BDI Logic II

Yudai Kubono

February 11, 2024

Graduate School of Science and Technology, Shizuoka University

Previous Presentation

- BDI theory claims that **belief, desire, and intention** are necessary to understand or explain practical reasoning.
  - Belief: we can buy a soda in the vending machine.
  - Desire: I want to drink something.
  - Plan: buying a soda in the vending machine.
  - Intention: I intend to fulfill the plan.
- Intention in Bratman's philosophy: a mental attitude characterized by some properties such as future-oriented, pro-attitude, and commitment ("I am planning to do it", "I will do it").

Intuition to The Notions

- Belief:「私の記憶では論文を出した」
- Desire:「今月中に論文を出したい」
- Intention:「(教授の前で) 今月中に論文を出します」

Rao & Georgeff's Formalization

- BDI Logic describes the mental attitudes belief, desire, and intention to analyze the theory.
- Rao & Georgeff's BDI logic is based on CTL^*, which is one of temporal logics.
- They proposed a logic for an agent with three mental attitudes.

Syntax

Let $\mathcal{P}$ be a countable set of atomic propositions. The language $\mathcal{L}_{\mathcal{P}}$ is the union of the two set of formulas $\mathcal{L}^S_{\mathcal{P}}$ and $\mathcal{L}^P_{\mathcal{P}}$ generated by the following grammar:

$$\mathcal{L}^S_{\mathcal{P}} \ni \varphi ::= p \mid \top \mid \neg\varphi \mid \varphi \wedge \varphi \mid \mathrm{A}\psi \mid \triangle\varphi,$$
$$\mathcal{L}^P_{\mathcal{P}} \ni \psi ::= \varphi \mid \neg\psi \mid \psi \wedge \psi \mid \mathrm{X}\psi \mid \psi \mathcal{U} \psi,$$

where $p \in \mathcal{P}$ and $\triangle \in \{\mathrm{BEL}, \mathrm{DES}, \mathrm{IND}\}$. Other temporal operators E, F, and G are defined by $\mathrm{E}\varphi := \neg\mathrm{A}\neg\varphi$, $\mathrm{F}\psi := \top\mathcal{U}\psi$, and $\mathrm{G}\psi := \neg\mathrm{F}\neg\psi$, respectively. Other logical connectives $\vee$, $\rightarrow$, and $\leftrightarrow$ are defined in the usual manner.

Semantics

Definition 1

A Kripke structure M is defined to be a tuple $\langle W, \{T_w, R_w\}_{w \in W}, V, B, D, I \rangle$, where:

- W *is a non-empty set of possible worlds;*
- $\{T_w\}_{w \in W}$ *is a non-empty set of time points in* w*;*
- $\{R_w\}_{w \in W}$ *is a serial binary relation on* T_w*;*
- $V(w, t_i)$ *is a subset of* $\mathcal{P}$ *for each* $w \in W$ *and* $t_i \in T_w$*;*
- B *is a serial, transitive, and Euclidean binary relation on* W*;*
- D, I *are serial binary relations on* W*.*

Definition 2

A path π_w in w is defined as a infinite sequence $(t_0, \ldots t_n)$ of time points such that for all $i \in \mathbb{N}$, $(t_i, t_{i+1}) \in R_w$. A sub-path in w that starts from t_i is denoted by $\pi_w[t_i]$. The initial time point in a path π_w is denoted as $\pi_w(0)$.

Satisfaction Relation

Definition 3-1 (State formula)

For each Kripke structure M, possible worlds $w \in W$, and $t_i \in T_w$, a satisfaction relation $\vDash$ is given as follows:

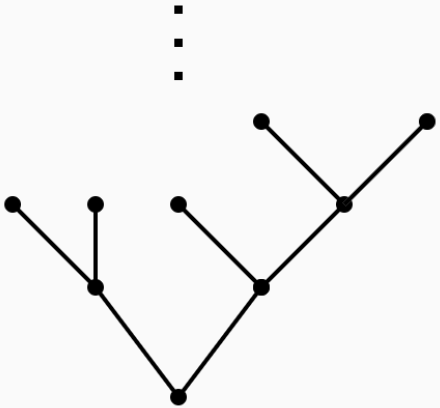
$$
\begin{aligned}
M, w, t_i &\vDash p \text{ iff } p \in V(w, t_i); \\
M, w, t_i &\vDash \top \text{ always holds;} \\
M, w, t_i &\vDash \neg\varphi \text{ iff } M, w, t_i \nvDash \varphi; \\
M, w, t_i &\vDash \varphi \wedge \chi \text{ iff } M, w, t_i \vDash \varphi, \text{ and } M, w, t_i \vDash \chi; \\
M, w, t_i &\vDash \mathrm{A}\psi \text{ iff } M, \pi_w[t_i] \vDash \psi \text{ for all } \pi_w[t_i]; \\
M, w, t_i &\vDash \triangle\varphi \text{ iff } M, v, t_i \vDash \varphi \text{ for all } v \text{ such that } (w, t_i, v) \in R(\triangle);
\end{aligned}
$$

where $\triangle \in \{\mathrm{BEL}, \mathrm{DES}, \mathrm{IND}\}$ and $R(\triangle) \in \{B, D, I\}$.

Definition 3-2 (Path formula)

For each Kripke structure M, possible world $w \in W$, and path π_w, a satisfaction relation $\vDash$ is given as follows:

$$
\begin{aligned}
M, \pi_w \vDash \varphi \ \textit{iff} \ & M, w, \pi_w(0) \vDash \varphi; \\
M, \pi_w \vDash \mathrm{X}\psi \ \textit{iff} \ & M, \pi_w[t_1] \vDash \psi; \\
M, \pi_w \vDash \psi \mathcal{U} \chi \ \textit{iff} \ & \text{there exists } n \text{ such that } M, \pi_w[t_n] \vDash \chi \text{ and} \\
& \text{for all } 0 \leq k \leq n, M, \pi_w[t_k] \vDash \psi.
\end{aligned}
$$



Describing Bratman's Analysis

- $\mathrm{IND}\varphi \rightarrow \mathrm{DES}\varphi$.
- $\neg\mathrm{BEL}(\mathrm{E}(\mathrm{F}\varphi)) \rightarrow \neg\mathrm{IND}\varphi$: an agent does not intend to do what she cannot believe to achieve.
- $\mathrm{IND}(\mathrm{A}(\mathrm{F}\varphi)) \rightarrow \mathrm{A}((\mathrm{IND}(\mathrm{A}(\mathrm{F}\varphi)))\mathcal{U}(\mathrm{BEL}\varphi \vee \neg\mathrm{BEL}(\mathrm{E}(\mathrm{F}\varphi))))$: if an agent intends to achieve it sometime in all the futures, then this intention continues until she believes that it holds or she cannot believe that there is a future that it holds anymore.

Other Properties

- $\mathrm{DES}\varphi \to \mathrm{BEL}\varphi$: an agent only desires to do what she believes.
- $\mathrm{BEL}\varphi \to \mathrm{DES}\varphi$: an agent only believes in what she desires to do.
- $\mathrm{IND}(\mathrm{A}(\mathrm{F}\varphi)) \to \mathrm{A}((\mathrm{IND}(\mathrm{A}(\mathrm{F}\varphi)))\mathcal{U}(\mathrm{BEL}\varphi))$: once an agent intends to achieve it, she never gives up.

BDI Architecture

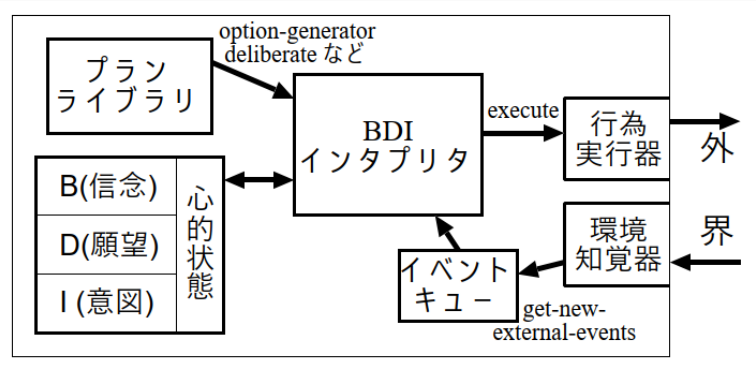


Figure 1: BDI Architecture quoted from (Nide (2014))

References

- Rao, A. S. and Georgeff, M. P. (1998). Decision Procedures for BDI Logics. *Journal of Logic and Computation*, 293-343.
- Meyer, J.-J. Ch., Broersen, J, and Herzig, A. (2015). BDI Logics. *Handbook of epistemic logic*, edited van Ditmarsch, H., van Der Hoek, W., Halpern, J. Y., and Kooi, B. College Publications.
- Nide, N. (2014). BDI -モデル、アーキテクチャ、論理—. 全脳アーキテクチャ勉強会.