

# BDI

## —モデル、アーキテクチャ、論理—

新出尚之(奈良女子大学)

nide@ics.nara-wu.ac.jp

全脳アーキテクチャ勉強会

2014年7月1日

– Typeset by Foil $\text{\TeX}$  –

# 概要

- 意図の理論
- BDI
  - ★ BDIモデル
  - ★ BDIアーキテクチャ
  - ★ BDI論理
- BDIと他の行為決定メカニズムの結びつき

## その前に**自分の立ち位置**

- 人間の思考の過程の実現に興味あり
- でも、実際の脳のしくみについては考えたことがほとんどなかった
  - ★ 「これまでに得られている理論」と「実際の脳のしくみとの関連」を、新たなアプローチの1つにしたい

## 自分の立ち位置 (continued)

- 「人間の思考の過程の実現」の手段として「記号論理」は依然有力だと思う
  - ★ JAWS2012でのR. Kowalskiの招待講演は記憶に新しい
  - ★ 記号論理とロボティクスの結びつきも重要視されつつある?(ICLP 2013の併設ワークショップKRRRR2013など)
- 人間の脳は論理だけで動いているわけではなさそうだが、軸としては論理を用いて考えたい
  - ★ 論理に基づく行為決定には、「なぜそうなったか」を説明できる利点があるのでは?

# 意図の理論

# 意図の理論

- 哲学者 Bratman が提唱 (1987)
- 人間の目標達成に向けての行為選択の説明
- 信念・願望・意図の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)

# 行為選択の過程

- 信念と願望から、達成すべき目標とその手段(計画、**プラン**)を選定(**熟考**)
- 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**

# 行為選択の過程(例)

(Singh他, 1999)  
一部改変

- 喉が渴いた 渴きを癒したい
  - ★ 達成すべき目標: 渴きを癒す
  - ★ 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

プラン名などの一部を講演時とは変更しております



# 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

## プランの**実行**: 本体を順に達成

- 副目標 … その目標を生成し、(手段を再帰的に選んで)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: 渴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

# プランの部分性(continued)

- 最初から目標達成までの全行為を決める必要はない
- 最初は大まかな計画を立て、必要になれば細部を決める
  - ★ 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

# 未来指向的 vs. 現在指向的

- 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
  - ★ (WBA勉強会があるから、7/1は新幹線で東京に行こう)
- 達成すべき時が来れば実行に移す… 現在指向的意図
  - ★ (WBA勉強会の日がきたから、これから新幹線で東京に行こう)
- 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  
(この道を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  
(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
- **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 (達成不能とわかったら意図を捨てる)
  - ★ opene-minded戦略 (達成不要になったら意図を捨てる)

# 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 他, 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, 1987)

**BDI**

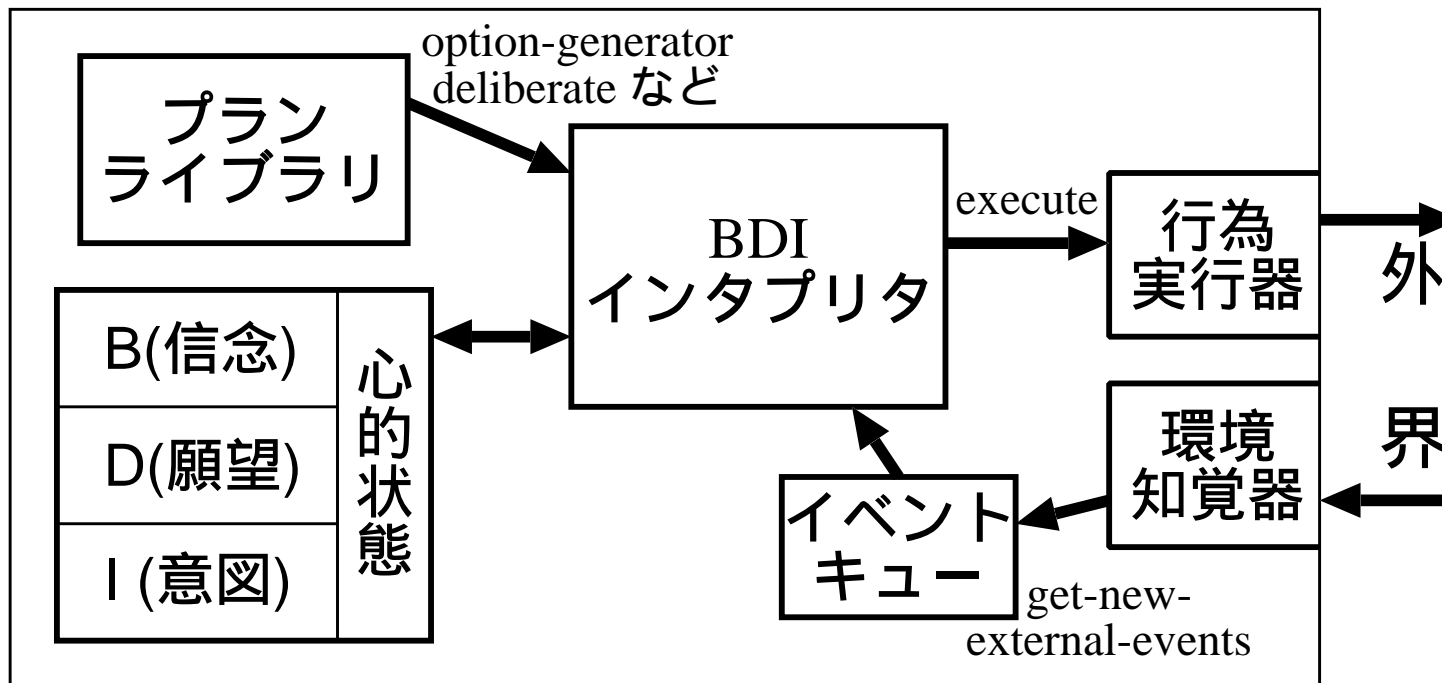


# BDIモデル

- 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
- 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
- これに基づくエージェントをBDIエージェントと呼ぶ

# 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・アクチュエータ...



# BDIインタプリタ

## 抽象インタプリタ

### **BDI-interpreter:**

initialize-state();

#### **do**

options := option-generator(event-queue,  $B$ ,  $D$ ,  $I$ );

selected-options := deliberate(options,  $B$ ,  $D$ ,  $I$ );

update-intentions(selected-options,  $B$ ,  $D$ ,  $I$ );

execute( $I$ );

get-new-external-events();

drop-successful-attitudes( $B$ ,  $D$ ,  $I$ );

drop-impossible-attitudes( $B$ ,  $D$ ,  $I$ );

**until** quit.

# 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 ... 渴きを癒したいので、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

## [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]

## [ $n$ 巡目]

⋮

- update-intentions ... 「ソーダを買って飲む」意図の続行を決定
- execute ... 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 ... 成功したという知覚を得る
- drop- $\{$ successful,impossible $\}$ -attitudes ... 「ソーダを買って飲む」意図は成功したので捨てる



# 実装

- PRS, dMARS (古典)
- AgentSpeak(L) (BDIモデル提唱者Raoによる) → Jason
- JADEX, JACK, etc.

# 応用

- PRS, dMARS... 大規模アプリケーションでの利用
  - ★ 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
  - ★ AAMAS2014でBDIに触れている発表: 約60件(約600件中)
    - \* Antunesら: 願望を自ら獲得・発展させるBDIエージェント
    - \* Leeら: 社会規範を遵守するBDIエージェント
    - \* Dignumら: マルチエージェント向けBDIモデル
    - \* Hindriks(招待講演): "... most work reported in ... ProMAS, AOSE, and DALT ... has taken its inspiration of ... Belief-Desire-Intention (BDI) agents."

# BDIモデルに元々ない部分

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

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 logic

- BDIエージェントを形式的に記述する様相論理体系
- オペレータいろいろ
  - ★  $\wedge$ (かつ),  $\vee$ (または),  $\neg$ (否定),  $\supset$ (ならば)
  - ★ BEL(信念), DESIRE(願望), INTEND(意図)
  - ★ A(全ての時間分岐で), E(ある時間分岐で)
  - ★ X(次の時刻に), G(永遠に), F(いつかは), U(あることが成り立つまで)
  - ★  $does(e)$  (基本行為  $e$  を実行する) など、イベント関連オペレータ

# BDI logic

さまざまな性質を記述し、それが成り立つか否かを形式的に議論できる

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 渴きが癒される))

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、渴きが癒されたと信じる

# BDI logic

さまざまな性質を記述し、それが成り立つか否かを形式的に議論できる

$A(\text{INTEND AF } s \text{ U BEL } q)$

- ( $s$ を「ソーダを買って飲む」、 $q$ を「渇きが癒される」とする)  
渇きが癒されたと信じるまでは、ソーダを買って飲む意図を持ち続ける
- **blind**コミットメント戦略の例

# BDI logic

さまざまな性質を記述し、それが成り立つか否かを形式的に議論できる

$$\text{BEL AG}(f \supset p) \supset \neg \text{INTEND EF}(f \wedge \neg p)$$

- ( $f$ を「歯を治療する」、 $p$ を「痛む」とする)  
「常に歯の治療をすると痛む」と信じるなら、「あわよくば痛まずに歯を治療する」ことは意図しない
- **合理的な行為選択**の説明ができる

## BDI logic(余談)

実のところ、BDI logicについては考え直してみたい点多数

- 心的状態ではなくその変化を書きたい場合が多い
- 心的状態が暗黙には保持されないのが不自然
- 意図は論理式に対する様相オペレータでいいのか？
- etc.



# BDIモデルの効用

- (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
- 行為の相互理解 —なぜそのような行動をとったのか

# BDIと他の行為決定メカニズムの 結びつき

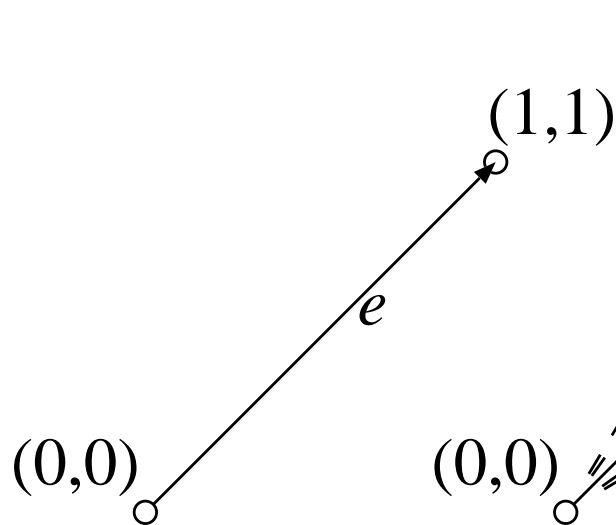
# 機械学習との融合

- BDI … 熟考の部分
- 機械学習 (特に強化学習) … 技能の獲得と利用の部分
- 人間は行為にあたって両者を併用している
  - ★ 例: 自転車で駅まで行く際
    - \* 最適なルートを経由するプランを熟考で選び、意図して実行
    - \* そのルートを巡航するには、倒れないで自転車を漕ぐ技能(学習で獲得)を使う
- 学習も「力任せ」ではなくある程度考えながら行っているかも(学習量から見て)
- ルールやプランの学習は? (帰納的学習?)

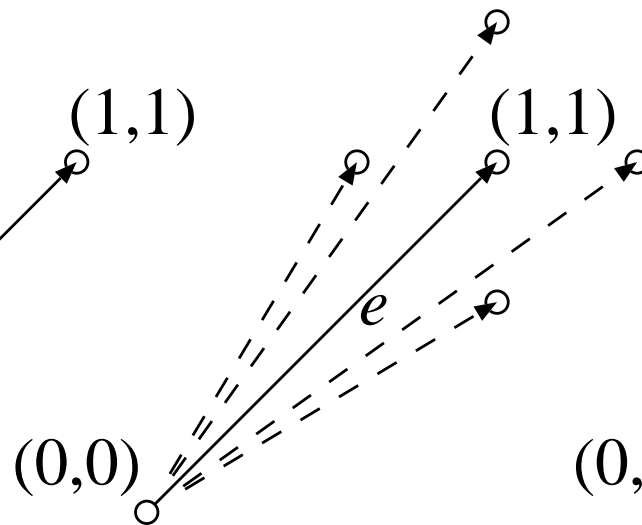
# 確率推論との融合

- 実世界の乱雑さ
  - ★ 行為の結果は確率的分布
  - ★ 確率分布つきBDI logic:  $C-TOMATES$  (新出他, 2013)

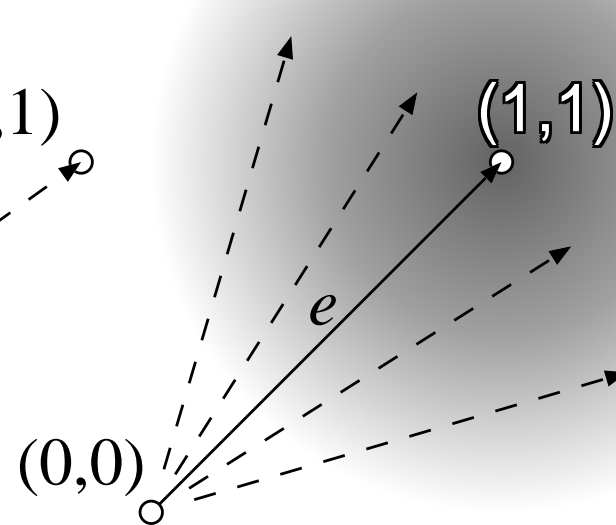
非確率的遷移



離散確率的遷移



連続確率的遷移



# 確率推論との融合

- 不確実な行為
  - ★ あるプランを実行すると、成功確率が $n\%$
  - ★ 成功確率が10%を下回るなら、別なプランを選んで意図をそのプランに切り替え
- 確率的な因果関係
  - ★ 例: 行為がうまくいかない原因の推測
  - ★ ベイジアンネットなどに関連?

# プランナとの結合

- プランライブラリにないプランの合成
  - ★ BDIの熟考ルーチンが必要に応じてプランナを呼び出す
  - ★ 例: 強化学習で得た方策をプランナで合成して他の問題に流用 (JAWS2010, 新出)
    - \* 人間が、学習で得たスキルを他の問題にも応用することと似ている?

# まとめ

- BDIモデルや、BDIと他の行為決定手法の融合は、人間も同じようなことをやっていると思えるものが多い
- 脳の仕組みとの対応が取れれば、いろいろなことが説明できたり実現に役立ったりするかも？

# 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