BDI
—モデル、論理、アーキテクチャ、そしてエージェント—

新出尚之(奈良女子大学)

2020年10月13日
VCPCオンラインセミナー

はじめに

- 「自律エージェント」とは、人間と同様に、自らの目的を持ち、その達成のために自らの行動を決定して動作するシステムである
- 自律エージェントの構築のためのモデルとして「BDIモデル」がある
- BDIモデルの特徴
  - ⋆ 人間の行為決定過程を基にしている
  - ⋆ 論理モデルを持ち、形式的記述が可能
- 本講演ではBDIモデルの紹介を行う

概　要

- 自律エージェント
- 意図の理論
- BDI
  - ⋆ BDIモデル
  - ⋆ BDIアーキテクチャ
  - ⋆ BDI論理(BDI logic)

自律エージェント

人間の行為決定

例 出張のため、飛行機に乗るべく空港バスを待っていたが、遅れそうなのでタクシーに切り替え。出発ターミナルが何番か知らなかったので、タクシー内から航空券記載の案内電話に電話して確認。

チェックインには長蛇の列。しかし、預け手荷物がなければエクスプレスチェックインが使えることを思い出し、そちらへ。セキュリティゲートを通ろうとすると引っかかる。携帯電話をポケットから取り出してOK。

出発ラウンジでカフェを探してサンドイッチで朝食。食べながらメールをチェック。出発時間のコールが聞こえたので搭乗口へ。(Bordini et al., *Programming multi-agent systems in AgentSpeak using Jason*, Wiley, 2007. 一部改変)

人間の行為決定(continued)

- 何らかの目的を持ち、その達成のための行動をとる
  - ⋆ 達成方法の案(プラン)を持っている
  - ⋆ プランは部分的にしか具体化されていないが、その場で具体化しつつ実行できる
- 何かうまくいかなければ回復のため、他のプラン(または目的)を選んで実行を試みる
  - ⋆ すぐに諦めたり家に戻ったりはしない
- 実行中に追加の情報が必要なことがある
  - ⋆ それを知るため追加の行為(アクション)を実行する
- それぞれの目標を持った複数の活動を並立させる

コンピュータの“行為”決定

- コンピュータシステムに、人間のような行為決定を行わせたい
  - ⋆ 「目的」レベルでの指示を与える
    - ∗ 手順を詳しく書いた「プログラム」をそのまま実行するのとは異なる
  - ⋆ 予期しない事態に対応し、失敗から回復
  - ⋆ 他者と知識レベルのコミュニケーションを行う
- そのような性質を持つシステムを作るにはどうすれば?
  - ⋆ 人間のやり方を模倣してみるのが有力な手

自律エージェントとは

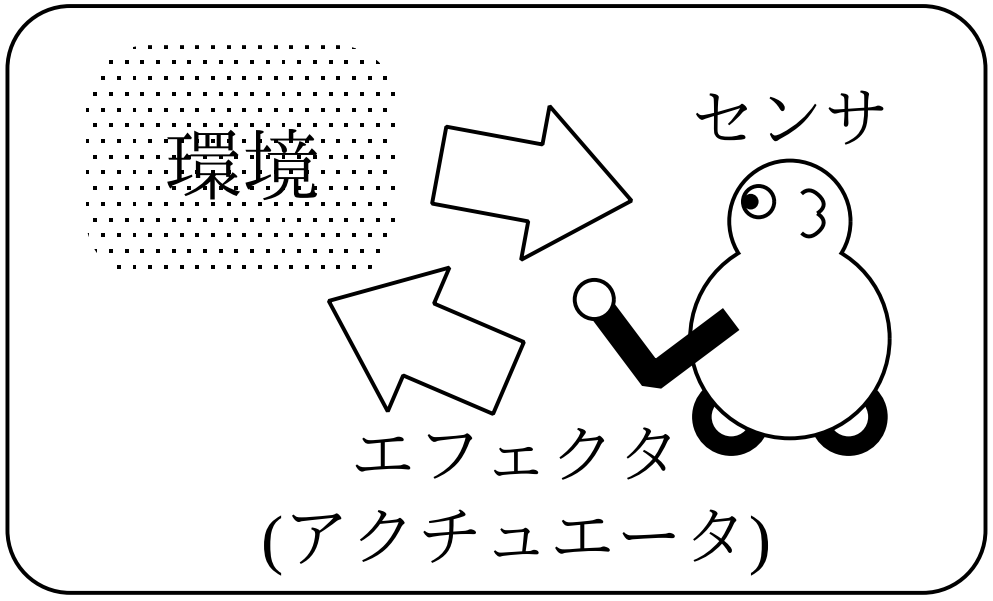
- 反応型(reactive)のシステム
  - ⋆ 動き続け、周辺環境と相互作用し続ける
    - ∗ 単純に「入力を得て出力を返して終わり」というプログラムはこれに当てはまらない
- 利用者がそのシステムに何らかの仕事を委託し、そのシステムが達成方法を決定する
  - ⋆ 利用者は達成すべき目的だけを与え、詳細に実現方法を指示することはしない
    - ∗ OS、銀行のATMシステム、Webサーバ、囲碁プログラム、などはこの性質を満たさない

エージェントの特徴

- 環境の中の存在である
- 自律性
- 目的志向性
- 反応性
- 社会性

エージェントの特徴(continued)

- 環境の中の存在である
  - ⋆ 環境と相互作用する
    - ∗ 環境から情報を知覚: センサ
    - ∗ 環境への働きかけ(行為): エフェクタ(アクチュエータとも)
    - ∗ 環境は実環境のこともあり、ソフトウェア内の世界のこともある



エージェントの特徴(continued)

- 環境からの情報(知覚)から、いかにして環境への働きかけ(行為)を決めるか
  - プラン(計画)を用いて決める
  - プランは既成の場合も、その場で自分で作る場合もある
- 環境の知覚も、環境への働きかけも不完全(成功するとは限らない)

エージェントの特徴

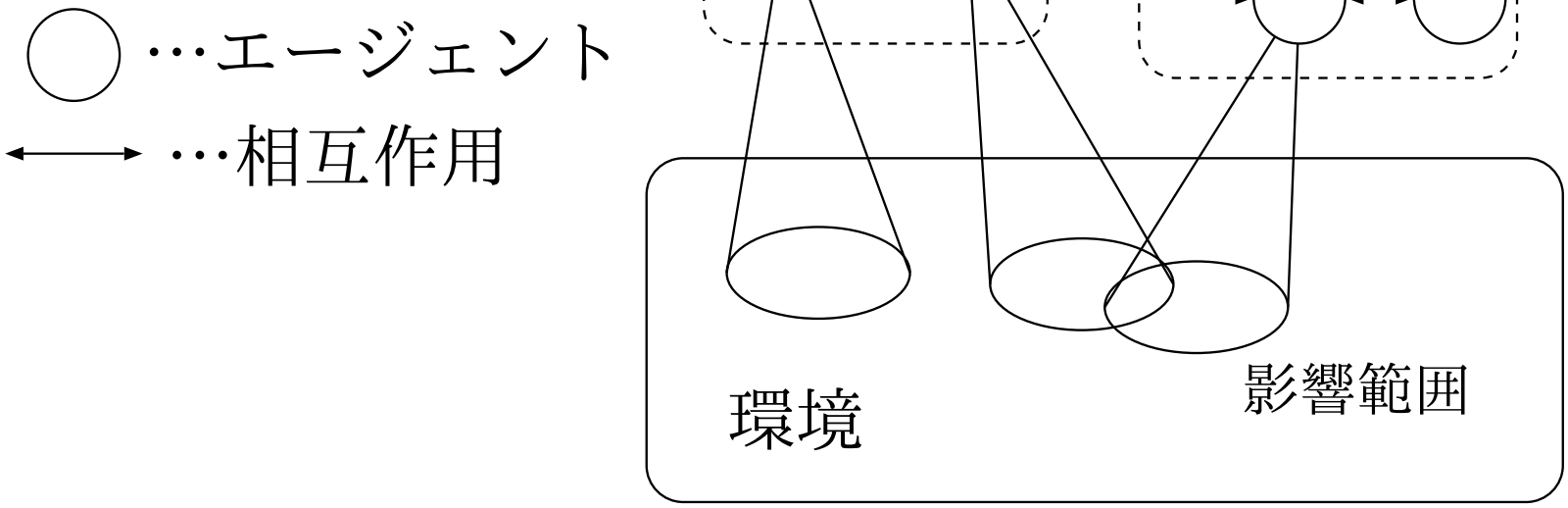
- 自律性
  - ⋆ 完全な「自律システム」は(例えば)人間
  - ⋆ 自ら目的を持ち、目的を達成する方法を選んで行為する
    - ∗ コンピュータシステムの場合「自ら目的を持つ」は難しい場合があるので、「利用者が目的を与え、システムが達成方法を選ぶ」と考える
- 目的志向性
  - ⋆ 「これこれをせよ」(手続き)ではなく「何かを達成せよ」(目的)

エージェントの特徴(continued)

- 反応性
  - ⋆ 環境からの情報に反応して目標や行動を変えたりする
- 社会性
  - ⋆ 他のエージェントと必要に応じて連携·協同する
  - ⋆ 情報(信念、目標、プランなど)を知識としてやりとりする

マルチエージェントシステム

- 単独のエージェントシステムは少ない
- 通常、環境内には他のエージェントが存在する
  - ⋆ 各エージェントにとって、影響を及ぼす範囲がある
  - ⋆ エージェント間で相互作用しあう関係や、組織的な関係もある



エージェントシステムの構築における要請

- 信念、目標、計画(プラン)といった概念を持つこと
- 目的を与えるという形で仕事が委託できること
- 目的指向での問題解決ができること
- 環境に反応して振る舞いを変える機能を持つこと
  - ⋆ 目的指向での問題解決との見通しよい統合も必要
- 知識レベルでのエージェント間のやりとりや協同ができること

特に、エージェントシステムを構築するための手段は、これらの機能を記述できることが求められる

エージェントシステムの構築に向けて

- BDIモデル
  - ★ 自律エージェントのモデル
  - ★ 信念·願望·意図の3つの心的状態を概念として持つ
  - ★ 行為選択の説明に「意図」の概念を使用
    - ＊ 目的達成のための意図を形成して持続し、その意図に基づいて行為を選ぶ
  - ★ 「**意図の理論**」がベース
    - ＊ 人間の目的達成に向けた行為決定機構の模倣

意図の理論

意図の理論

- 哲学者Bratmanが提唱(1987)
- 人間の目的達成に向けての行為選択の説明
- **信念**(belief)・**願望**(desire)・**意図**(intention)の3つの心的状態・・・BDI
- 「**意図**」の概念が鍵
  - ★ 特定の計画のもとに行動しようという心の働き
  - ★ 信念・願望には還元できない
  - ★ **持続性**を持つ(行為の**一貫性**を生む働き)
  - ★ 「未来指向的」意図(この計画でいこう)と「現在指向的」意図(今がその時だ)
- 自律エージェントが備えるべき要件の分析と見ることもできる

意図の(言葉としての)定義

- (1)考えていること。おもわく。つもり。(2)行おうとめざしていること。また、その目的。(広辞苑による)
- 目的を達成するために立案した大きな計画の部分であり、未来または現在、行おうと目指している心的状態
- 意図の理論
  - ⋆ 我々人間は、目的を持って生きている。そして、その目的を達成するために大きな計画を立案する。そして、それら計画から選択して「行おうと目指していること」を意図という心的状態として形成する。

(p. 18, p. 20 は近畿大学「エージェント工学」の授業資料(高田) より、一部改変)

3つの心的状態

信念 エージェントが環境に関して持つ情報。正確・完全とは限らない

- 知覚、あるいは他エージェントとのコミュニケーションなどから得られる
- プランも広義で(目的達成手段に関する)信念の一部

願望 エージェントが到達できたらよいなと思う状況。これだけではエージェントの行為には直結しない

意図 エージェントが目的達成に向かって特定の手段(計画)で取り組もうと決めた状況、ないしはその計画

行為者としての人間

- 我々人間は計画立案する社会的行為者
- 計画立案の問題点
  - ⋆ 信念の不完全性
  - ⋆ 推論能力の有限性
  - ⋆ 個人的および社会的調整の必要性
  - ⋆ 予期せぬ事態が起きうる (詳細な計画は無駄)
- 部分的に計画立案し複雑な目的を達成

実践的推論(practical reasoning)

- 目的達成に向けた行為を決めるための推論
- 以下の2つからなる
  - ⋆ 熟考(deliberation)
    - ∗ まず、どのような目的を達成すべきか決定
  - ⋆ 手段–目的推論(means-ends reasoning)
    - ∗ 次いで、その目的をどのような手段で達成するかを決定
- **意図**とは、選んだ手段について「これで達成しよう」と決めた心的状態で、これが行為の原動力となる

実践的推論による行為選択の過程

- 信念と願望をもとに、達成すべき目的を(複数のうちから) 選ぶ(**熟考**)
- その目的を達成するための手段(計画、プラン)を決定(**手段–目的推論**)*

  *計画を複数から選ぶ過程も熟考に含める文献もある

- 手段として選定したプランに対し、「そのプランを実行する」という「**意図**」を形成·保持
- その意図に沿って**行為**
  - ⋆ その意図は(基本的に)達成まで持続するが、状況に応じては捨てられたり他の意図を選び直したりもする

行為選択の過程(例)

- 喉が渇いた→渇きを癒したい
  - ⋆ 達成すべき目的: 渇きを癒す
  - ⋆ その手段: プラン「ソーダを買って飲む」「紅茶を淹れて飲む」など
    →ここでは「ソーダを買って飲む」を選んだ

- プラン「ソーダを買って飲む」を実行するという意図を形成·保持し、それに沿って行為

(Singh et al., *Formal Methods in DAI: Logic-Based Representation and Reasoning*, 1999. 一部改変)

プラン

プランの構造

- (ラベル ··· プランの名前)
- トリガ·イベント ··· 目標の発生など
- 前提条件 ··· そのプランが選ばれる条件
- **本体** ··· 副目標あるいは基本行為の列
  - ⋆ 副目標 ··· さらに他のプランで再帰的に達成する目標
  - ⋆ 基本行為 ··· それ以上分解できない、直接実行可能な行為

プランの**実行**: 本体を順に達成

- 副目標 ··· その目標を生成し、手段を選んで達成
- 基本行為 ··· 実行

プラン(continued)

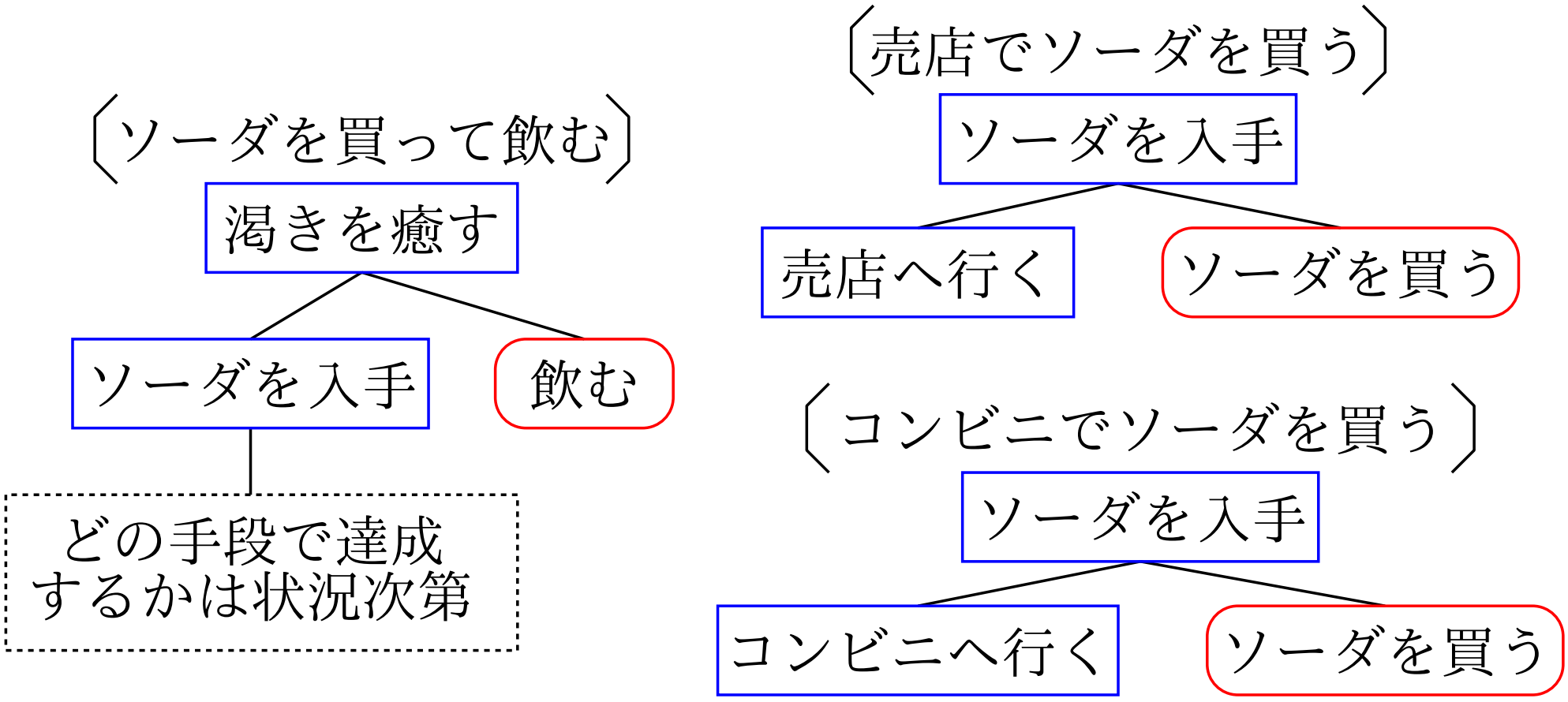
「ソーダを買って飲む」プランの場合

- ラベル: “ソーダを買って飲む”
- トリガ·イベント: 渇きを癒す目標の発生
- 前提条件: ソーダを買うお金がある、など
- **本体**:
  - ⋆ ソーダを入手 ···副目標
  - ⋆ 飲む ···基本行為

その**実行**

- 副目標「ソーダを入手」を達成
- 基本行為「飲む」を実行

プラン(continued)



〈 〉=ラベル　　=(副)目標　　=基本行為

プランの性質

- 部分性
  - ⋆ 最初から目標達成までの全行程を決める必要はない
  - ⋆ 部分的な計画を前もって定め、必要に直面すれば細部を決定する
  - ⋆ ソーダの例
    - ∗ 最初は「ソーダを入手」→「飲む」だけ決定
    - ∗ 「ソーダを入手」する具体的な手段はその時になって決める
  - ⋆ 旅行の例
    - ∗ 最初は「目的地に行く」→「観光する」→「帰る」だけ決定
    - ∗ 観光の具体的な内容は別途決める。あるいは現地でその時に決める

プランの性質(continued)

- 階層性
  - ⋆ 他の副目標を考慮することなく、ある副目標のみ考慮の対象として細部を決定する
- 慣性
  - ⋆ 実行中は再考慮への抵抗が起こる
- 整合性(要請)
  - ⋆ 計画は動的環境でうまく遂行できることが必要
  - ⋆ 信念との整合性があること

未来指向的意図と現在指向的意図

- 達成すべき目標と手段が決まっても、達成すべき時は今でないかもしれない . . . 未来指向的意図
  (例: 帰省するので、12/30は新幹線で広島に行こう)
- 達成すべき時が来れば実行に移す . . . 現在指向的意図
  (例: 12/30が来たから、これから新幹線で広島に行こう)
- 複数の(未来指向的)意図の並立　(マルチタスク)
  - ⋆ 実行すべき時が来れば意識に上る

意図の働き

- 一貫した行為
  - ⋆ 一度手段を決めれば、他の手段は当面考慮しない(再考慮への抵抗·持続性)
    (例: 駅へはこの道で行くと決めたのだから、しばらくはそうしてみよう)
- 失敗からの回復
  - ⋆ 目標が残っていれば、達成のための意図が選び直される
    (例: 駅への道が工事で通れない、駅へ行くには回り道しよう)

意図の再考慮

- 実世界は動的
- 事情が変われば、手段を選び直さねばならないこともある
- **コミットメント戦略** ··· ある意図にどの程度こだわるか
  - ⋆ blind戦略　(達成するまで意図を捨てない)
  - ⋆ single-minded戦略　(達成したか、あるいは達成可能という信念がなくなったとき、意図を捨てる)
  - ⋆ open-minded戦略　(達成したか、あるいは達成するという願望がなくなったとき、意図を捨てる)

合理的な行為選択

心的状態の**整合性**

- 達成できないと信じる意図は形成しない
  - ⋆ 「歯の治療をすれば痛い」と信じるなら「痛くしないで歯の治療をする」という意図は形成しない(Rao et al., *Modeling Rational Agents within a BDI-Architecture*, 1997.)
- 互いに矛盾する意図は形成しない
  - ⋆ 車が1台しかないという信念があるなら、「車で買い物に出かける」と「家族のために車を残して出かける」の両方を意図することはない(Bratman, *Intention, Plans, and Practical Reason*, 1987.)

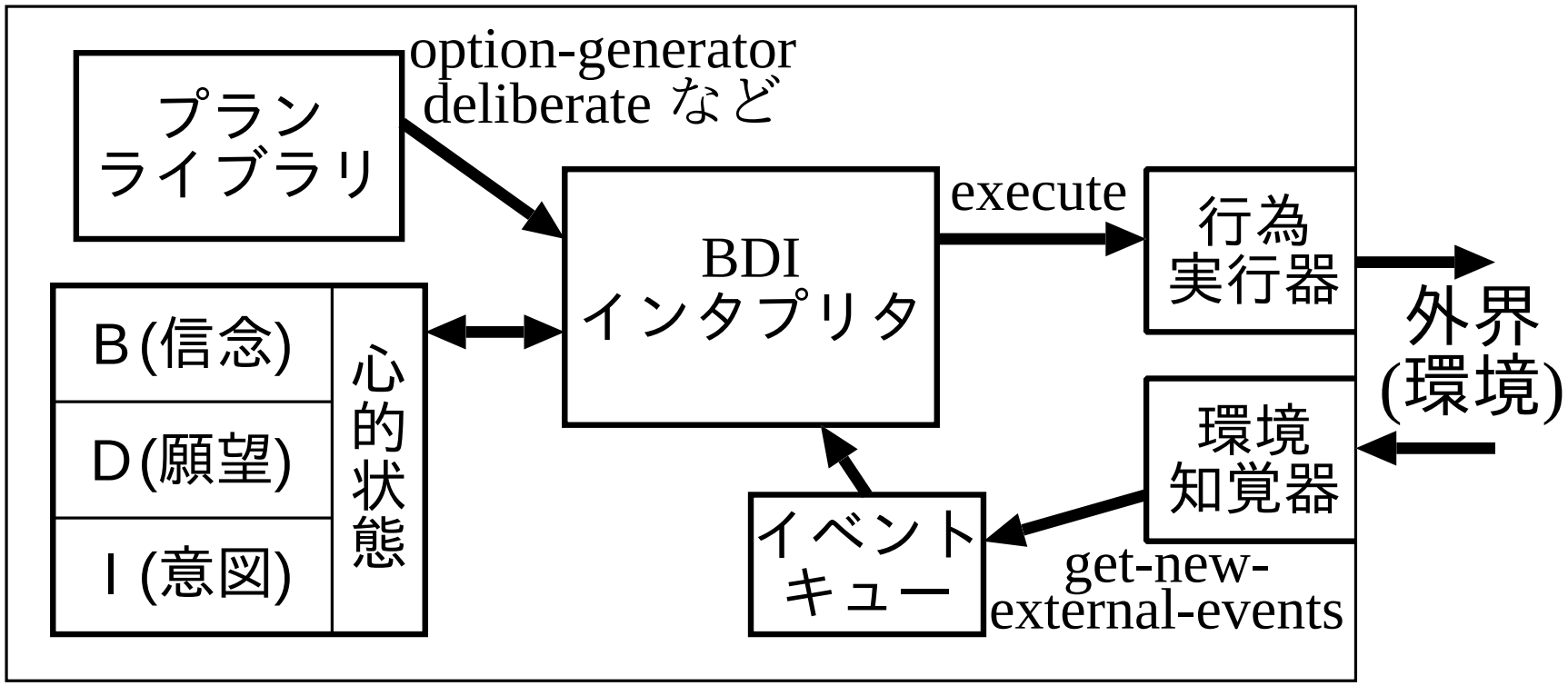
BDI

BDIモデル

- Rao, Georgeffらが提唱(1991～)
- 意図の理論に基づく、自律·合理的エージェントのモデル
  - ⋆ これに基づくエージェントをBDIエージェントと呼ぶ
- 意図の理論をエージェントの行為決定に適用
  - ⋆ 信念·願望·意図を明示的に持ち、これらを用いて行為決定を行う
- エージェントの振る舞いについて形式的に議論する手段(BDI logic)を持つ

BDIアーキテクチャ

- BDIエージェントの基本アーキテクチャ
- 3つの心的状態、プランライブラリ、センサ(環境知覚器)・アクチュエータ(行為実行器)



⋆ BDIモデルではプランは基本的に既成のものを用いる

BDIインタプリタ

BDIエージェントの基本設計となる抽象インタプリタ

BDI-interpreter:
initialize-state();
do
options := option-generator(event-queue, *B*, *D*, *I*);
/* 願望を実現するプランを選択し意図とする */
selected-options := deliberate(options, *B*, *D*, *I*);
/* 次に実行すべき意図を決定 */
update-intentions(selected-options, *B*, *D*, *I*);
/* 手段を選択し、実行する行為を特定 */
execute(*I*); /* 実行 */
get-new-external-events();
/* イベントキューからイベント(知覚情報など)の取り込み */
drop-successful-attitudes(*B*, *D*, *I*);
/* 成功した意図の削除 */
drop-impossible-attitudes(*B*, *D*, *I*);
/* 達成不能になった意図の放棄 */
until quit.

(*B*: 信念 *D*: 願望 *I*: 信念)

BDIインタプリタ(continued)

- option-generator
  - ⋆ 願望と信念から、達成すべき目標が発生し、その達成手段の候補を求める(実際には**プランライブラリ**から探す)
- deliberate
  - ⋆ option-generetaorで候補となったものから、実際にどれを選ぶか**熟考**で決定
- update-intentions
  - ⋆ 手段として選定したプランを**意図**とし、現在実行する意図を決定

BDIインタプリタ(continued)

- execute
  - ⋆ 現在実行する意図の(プランの)本体の１ステップを見て、基本行為なら**実行**。副目標ならその目標を発生
- get-new-external-events
  - ⋆ 外部からの**知覚**を受け取る(行為の成否、新たな刺激など)
- drop-{successful,impossible}-attitudes
  - ⋆ 成功した、あるいは不可能とわかった意図の破棄

例題

ソーダの例 [インタプリタ1巡目]

- option-generator ･･･ 渇きを癒したいので、その手段の候補となるプランを列挙
- deliberate ･･･ 列挙した中からプラン「ソーダを買って飲む」に決定
- update-intentions ･･･ それを意図とし、現在実行する意図に選ぶ
- execute ･･･ 意図の(プランの)本体の1ステップ目は副目標「ソーダを入手」なので、それを目標として生成
- get-new-external-events ･･･ 外部の知覚。特に何もなし
- drop-{successful,impossible}-attitudes ･･･ 特に何もせず

例題(continued)

[2巡目]

- option-generator・・・「ソーダを入手」を達成する方法の候補を列挙
- deliberate・・・近くのコンビニで買うプランに決定

⋮

[このあと何巡かするうち、「ソーダを入手」という目標が達成されたとする]

例題(continued)

[n 巡目]

$\vdots$

- update-intentions $\cdots$「ソーダを買って飲む」意図の続行を決定
- execute $\cdots$ 意図の(プランの)本体の次の1ステップは基本行為「飲む」なのでそれを実行
- get-new-external-events $\cdots$ 成功したという知覚を得る
- drop-{successful,impossible}-attitudes $\cdots$「ソーダを買って飲む」意図は成功したので捨てる

応用

- 実装
  - ⋆ PRS, dMARS(古典)
  - ⋆ AgentSpeak(L) (BDIモデル提唱者Raoによる) → Jason
  - ⋆ Jadex, JACK, etc.
- アプリケーション
  - ⋆ 航空管理システム、スペースシャトル診断 etc.
- 研究面でも基盤として広く用いられる
  - ⋆ AAMAS2020でBDIに触れている発表: 約10件
  - ⋆ EMAS2014のK. Hindriksによる招待講演:

    “· · · most work reported in · · · ProMAS, AOSE, and DALT · · · has taken its inspiration of · · · Belief-Desire-Intention (BDI) agents.”

システムのモデル化例

- 鵜飼いの例
  - ⋆ 鵜
    - ∗ 自分の動ける範囲、魚との位置関係などの知覚情報と、魚の獲り方に関する知識などを信念として持つ
    - ∗ 自分が(あるいは自分の属する集団が)魚を獲るという願望を実現するプランを選定して意図とする
    - ∗ 意図となったプランを用いて、魚を獲ろうとする行動をとる。状況によって意図の再考慮などを行う
    - ∗ 場合によっては仲間に魚を獲らせるために自分は追い込み役をするなどの協調行為をとるかもしれない

システムのモデル化例(continued)

- ★ 鵜匠
  - ∗ 使える鵜の数、河の広さや形、その他の知覚情報と、状況による鵜の配置戦略などを信念として持つ
  - ∗ 自分が鵜を使って魚を獲るという願望を実現するプランを選定して意図とする
  - ∗ 意図となったプランを用いて、鵜を操り魚を獲る行動をとる
  - ∗ 鵜の様子がおかしい、鵜の数が足りないなど、このままでは魚を獲る願望が満たせないと信じる場合、鵜を治癒したり増やしたりする願望が生まれ、そちらを達成する意図が先に選択されてその行動をとる

システムのモデル化例(continued)

- 山火事消火の例
  - ⋆ 司令部、消火部隊、救助部隊など多数のエージェント
    - ∗ 司令部は犠牲者を出さない鎮火を願望し、状況によってそのための最善のプラン(部隊の人数配置、優先順位など)を定め、司令を出すという行動をとる
    - ∗ 消火部隊は消火、救助部隊は救助など、司令された役割の完遂という願望を持ち、その実現のプランを意図として選んで行動
    - ∗ いずれも状況によって意図を再考慮し、配置転換や優先度の変更、消火や救助ができない場合応援を呼ぶ、などの行動をとる

BDI logic

BDI logic

- **記号論理学**の手法を用いて、BDIエージェントの性質を議論するための論理モデル
- **様相論理**を用いて、信念、願望、意図などBDIエージェントに必要な概念を形式的に表現できる

記号論理学

記号による論理学(数理論理学とも)

- 言明を記号で表現
  $\mathsf{father}(\mathsf{namihei}, \mathsf{sazae})\cdots$ 波平はサザエの父
  $\vee\cdots$「または」，$\mathsf{BEL}\,\alpha\cdots$「$\alpha$を信じる」，など
- 自然言語に起因する曖昧さの排除
- 計算機科学との親和性
  自動推論$\cdots$ 記号データの操作

記号論理学(continued)

言明を記号で表現

- 命題論理

  $\phi \vee \psi \cdots \phi$または$\psi$　　$\phi \wedge \psi \cdots \phi$かつ$\psi$

  $\phi \supset \psi \cdots \phi$**ならば**$\psi$　　$\neg\phi \cdots \phi$でない

- 述語論理

  $\forall x\phi \cdots$ 全てのxはϕを満たす

  $\exists x\phi \cdots$ あるxはϕを満たす

- 様相論理(様相命題論理・様相述語論理)

  $\Box\phi \cdots \phi$は必然的に成り立つ

  $\Diamond\phi \cdots \phi$は成り立つ可能性がある　etc.

記号論理学(continued)

自然言語に起因する曖昧さの排除

- 「晴れる」をp、「曇る」をqで表すとき、

  晴れるかまたは曇ることはない

  は$p \vee \neg q$, $\neg(p \vee q)$のどっち?

記号論理学(continued)

計算機科学との親和性

- 「推論」が記号データ上の操作として表現できる
  ...知的システム実現の道具
  - ⋆ 自動演繹
    任意の記号列ϕ, ψに対し、ϕと$\phi \supset \psi$からψを導く
  - ⋆ 論理プログラミング

```
/* exp(X, Y, Z)で「XのY乗がZ」を表す */
exp(X, 0, 1).
exp(X, Y, X*W) ← exp(X, Y-1, W).
```

このプログラムから2^3が下記の過程を経て求まる

exp(2, 3, 8) ← exp(2, 2, 4) ← exp(2, 1, 2) ← exp(2, 0, 1)

意味論

記号論理学での言明(論理式)の意味(真偽)の定め方

- まず、原子命題($\vee$や$\wedge$などのない、すなわちそれ以上分解できない論理式)の真偽を先に決めておき、

ϕ	ψ	$\neg\phi$	$\phi \wedge \psi$	$\phi \vee \psi$	$\phi \supset \psi$	$\forall x\phi$	$\exists x\phi$
偽	偽	真	偽	偽	真	全てのxについてPが真なら真	あるxについて P が真なら真
偽	真	真	偽	真	真		
真	偽	偽	偽	真	偽		
真	真	偽	真	真	真		

によって他の論理式の真偽を決める

$\phi \supset \psi$は「ϕが真でかつψが偽であることはない」と言い換えられる

意味論(continued)

例えば$p \supset q$や$p \wedge (p \supset q)$や$p \wedge (p \supset q) \supset q$の意味は、原子命題$p$や$q$の真偽によって以下のようになる

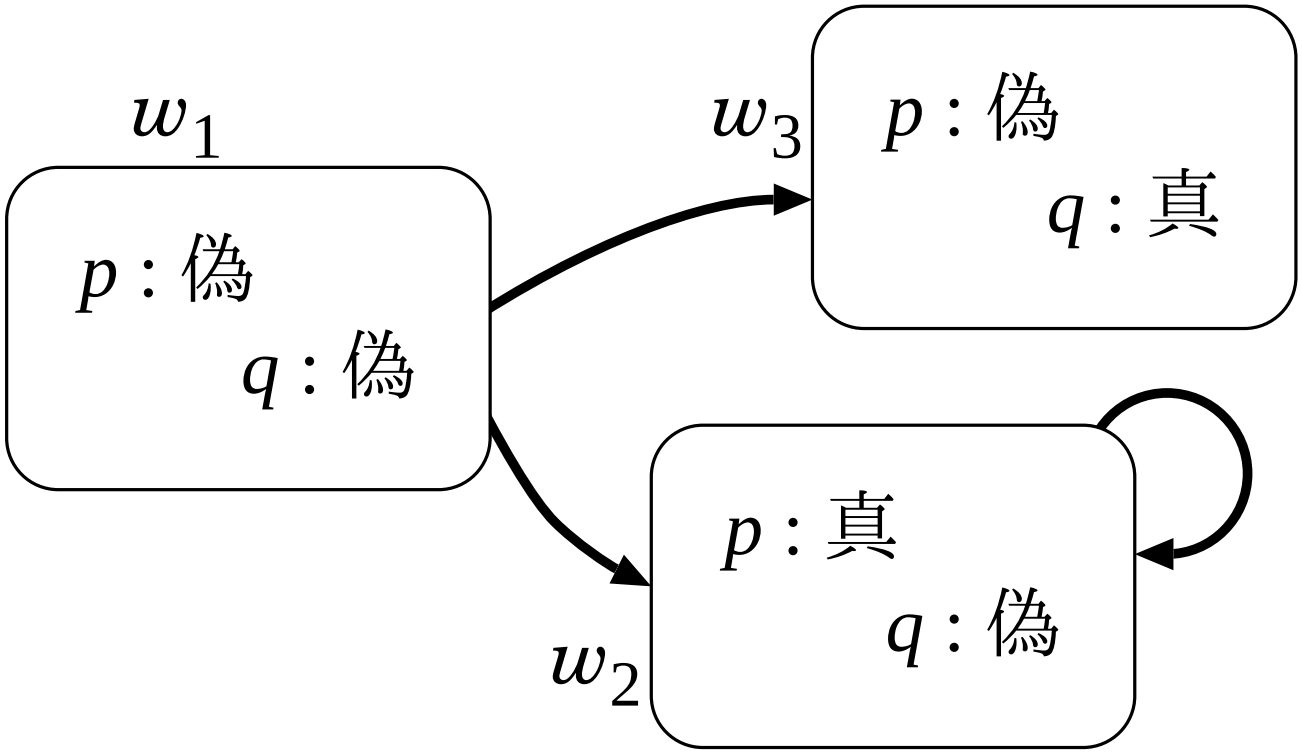
p	q	$p \supset q$	$p \wedge (p \supset q)$	$p \wedge (p \supset q) \supset q$
偽	偽	真	偽	真
偽	真	真	偽	真
真	偽	偽	偽	真
真	真	真	真	真

(ちなみに結合の優先順位は$\neg$, $\wedge$, $\vee$, $\supset$の順)

$p \wedge (p \supset q) \supset q$は、$p$や$q$の真偽に関わらず**常に真(恒真)**。他には$p \vee \neg p$などもそう。$\phi$が恒真であることを$\models \phi$と書く(例えば$\models p \wedge (p \supset q) \supset q$と書く)。

様相論理の意味論

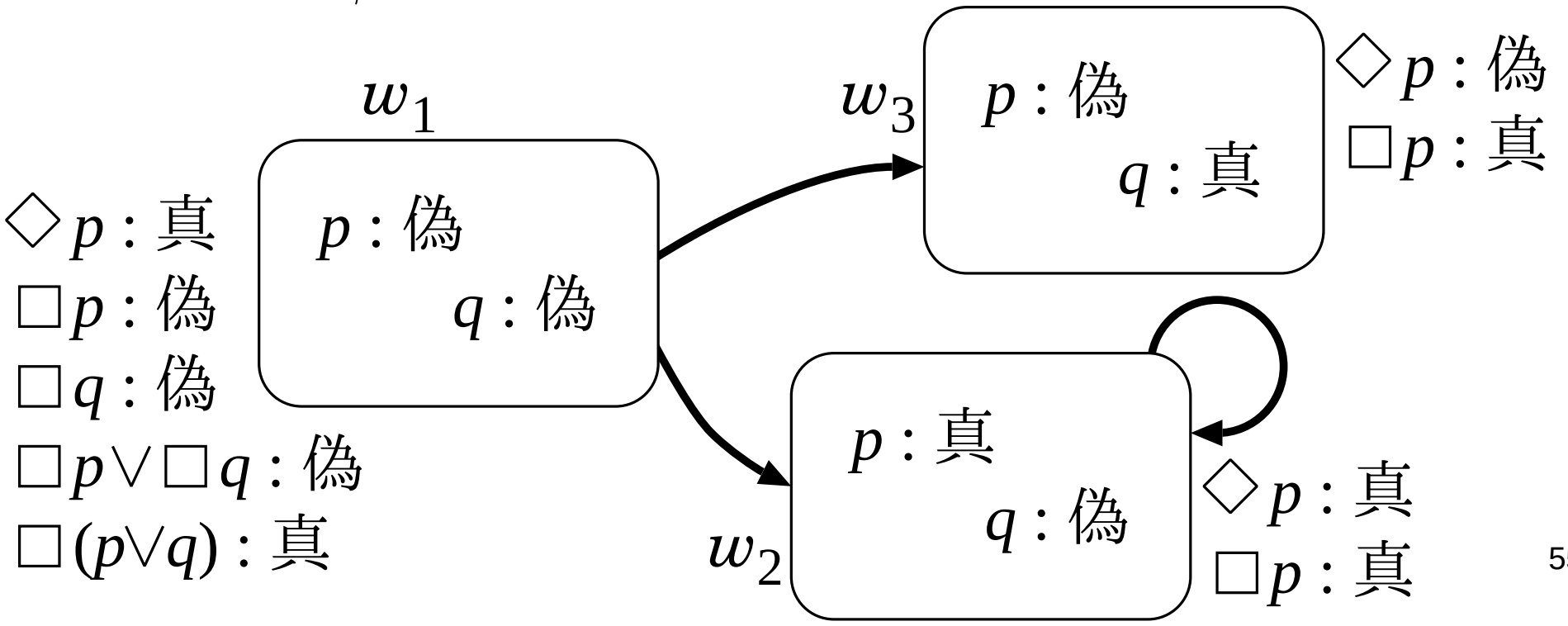
- たくさんの「世界」を考える(**多世界**意味論)
- それぞれの世界で原子命題の真偽を独立に決めておく
- 各世界からどの世界が「見える」かを決める(**可視関係**)



様相論理の意味論(continued)

こうした上で以下のように決める

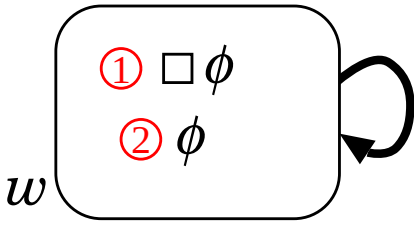
- ある世界wで$\Box\phi$が真であるのは、wから見える**全ての**世界でϕが真であるとき
- ある世界wで$\Diamond\phi$が真であるのは、wから見える**いずれかの**世界でϕが真であるとき



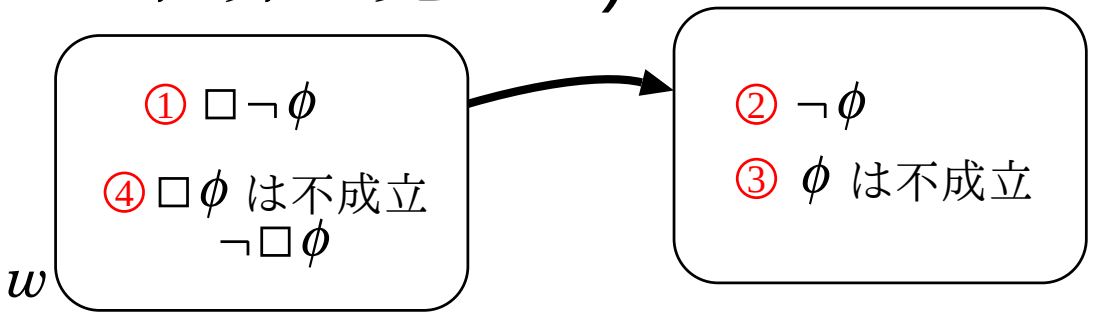
様相論理の意味論(continued)

可視関係への制限の入れ方により、□はさまざまな性質を持つ。例えば

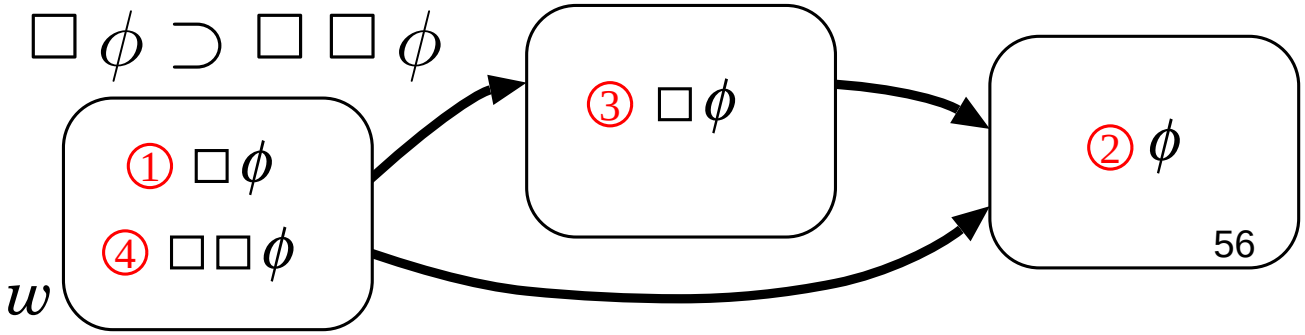
1. 可視関係が反射的(自分自身は必ず見える)なら、$\models \Box\phi \supset \phi$



2. 可視関係がSerial(最低1つの世界が見える)なら、$\models \Box\neg\phi \supset \neg\Box\phi$



3. 可視関係が推移的(2段階で行けるところには必ず1段階で行ける)なら、$\models \Box\phi \supset \Box\Box\phi$



様相論理の意味論(continued)

「□」を

- 「信念」と捉えたい場合は2. と3.
- 「願望」や「意図」と捉えたい場合は2.
- 「知識」と捉えたい場合は1. と3.
- 「次の時刻」と捉えたい場合は2.

を仮定することが多い

例: 「□」を「信念」と捉え、$\Box\phi$を$\mathrm{BEL}\,\phi$と書くことにし、2. と3. を仮定すると、以下が言える

- $\models \mathrm{BEL}\,\neg\phi \supset \neg\,\mathrm{BEL}\,\phi$ (相反する 2つのことは信じない)
- $\models \mathrm{BEL}\,\phi \supset \mathrm{BEL}\,\mathrm{BEL}\,\phi$ (あることを信じていれば、それを信じていると信じる)

様相論理の意味論(continued)

可視関係を2種類以上導入すると、さまざまな様相を同時に扱える

例: 可視関係を2種類導入し

- 1つ目の可視関係による□を「次の時刻」と捉えてAXと書く
- 2つ目の可視関係による□を「信念」と捉えてBELと書く

すると

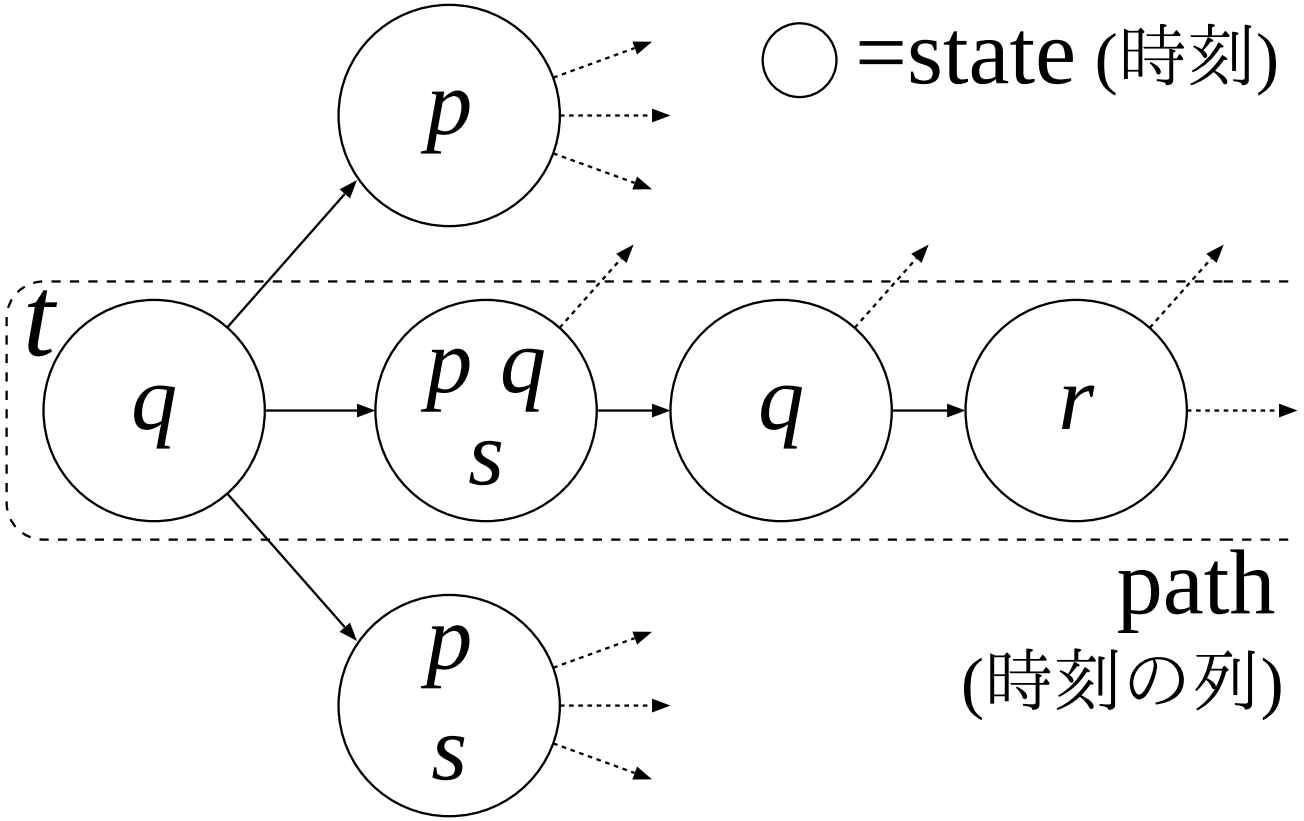
- AX BEL ϕは「ϕを次の時刻に信じる」を表す
- BEL AX ϕは「次の時刻にϕであると今信じる」を表す

BDI logic

- BDIモデルの記述に用いる様相論理
- エージェントの性質を形式的に議論
- 以下の様相(心的状態様相・時相)オペレータを持つ
  - ⋆ BEL$\phi \cdots \phi$を信念に持つ
  - ⋆ DESIRE$\phi \cdots \phi$を願望する
  - ⋆ INTEND$\phi \cdots \phi$を意図する
  - ⋆ A$\phi \cdots$ 全ての未来でϕ
  - ⋆ E$\phi \cdots$ ある未来でϕ
  - ⋆ X$\phi \cdots$ 次の時刻にϕ
  - ⋆ G$\phi \cdots$ 永遠にϕ
  - ⋆ F$\phi \cdots$ いつかϕ
  - ⋆ ϕ U $\psi \cdots \psi$が成り立つまでϕ
  - ⋆ $does(e) \cdots$ 基本行為eを実行する

分岐時間様相

BDI logicでは、時間は離散で、未来方向に木構造

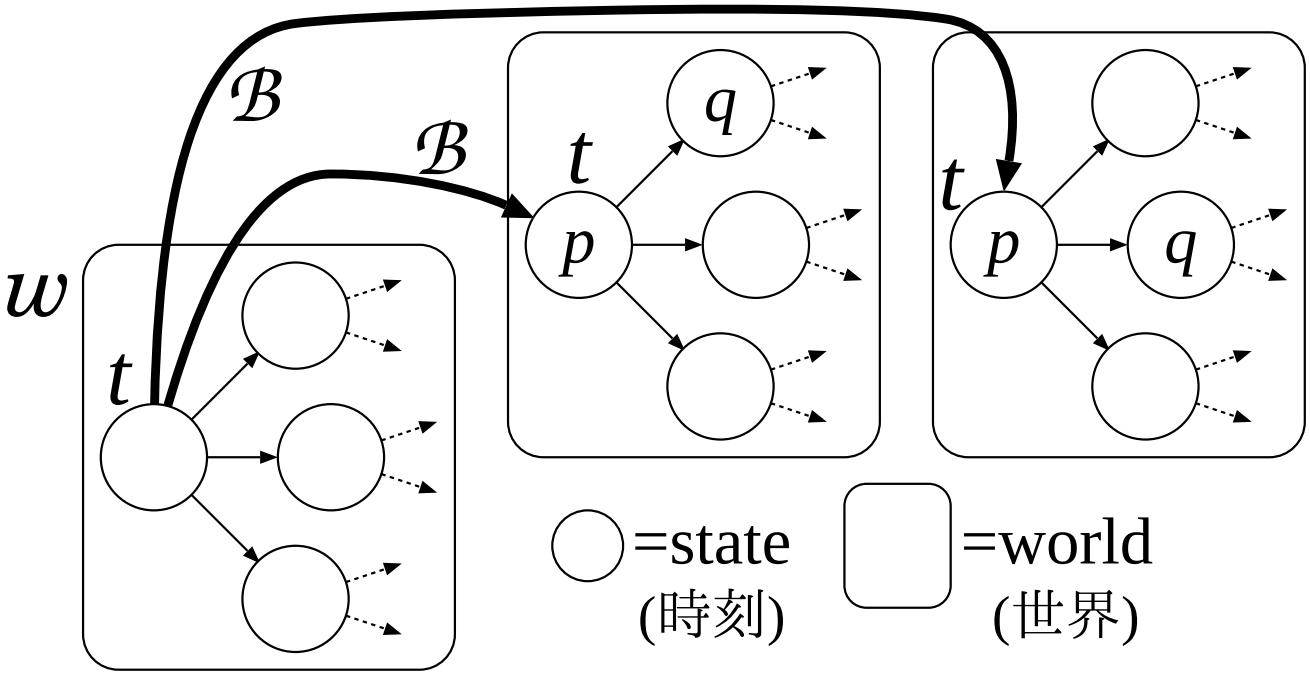


時刻tで真になる論理式の例: $\mathsf{AX}\,p$, $\mathsf{E}(q\,\mathsf{U}\,r)$

心的状態様相(信念・願望・意図)

BDIストラクチャ

- 時間の木が1つの「世界」
- 世界間の可視関係$\mathcal{B}$で表される様相を「信念」、$\mathcal{D}$が同「願望」、$\mathcal{I}$が同「意図」と考える



世界wの時刻tで真となる論理式の例: BELp, BEL EX q

BDI logicによる記述例

INTEND ソーダを買って飲む $\supset$
　INTEND AF ソーダを入手 $\wedge$
　AG(BEL ソーダを入手 $\supset$ INTEND $does$(飲む) $\wedge$
　　AX(BEL $succeeded$(飲む) $\supset$ BEL 渇きが癒される))

INTEND $does$(飲む) $\supset$ $does$(飲む)

- 今「ソーダを買って飲む」意図を持つと、まず「ソーダを入手」する意図を形成する。
その後、「ソーダを入手」したという信念が得られた時、基本行為「飲む」を意図し(従って実行し)、次の時刻に「飲む」が成功したという信念が得られれば、渇きが癒されたと信じる

BDI logicによる記述例(continued)

コミットメント戦略 · · · 意図の持続性

$\mathrm{INTEND}\,\phi \supset \mathrm{A}(\mathrm{INTEND}\,\phi\ \mathrm{U}\ \mathrm{BEL}\,\phi)$

· · · blindコミットメント戦略

$\mathrm{INTEND}\,\phi \supset \mathrm{A}(\mathrm{INTEND}\,\phi\ \mathrm{U}\ (\mathrm{BEL}\,\phi \vee \neg\,\mathrm{BEL}\,\mathrm{EF}\,\phi))$

· · · single-mindedコミットメント戦略

$\mathrm{INTEND}\,\phi \supset \mathrm{A}(\mathrm{INTEND}\,\phi\,\mathrm{U}\,(\mathrm{BEL}\,\phi \vee \neg\,\mathrm{DESIRE}\,\mathrm{EF}\,\phi))$

· · · open-mindedコミットメント戦略

下に行くほど専念の度合が低い

BDI logicによる記述例(continued)

心的状態の整合性(これらを成り立たせるような可視関係への制限を仮定する)

$\models \mathrm{DESIRE}\,\phi \supset \mathrm{BEL}\,\mathrm{DESIRE}\,\phi$

・・・内省公理 (INTENDやBELに対しても)

$\models \mathrm{DESIRE}\,\mathrm{EF}\,\phi \supset \mathrm{BEL}\,\mathrm{EF}\,\phi$

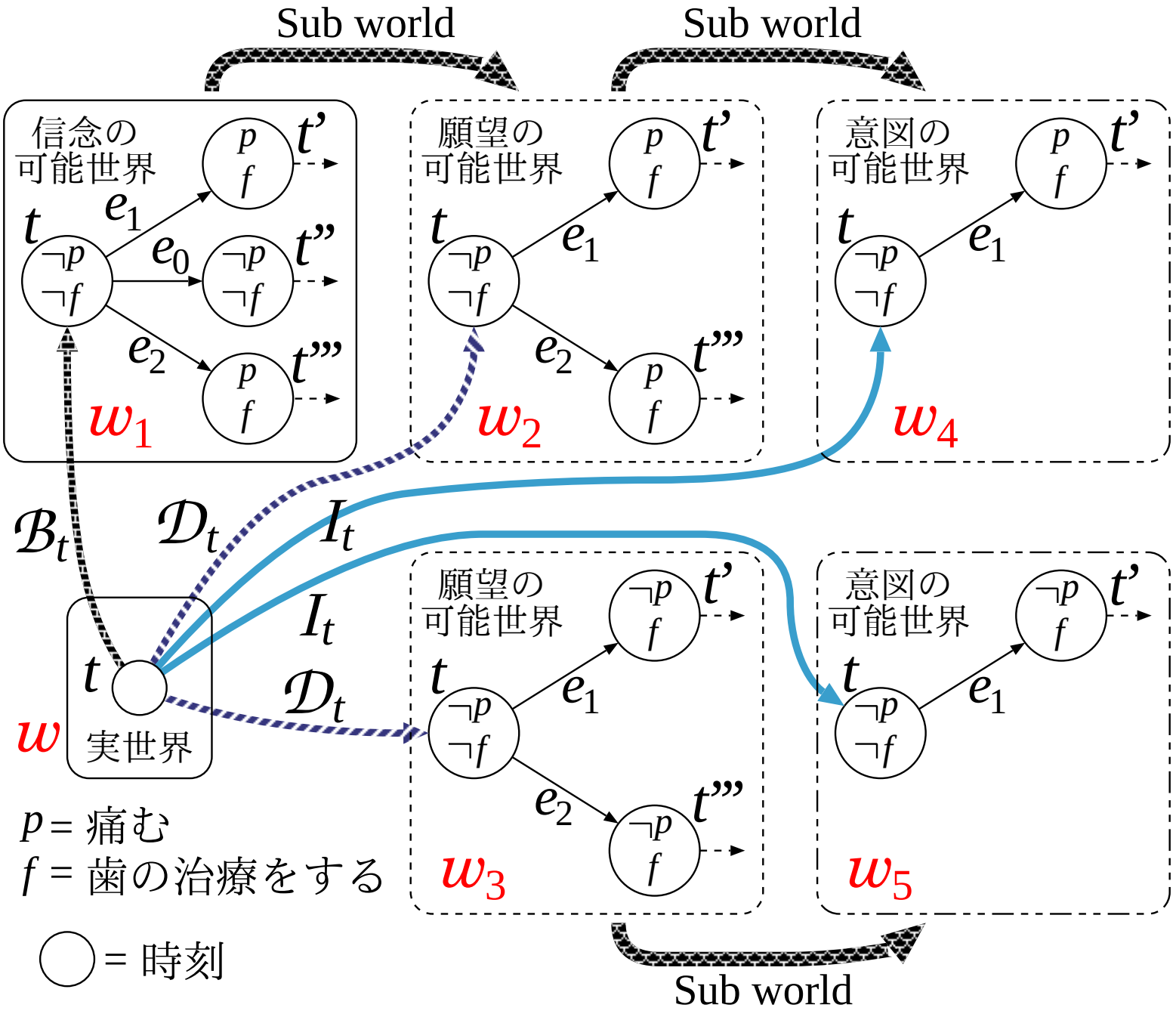
・・・願望と信念の整合性 (INTENDとBELに対しても)

下のものの対偶$\models \neg \mathrm{BEL}\,\mathrm{EF}\,\phi \supset \neg \mathrm{DESIRE}\,\mathrm{EF}\,\phi$と、信念の性質$\models \mathrm{BEL}\,\neg \mathrm{EF}\,\phi \supset \neg \mathrm{BEL}\,\mathrm{EF}\,\phi$から

$\models \mathrm{BEL}\,\neg \mathrm{EF}\,\phi \supset \neg \mathrm{DESIRE}\,\mathrm{EF}\,\phi$

が出る

BDI logicによる意思決定のモデル化



BDI logicによる意思決定のモデル化(continued)

- 信念として持つ時系列中から願望を、さらにその中から意図を選択(Sub world; 時間の流れを制限した世界)
- 信念と整合しない願望(痛まずに歯を治したい)や意図は持たない

  $\models \mathsf{BEL} \neg \mathsf{EF}(f \wedge \neg p) \supset \neg \mathsf{DESIRE}\, \mathsf{EF}(f \wedge \neg p)$
- 願望は信念に引きずられない(痛むことは願望しない)

  $\mathsf{BEL}\, \mathsf{AG}(f \supset p) \wedge \mathsf{DESIRE}\, \mathsf{EF}\, f \supset \mathsf{DESIRE}\, \mathsf{EF}\, p$は恒真ではない

エージェントの意思決定過程やその合理性をうまくモデル化できている

BDIモデルの有効性

- (意図の理論で述べられた)人間のものに近い行為決定
  - ⋆ 問題解決に向けた一貫した行為
  - ⋆ プランの部分性
  - ⋆ 複数意図の並行処理
  - ⋆ 合理性、再考慮etc.
- 行為決定のモデル化、(BDI logicによる)形式化

BDIモデルに元々ない部分

- 技能(反射的行為)の獲得と利用(i.e. 機械学習)
- 社会性(社会的義務など)
- 非合理的な行為(感情、etc.)
- プランニング(プランを動的に作成)

BDIモデルは、人間の行為を決めるメカニズムのうち主に「頭で考える」部分を取り上げたもの

BDIモデルに対する拡張

- BDIエージェントに対する拡張の要求
  - ⋆ 例えば
    1. 概念の追加(「技能」「権利」「義務」や「感情」など)
    2. 異なる行為決定機構
  - ⋆ 2.の例として考えうるもの
    - ∗ 機械学習(例えば強化学習)との結合
    - ∗ 確率的な推論の導入(不確かさを考慮した推論)
    - ∗ プランナの導入

BDIモデルに対する拡張(continued)

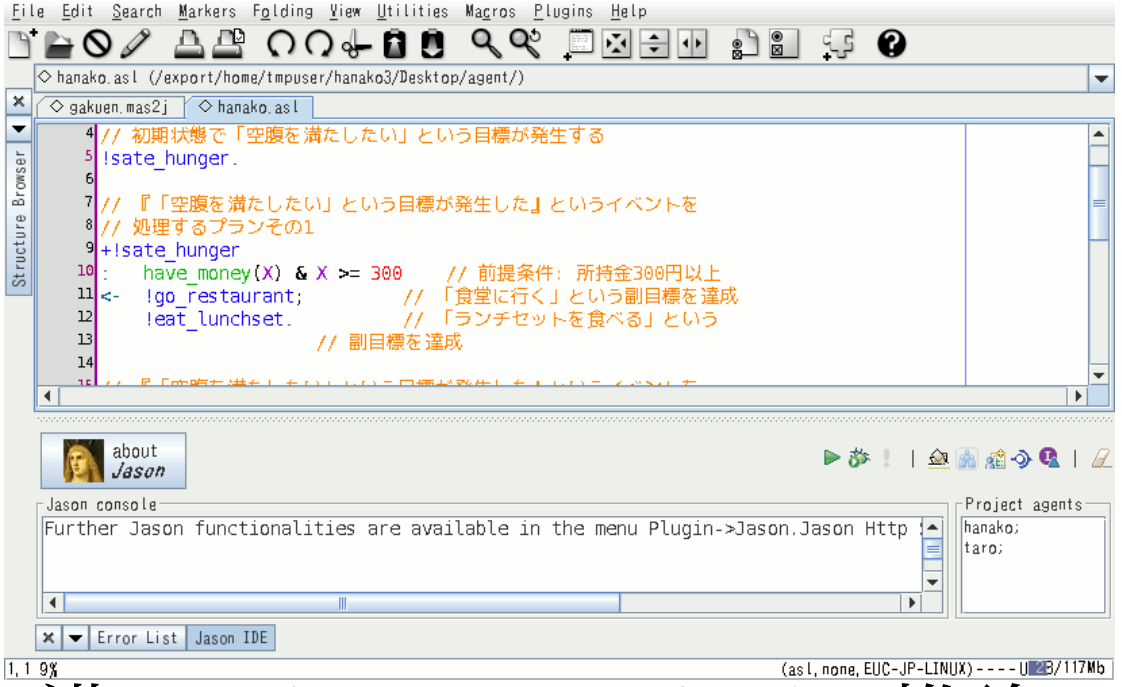
- BDI logicへの拡張
  - ⋆ 例えば
    1. 様相オペレータの追加(「権利」「義務」などを記述)
    2. 確率的遷移の記述(強化学習などの記述が可能に)
    3. マルチエージェントへの拡張
    4. ベースとなる論理体型の変更(矛盾許容論理、線形論理、直観主義論理など)

Jason

—BDIエージェント構築プラットフォームの例—

Jason

- BDIアーキテクチャに基づくエージェント構築の道具として、(BDIアーキテクチャ提唱者の)Raoらが開発
- AgentSpeak(L)インタプリタ+IDE
- エージェントのプランや信念などを記述しておき、エージェントを動かす



- プランの前提条件が満たされているかなどの推論には、Prolog言語と同様の後向き推論が使われている

Jasonでのエージェント記述

- エージェントのプランや、初期状態での信念・目標などを記述
- 1つのプランは原則として「トリガイベント：前提条件 <- 本体」の形(「：前提条件」は省略する場合もあり、その場合は無条件)
- プランの本体の中身は

  基本行為または副目標;

  ...

  基本行為または副目標.

  の形
- 「+!」で始まるものは、目標の発生を表す「イベント」
- 「!」で始まるものは、目標(または副目標)の発生

Jasonでのエージェント記述(continued)

エージェントhanako

```
// 初期状態で「400円持っている」という信念を持っている
have_money(400).

// 初期状態で「空腹を満たしたい」という目標が発生する
!sate_hunger.

// 『「空腹を満たしたい」という目標が発生した』という
// イベントを処理するプランその1
+!sate_hunger
:   have_money(X) & X >= 300   // 前提条件: 所持金300円以上
<-  !go_restaurant;            // 「食堂に行く」という副目標を達成
    !eat_lunchset.             // 「ランチセットを食べる」という
                               // 副目標を達成
```

```
// 『「空腹を満たしたい」という目標が発生した』という
// イベントを処理するプランその2
+!sate_hunger
:   have_money(X) & X < 300
<-  !go_restaurant;
    !eat_sandwich.

// 『「食堂に行く」という目標が発生した』という
// イベントを処理するプラン
+!go_restaurant    // 前提条件が省略されると「無条件」扱い
<-  .print("going to the restaurant").
    // ここでは実際は画面に何がしか出力しているだけ
    // .printは出力を行う基本行為

// 『「ランチセットを食べる」という目標が発生した』という
// イベントを処理するプラン
+!eat_lunchset
<-  .print("eating lunch set").

+!eat_sandwich  <-  .print("eating sandwich").
```

Jasonでのエージェント記述(continued)

エージェントtaro

```
have_money(300).
!sate_hunger.

+!sate_hunger
:   have_money(X) & X >= 200
<-  !go_restaurant;  !eat_curry;  !eat_noodle.

+!sate_hunger
:   have_money(X) & X < 200
<-  !go_restaurant;  !eat_noodle.

+!go_restaurant <- .print("going to the restaurant").
+!eat_curry     <- .print("eating curry").
+!eat_noodle    <- .print("eating noodle").
```

Jasonでのプランの選択

- 目標が発生したら、その目標が「発生した」という**イベント**が起きる
- そのイベントと一致する**トリガイベント**を持つプランが、意図として選ばれる候補となる
- 候補の中から、**前提条件**が現在の状況下で真であるもの1つが「意図」として選ばれる
  - 2つ以上ある場合は、標準的にはそのうち最初のものが選ばれるが、変更も可能

以上は、現在の状況から、プランに記述された規則によって行為を決めている。このような「行為を決めるための推論」が「**実践的推論**」であり、事実のみに関する推論と区別される。

Jasonでのプランの選択と実行の例

エージェントhanakoの例だと

- sate_hungerという目標が発生(!sate_hanger)
- それによって、+!sate_hungerというイベントが発生
- そのイベントをトリガイベントに持つプランを探す
- そのようなプランの中で、前提条件が満たされるものが選ばれ「意図」となる
- 意図はそれ1つしかないので、本体を実行する
- 本体の実行は、まず副目標「go_restaurant」を発生させて達成
- 次に「eat_lunchset」という副目標を発生させて達成

目標を副目標に置き換えていく過程は、「ある目標の達成を、プランを使って、別な目標の達成に帰着する」というものであり、Prolog言語での後向き推論の過程(「ある質問に答えることを、規則を使って、別なある質問に答えることに帰着する」)に類似

まとめ

- BDIモデルは、人間の行為決定を模した、自律エージェントのモデル
- BDI logicにより、その振る舞いや性質を形式的に記述・議論できる
- ただしあくまでモデルでありフレームワークであって、具体的なシステムでどのように意図を選択したり再考したりするかは、設計者に委ねられる

参考文献

- 山川他: 特集: 意図研究のスペクトル, 人工知能学会誌, Vol. 20, No. 4, pp. 357–455, 2005. 「意図」に関する解説記事、BDIモデルの解説もあり
- Bratman: *Intention, Plans, and Practical Reason*. Harvard University Press, 1987. (門脇他訳, 意図と行為—合理性、計画、実践的推論—, 産業図書, 1994.) 「意図の理論」の原典
- Rao et al.: Modeling Rational Agents within a BDI-Architecture. In *Reading in Agents*, Morgan Kaufmann, pp. 317–328, 1997. BDIモデルによる合理的エージェントの形式化
- Singh et al.: Formal Methods in DAI: Logic-Based Representation and Reasoning. In *Multiagent Systems*, The MIT Press, pp. 331–376, 1999. BDIアーキテクチャの解説
- Bordini et al.: *Programming multi-agent systems in AgentSpeak using Jason*, Wiley, 2007. BDIエージェントとJasonについて