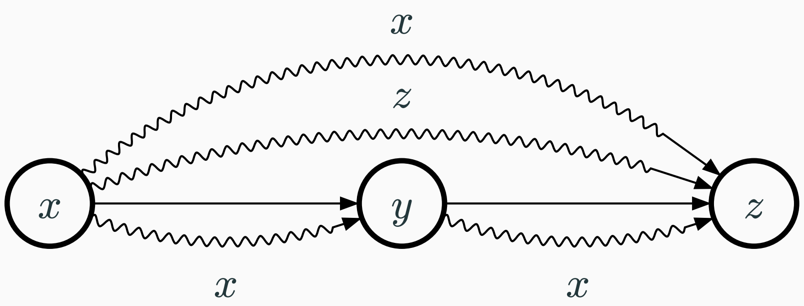

Figure 1: Veltman モデルの例

- (b) : $xS_w y \Rightarrow wRx$
- (j1) : $wRx \Rightarrow xS_w x$
- (j2) : $xS_w y, yS_w z \Rightarrow xS_w z$
- (j4) : $xS_w y \Rightarrow wRy$
- (j5) : $wRx, xRy \Rightarrow xS_w y$

[1] R は推移的かつ逆整礎的: $x_0 R x_1 R x_2 R x_3 \cdots$ は存在しない．

[2] 正確には右の性質を全て満たすものを Veltman フレームと呼び，これはIL^--フレームや Veltman prestructure と呼ぶ．

2.4 Veltman フレームの形式化

```
structure Modal.Kripke.Frame where
  World : Type
  Rel : HRel World
  [world_nonempty : Nonempty World]

structure InterpretabilityLogic.Veltman.Frame extends toKripkeFrame : Modal.Kripke.Frame where
  [isGL : toKripkeFrame.IsGL]
  S : (w : World) → HRel World
  S_cond {w x y} : S w x y → w ≺ x

abbrev SRel' {F : Veltman.Frame} (w : outParam F.World) (x y : F.World) := F.S w x y
notation:45 x:max " ≺[" w "] " y:max => SRel' w x y

class HasAxiomJ1 (F : Frame) : Prop where
  S_J1 : ∀ {w x : F.World}, w ≺ x → x ≺[w] x

class HasAxiomJ2 (F : Frame) extends F.HasAxiomJ4 where
  S_J2 : ∀ {w x y z : F.World}, x ≺[w] y → y ≺[w] z → x ≺[w] z

class HasAxiomJ4 (F : Frame) : Prop where
  S_J4 : ∀ {w x y : F.World}, x ≺[w] y → w ≺ y

class HasAxiomJ5 (F : Frame) : Prop where
  S_J5 : ∀ {w x y : F.World}, w ≺ x → x ≺ y → x ≺[w] y
```

Listing 1: Veltman フレームとその上の条件の Lean 上の形式化

命題 1，命題 2，命題 3 を Lean で形式化した．以下の IL の部分論理 $\mathbf{IL}^-(X)$ は然るべきフレームクラスに対して健全であり，反例モデルを作って以下の包含関係が成り立つことを検証した[1]．

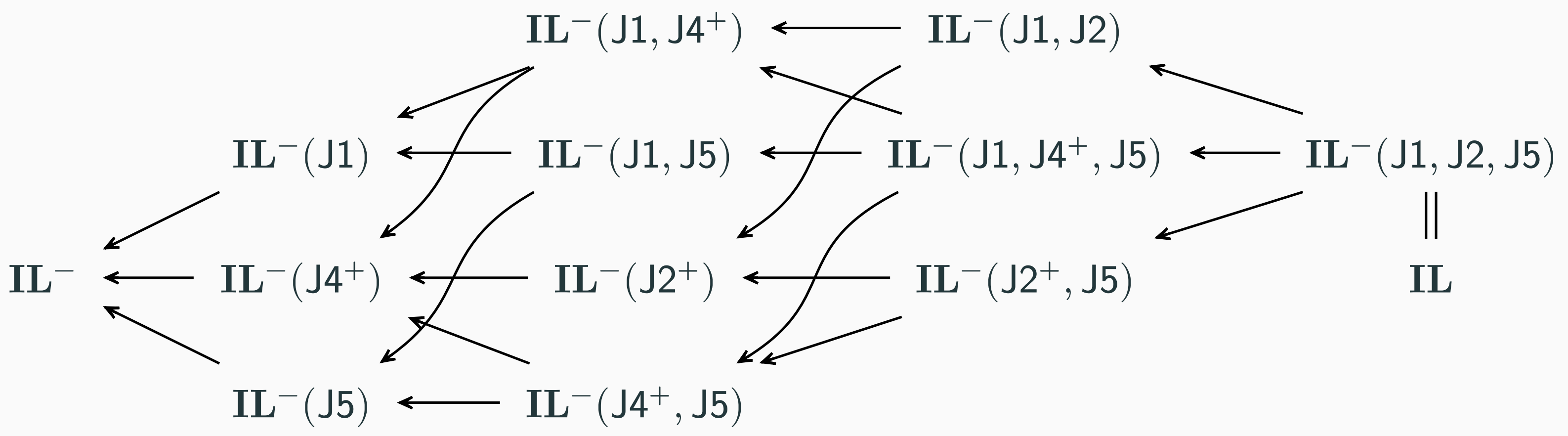

Figure 2: Interpretability Logic Zoo (1)

[1] 例えば $\mathbf{IL}^-(\mathrm{J4})$ と $\mathbf{IL}^-(\mathrm{J4}^+)$ は Veltman 意味論では分離できない．その意味で $\mathbf{IL}^-(\mathrm{J4})$ は Veltman 意味論に対して完全ではない．cf: T. Kurahashi and Y. Okawa [1]

2.6 IL の部分論理についての分離(2)

反例モデルの構成は自然数に順序を入れる．この検証は `omega` や `grind` など自動化タクティクが非常に有効.

```
use {
  toKripkeFrame := ⟨Fin 4, (· < ·)⟩
  isGL := Modal.Kripke.Frame.isGL_of_isFiniteGL {
    trans := by omega;
    irrefl := by omega;
  }
  S w x y := (w < x ∧ x = y) ∨ (w < x ∧ x < y ∧ ¬(w = 0 ∧ x = 1 ∧ y = 3)),
  S_cond := by tauto;
}
constructor;
. constructor;
  exact {
    S_J1 := by tauto;
    S_J4 := by grind;
  }
. by_contra hC;
  have := Veltman.Frame.HasAxiomJ2.of_validate_axiomJ2 hC ▷.S_J2 (w := 0) (x := 1) (y := 2) (z := 3) (by tauto) (by tauto);
  contradiction;
```

Listing 2: $\mathbf{IL}^{-}(\mathrm{J1}, \mathrm{J4}^{+}) \subsetneq \mathbf{IL}^{-}(\mathrm{J1}, \mathrm{J2})$ となる反例モデルのコードの抜粋

2.7 IL の拡大論理について

IL の拡張論理についても同様に Lean で検証した.
破線は Veltman 意味論では分離できないことを示す.

- M : $(p \rhd q) \to (p \land \Box r) \rhd (q \land \Box r)$
- M_0 : $(p \rhd q) \to (\Diamond p \land \Box r) \rhd (q \land \Box r)$
- P : $(p \rhd \Diamond q) \to \Box(p \rhd q)$
- P_0 : $(p \rhd \Diamond q) \to \Box(p \rhd q)$
- W : $(p \rhd q) \to (p \rhd (q \land \Box\neg p))$
- $\mathsf{KW1}^0$: $((q \land p) \rhd \Diamond q) \to (q \rhd (q \land \Box\neg p))$
- KW2 : $(p \rhd \Diamond q) \to (q \rhd (q \land \Box\neg p))$
- W* : $(p \rhd q) \to ((q \land \Box r) \rhd (q \land \Box r \land \Box\neg p))$
- R : $(p \rhd q) \to \neg(p \rhd \neg r) \rhd (q \land \Box r)$
- R* : $(p \rhd q) \to \neg(p \rhd \neg r) \rhd (q \land \Box r \land \Box\neg p)$
- F : $(p \rhd \Diamond p) \to \Box\neg p$

詳細は割愛する.

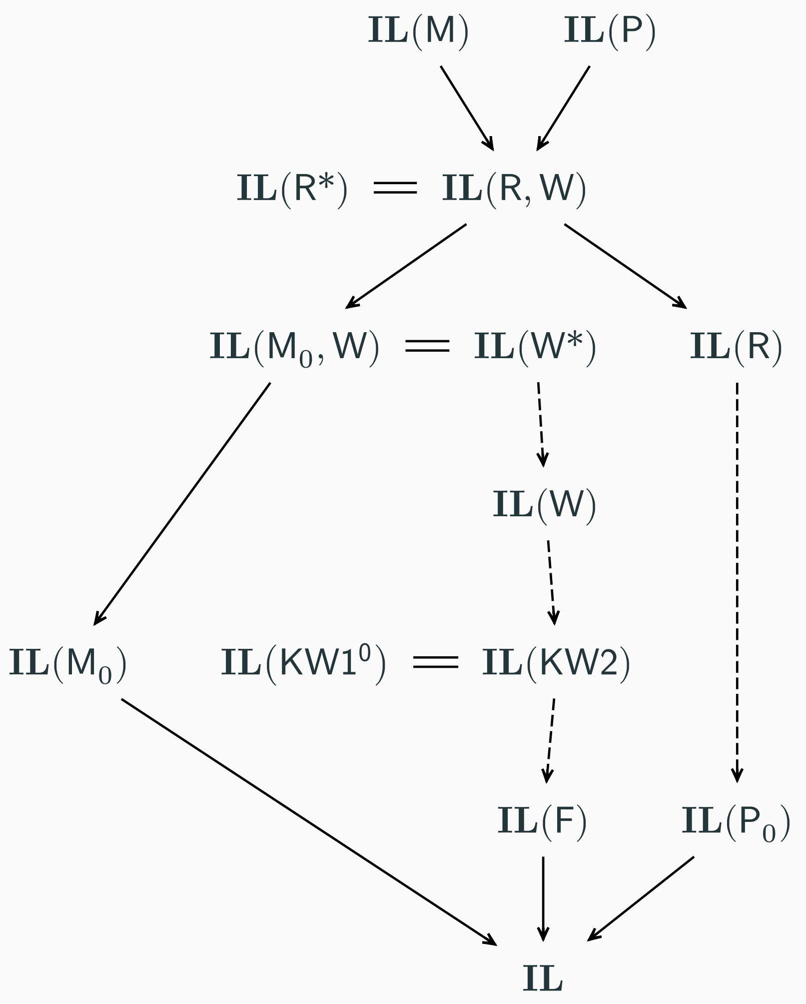

Figure 3: Interpretability Logic Zoo (2)

3. 展望

3.1 展望: 完全性について

```
class Sound (𝓢 : S) (𝓜 : M) : Prop where
  sound : ∀ {φ : F}, 𝓢 ⊢ φ → 𝓜 ⊧ φ

class Complete (𝓢 : S) (𝓜 : M) : Prop where
  complete : ∀ {φ : F}, 𝓜 ⊧ φ → 𝓢 ⊢ φ

protected class Frame.IsIL (F : Frame) extends F.IsILMinus_J1_J2_J5
protected alias FrameClass.IL := FrameClass.ILMinus_J1_J2_J5

instance : Sound InterpretabilityLogic.IL FrameClass.IL := by ... -- done

instance : Entailment.Consistent InterpretabilityLogic.IL := by ... -- done

instance : Complete InterpretabilityLogic.IL FrameClass.IL := by sorry
```

完全性に関しては何も形式化できていない．

定理 4: $\forall F. F \vDash \mathbf{IL} \Rightarrow F \vDash \varphi$ なら $\mathbf{IL} \vdash \varphi$．同様のことが $\mathbf{IL}^{-}$ などについても成り立つ．

カノニカルモデルに似た振る舞いをするモデルを作るが，S_w の定義はかなり技巧的で形式化が大変．しかもそれぞれの論理に対して別個に構成する必要があり，量もかさばる．

いくつかの公理は Veltman 意味論では分離できない.

例: $F \vDash (\mathrm{j2}) \iff F \vDash \mathsf{J2} \iff F \vDash \mathsf{J2}^{+}$

例: $F \vDash \mathbf{IL}$ のとき次は同値.

1. 公理 W: $F \vDash p \rhd q \to p \to (q \land \Box\neg p)$
2. 公理 F: $F \vDash p \rhd \Diamond p \to \Box\neg p$
3. 任意の $w \in W$ に対し, R と S_w の合成 $(R; S_w)$ は逆整礎的.

```
abbrev RS {F : Frame} (w : F.World) := Relation.Comp (· ≺ ·) (· ≺[w] ·)
notation:50 x:max " ≺≺[" w "] " y:max ⇒ RS w x y

class HasAxiomW (F : Frame) : Prop where
  S_W : ∀ w : F.World, ConverseWellFounded $ (· ≺≺[w] ·)

lemma TFAE_HasAxiomW [F.IsIL] : [
  F.HasAxiomW,
  F ⊧ Axioms.W (.atom 0) (.atom 1),
  F ⊧ Axioms.F (.atom 0)
].TFAE := by ...
```

Veltman 意味論の近傍意味論的な拡張である Verbrugge 意味論では上の例を分離できる．技術的取り扱いは煩雑.

J. M. Rovira [2] は Agda で Verbrugge 意味論を形式化し，いくつかのフレーム条件などを形式化しているが，完全性などは未完．この結果を拡張する.

3.3 展望: 算術的完全性

証明可能性述語 $\mathrm{Pr}_T(x)$ の挙動は様相演算子 $\Box$ を通じて様相論理 **GL** と対応する.

解釈可能性述語 $\mathrm{Int}_T(x, y)$ の挙動は様相演算子 $\rhd$ を通じて **IL** の拡張論理に対応する.

事実 5:
- T が **PA** の拡大理論ならその証明可能性論理は $\mathbf{IL}(\mathrm{M})$ である.
- T が有限公理化可能ならその解釈可能性論理は $\mathbf{IL}(\mathrm{P})$ である.

解釈そのものの形式化が現在出来ていない. 例えば, $\mathbf{ZFC} \rhd \mathbf{PA}$ といったことも形式化出来ていない.

様相論理的には, 逆に Veltman 意味論を単純化して技術的に扱いやすくした Visser 意味論などの下準備も必要.

やることは多い.

4. 参考文献

[1] T. Kurahashi and Y. Okawa, "Modal Completeness of Sublogics of the Interpretability Logic IL," Mathematical Logic Quarterly, vol. 67, no. 2, pp. 164–185, May 2021, doi: 10.1002/malq.202000037.

[2] J. M. Rovira, "Interpretability Logics and Generalized Veltman Semantics in Agda," Master's Thesis, Universitat de Barcelona, 2020.