マルチエージェントシステムのための論理, *ver. 0.5*

Logics for Multiagent Systems, ver. 0.5

内容

1. 概論
2. なぜ理論が必要なのか>?
3. エージェント＝意図するシステム
4. 態度の理論
5. 態度の形式化
6. 標準様相論理

   (以下は準備中)
7. 意図
8. 発話行為の意味論
9. 協調の一理論
10. 形式的枠組み
11. 4ステージモデル
    (1) 認識，(2) チーム形成，(3) プラン形成，(4) チーム行動

1. 概論 (1/1)

- この章の目的は理論専門家がエージェントを概念化する方法を概観し，エージェント理論の重要な開発過程を要約すること．
- 「なぜ，理論が必要なのか?」という質問に答えることから始める．
- エージェントを特徴付けるために用いられる様々な異なる態度に関して議論する．
- 態度の形式化に伴ういくつかの問題を紹介する．
- 様相論理を，知識に焦点をあてて態度に関する推論をするための道具として紹介する．
- Moor の能力理論に関して議論する．
- エージェント理論のケーススタディとして Cohen-Levesque の意図理論を紹介する．

2. なぜ理論が必要なのか? (1/2)

- 形式的方法論は一般的なソフトウェア開発の実践過程に対しては，ほとんど影響をおよぼしてこなかった．では，なぜ，エージェントに基づくシステムでは形式的方法論が必要なのか?
- それに対する答えは,「エージェント開発に用いるアーキテクチャ，言語，そしてツールに対して意味論を与えることができる必要があるからだ」
- そのような意味論がなければ，何が起きているのか，また，なぜそれがうまくいくのかを決して明確には説明できない．
- エンドユーザ (つまりプログラマ) は一生，これらの意味論を読んだり理解したりする必要はないだろうが，これらの意味論なしでは，プログラミング言語開発において進歩がなくなるのだ．

2. なぜ理論が必要なのか? (2/2)

- エージェントに基づくシステムには多くの概念やツールがあり，それらは (比喩や類推により) 直感的には理解しやすく，そして，明らかに可能性を持っている.
- しかし，これらのツールを学術的に深く理解するためには理論が必要だ!

3. エージェント＝意図するシステム (1/1)

- 理論専門家の開始点はどこにあるのか?
- 意図するシステムとしてのエージェントの概念
- エージェントの理論専門家は，エージェントを意図するシステムとみなすことから始める．もっとも単純で無矛盾な記述は意図的な立場を求める．

4. 態度の理論 [Theories of Attitude](1/2)

- 計算機システムを「知能的」概念に基づく言葉で設計し 構築したい
- そうではあるが，これが出来るようになる前に，そういう態度に対する計算的に扱いやすい理論とシステムの挙動を生成するための相互作用に関する方法論を同定する必要がある
- よって，最初には，どの態度をとるのかが重要になる

4. 態度の理論 [Theories of Attitude](2/2)

- 2 つのカテゴリ
  - 情報的態度 (information-attitude)
    - 信念
    - 知識
  - 計画的態度 (pro-attitude)
    - 願望
    - 意図
    - 義務
    - コミット メント (責務)
    - 選択
    - . . . . . .

5. 態度の形式化 (1/2)

- どのようにすれば態度を形式化できる?
- 以下を考えてみる

Taro Believes Ei-Ichi is father of Ikkei.

- これをとりあえず一階論理に翻訳してみると以下のようになる

$$Bel(Taro,\ Father(Ikkei,\ Ei-Ichi))$$

  - しかし，Bel 述語の第二引数は一階述語の論理式であり，項ではない．

    “Bel” を論理式に適用できるようにする必要がある

  - 項を同じ外延で置き換えることが可能 ($Ei-Ichi$ は $Osawa$ と同じ)

    意図概念は参照的に不透明 (referentially opaque)

5. 態度の形式化 (2/2)

- よって，意図的概念を論理的に形式化するためには，以下の二つの異なる種類の問題を解かなければならない
  - 統語論:
  - 意味論:同値関係による置き換えの禁止
- 統語論問題に対する二つの基本的な方法
  - 論理式に適用される様相演算子を含む様相言語の利用
  - メタ言語の利用．目的言語の論理式を記述する項を含む一階言語．

6. 標準様相論理 (1/8)

- 知識と信念に関する様相論理の導入

- 統語規則は，古典命題論理に「知っている」ことを表す演算子 K を加えたもの．
  用語:
  $\Phi = p, q, r, \ldots$: 素命題
  $\wedge, \vee, \neg, \ldots$: 結合演算子
  K: 様相演算子
  統語規則:

$$
\begin{array}{lcl}
(wff) & ::= & any\ member\ of\ \Phi \\
 & | & \neg(wff) \\
 & | & (wff)\ \vee\ (wff) \\
 & | & K(wff)
\end{array}
$$

  K の入れ子構造も許される．

6. 標準様相論理 (2/8)

- (例)

$$K(p \wedge q)$$
$$K(p \wedge K(q))$$

6. 標準様相論理 (3/8)

- 意味論は，統語論よりもさらに扱いにくい．考え方としては，以下のように，エージェントの信念は可能世界の集合により特徴付けられるというもの

- トランプのポーカーをプレイするエージェント A(このエージェントはスペードのエースを持っていると仮定する) を考える．このエージェント A はどのようにして敵が持っているカードを演繹できるだろうか?

- エージェント A が知っていることに基づき，すべての不可能な構成 (相互のカードの組み合わせ) を系統的に削除していく．(例えば，エージェント A がスペードのエースを持っていないという構成はすべて取り除かれる)

6. 標準様相論理 (4/8)

- エージェント A が知っていることを前提として，これらの思考の後に残っているすべての構成が世界，つまり可能な様態である

- エージェント A にとってのすべての可能性のなかで真であることは，そのエージェントにより信じられている．
  例えば，エージェント A のすべての認識的選択肢の中で，そのエージェントはスペードのエースを持っている．

- 二つの利点
  - エージェントの認知構造において中立であり続ける
  - 関係する数学理論が非常によくできている

6. 標準様相論理 (5/8)

- これを形式化するために W を世界の集合，そして $R \subseteq W \times W$ を W 上の二項関係とする．この二項関係によりエージェントがどの世界に関して考慮可能かが特徴付けられる．

- 例えば，もしも $(w, w') \in R$ であり，さらに，もしもそのエージェントが本当に世界 w にいるなら，そのエージェントに関する限り，それは w' にいるかも知れない．

- 論理式の意味は世界に相対的に与えられる．特に，$K\phi$ が世界 w で真であることは，ϕ が $(w, w') \in R$ であるすべての w' において真であることと同値である．

6. 標準様相論理 (6/8)

- この定義における二つの基本性質
  - 以下の公理は妥当 (恒真) である
    $K(\phi \Rightarrow \psi) \Rightarrow (K\phi \Rightarrow K\psi)$
  - もしも ϕ が妥当なら，$K\phi$ は妥当
- よって，エージェントの知識は論理的帰結のもとに閉じている．
  これは論理的全知を意味するが，実は，これは望ましい特性ではない (そんなエージェントは実装不可能だから)．
  (注 論理的帰結 (logical consequence): 論理式 $G_1, \ldots, G_n$ と論理式 H が与えられたとする．$G_1 \wedge \ldots \wedge G_n$ を真にするすべての解釈 I に対して H もまた真になれば，H を $G_1, \ldots, G_n$ の論理的帰結と呼ぶ．詳細は「人工知能 II」の講義ノートの 論理的帰結(← ここをクリック) の節を参照のこと)

6. 標準様相論理 (7/8)

- この論理のもっとも興味深い特性は到達可能性関係 R に関するものである．
  いろいろな制約を課すことで，さまざまな公理を得ることができる．これらの公理はたくさんあるが，もっとも重要なのは以下である．

T	$K\varphi \Rightarrow \varphi$
D	$K\varphi \Rightarrow \neg K\neg\varphi$
4	$K\varphi \Rightarrow KK\varphi$
5	$\neg K\neg\varphi \Rightarrow K\neg K\neg\varphi$

6. 標準様相論理 (8/8)

- 公理 T は知識公理．(「知っていることは真である」ことを意味する)
- 公理 D は無矛盾性公理．(「もしも φ を知っていれば，同時に $\neg\varphi$ を知ることは出来ない」ことを意味する)
- 公理 4 は肯定的内省．(「もしも φ を知っていれば，φ を知っていることを知っている」ことを意味する)
- 公理 5 は否定的内省．(「知らないことに関する認識がある」ことを意味する)
- エージェントを表現するときにどの公理を選ぶかは (ある程度) 選択の余地がある
- KTD45 すべてを採用した論理システムは S5 となる．(理想的知識の論理として選択される)
- T を除いた S5 は弱 S5(weak-S5)，もしくは KD45 という．(理想的信念の論理として選択される)

7. 意図 (Intention)(1/13)

- 知識や信念だけではエージェントを完全に特徴付けることはできない．
- エージェントには態度の合理的バランスが必要．
  - 過度にコミット (over-commitment) しない．
  - 過小にコミット (under-commitment) しない．
- 以下では，エージェントの認知状態を構成する要素がどのようであるべきかに関する整合的な説明を概観する．(Cohen & Levesque の意図の理論)

7.1 意図とは何か?(2/13)

Bratman の理論に基づき，Cohen-Levesque は意図が充足しなければならない 7 つの特性を挙げた．

1. 意図はエージェントに問題をもたらす．エージェントはその問題を達成するための方法を決定しなければならない．
   If I have an intention to ϕ, you would expect me to devote resource to deciding how to bring about ϕ

2. 意図は他の意図を採り入れるための「フィルタ」を提供する．他の意図は競合するものであってはならない．
   If I have an intention to ϕ, you would expect me to adopt an intention ψ such that ϕ and ψ are mutually exclusive.

3. エージェントは自己の意図の成功を観察し，試みが失敗した場合は再度試すことに傾倒する．
   If an agent's first attempt to achieve ϕ fails, then all other things being equall, it will try an alternative plan to achieve ϕ.

7.1 意図とは何か?(3/13)

4. エージェントは自分の意図が可能であると信じている.
   That is, they believe there is at least some way that the intentions could be brought about.(CTL notation: $E\Diamond\phi$*)*

5. エージェントは不可能であると思っていることに関しては意図を持たない.
   It would not be rational of me to adopt an intention to ϕ *if I believe* ϕ *was not possible.(CTL notation:* $A\Box\neg\phi$*)*

6. ある状況において，エージェントは自己の意図をもつ.
   It would not normally be rational of me to believe that I would bring my intentions about; intentions can fail. Moreover, it does not make sense that if I believe ϕ *is inevitable (CTL:* $A\Diamond\phi$*) that I would adopt it as an intention.*

7.1 意図とは何か?(4/13)

7. エージェントは自己の意図のすべての予期される副作用を意図する必要はない.

   If I believe $\varphi \Rightarrow \psi$ and I intend that φ, I do not necessarily intend ψ also. (Intentions are not closed under implication.) This last problem is known as the the side effect or package deal problem. I may believe that going to the dentist involve pain, and I may also intend to go to the dentist – but this does not imply that I intend to suffer pain!

7.1 意図とは何か?(5/13)

- Cohen-Levesque は以下の主要構成子を伴う多様相理論を用いる.

Operator	Meaning
$(\text{Bel}\ i\ \phi)$	agent i believes ϕ
$(\text{Goal}\ i\ \phi)$	agent i has goal of ϕ
$(\text{Happens}\ \alpha)$	action α will happen next
$(\text{Done}\ \alpha)$	action α has just happened