

人工知能A

Topic 1: Introduction (History and agents)

1

自己紹介

- 1982年 早稲田大学 数学専攻修了
- NTT基礎研究所→
→ NTTコミュニケーション科学基礎研究所
- 人工知能の研究 (1982 ~) :
 - 知識表現・機械学習・マルチエージェントシステム
 - インターネットのボランティア (1983 ~)
- 2007年から早稲田大学へ移動 (異動)
- 現在の研究分野 (は変わらず) :
人工知能、特にマルチエージェントシステム

2

オリエンテーション

- 講義の内容
 - 人工知能に関する技術を広く紹介する。
 - 人工知能という分野はとても広く、人それぞれに思い入れもある
 - 特定の技術を深く紹介するものではない。これは卒論時あるいは大学院で
 - 人工知能に関連する技術を紹介し、次のステップに進めるように
 - なるべく最近話題には触れる。しかし、長い歴史でAIの研究の中心は常に変化し、次のブレークスルーも何があるか分からない
 - 人工知能研究の基本、考え方などを習得
 - 知能・知性の解明や機械による実現は2000年以上続く人類の夢であるとともに、他の計算機科学への波及も多い。

3

オリエンテーション

- 講義資料
 - スライド資料などがWaseda Moodleにあり。
 - 将来ノートとして活用できるよう、多めに書いてある。
 - なるべく数日前には公開 (予習できるように)。ただし、少々の変更はあるので、直前に再掲する。
- 教科書はない。が、参考書を後で示す。
- 成績
 - レポート(30%)、
 - テスト (教場試験) (90%)
 - 出席 (10%)

5

参考書

- S. Russell and P. Norvig,
Artificial Intelligence: A Modern Approach,
4th edition (3rd editionもあるので注意)
 - 1000ページ以上。しっかり勉強するならこれがお勧め30000円ぐらい (正式版、ハードカバー、円安で高い)
 - Amazonでpaperbackは8000円ぐらい。Kindleは12000円ぐらい (Kindle版は使い勝手が悪いらしい)
- David L. Poole and Alan K. Mackworth
Artificial Intelligence: Foundations of Computational Agents (2nd edition)
 - 上記の本に比べ新しい。ややコンパクト。
 - 実は… <http://artint.info/>
 - でも本も可能な限り買しましょう。Amazon: 9000円



6

参考書

- 谷口忠大、人工知能概論 第2版 (講談社) 2860円
 - 大学2年向け入門書
 - ホイールダック2号というエージェントが主役
- 趙・樋口、人工知能: AIの基礎から知的探索へ (共立出版)
 - 新しい本。内容は標準的・保守的 (本講義とは方向性が違うかも)
- その他「人工知能」や「機械学習」を含むタイトルだが内容は統計的推定・検定・回帰分析のみの本もある。
- またライブラリの使い方で、技術的中身を記述していない本もある。コーダー向け?
- AIの一分野の教科書と考えて読むこと。このような本の内容が人工知能のすべてではない。

7

8 人工知能の歴史と 考え方

人工知能研究の目的 (当初)

- 認知科学:
 - 人間の認知過程や認知手法を理解するために
 - 例. 画像の認識、記憶、学習、言語習得 ……
- 哲学: 知能・知性に関する研究の起源。古代ギリシャ (ソクラテス) など2000年の歴史
 - 基本的で重要な哲学的な課題を探究する一手段として
 - 例. 意識ってなんだろう……、理解したとは?
- 計算機科学 (Computer Science):
 - 知性の解明: 機械 (計算機) により賢く、柔軟な仕事をさせるために。
 - 例. 言葉の理解、画像から個人を認識する、楽しそうな旅行のプランを立てる、など……当初は気軽に考えた

9

人工知能とは

- 人工知能 = Artificial Intelligence (AI)
- AIって何? Intelligenceって何?
- Intelligence (知能)
 - 一般的な意味 (辞書)
 - 外界の状況の知覚を元に認識・理解を行う
 - 得られた知識・情報を使用して問題を適切に処理し、問題に適応する能力
 - 哲学辞典:
 - 学習能力、抽象思考能力、環境適応能力

→ でも何ができたら学習 (抽象思考、環境適応) 能力があると言えるのだろうか?

10

人工知能とは

- Webで検索すると、AIとは……
 - 知的な人間の思考過程や行為を模倣 (シミュレート) すること
 - 人間の推論方法や推論結果を計算機上で再現すること
 - 計算機科学的手法を用いて人間の認知過程を研究すること
 - 人間の行為を模倣する計算機モデルの研究
 - 人間をまねた機械を作る, etc.
- どれも正しいが、すべては表現できていない。。。
 - 人間の夢の一つ → 多くの思惑・思い入れが…
- 本授業の目的: 計算機科学の観点から、AIに関する基本的な概念、方法論、アルゴリズムを習得する。

11

考えたいこと(1) --- プロセス

- 機械は飛ぶか? ----- Yes?
 - でも動物 (蝙蝠、鳥、昆虫) のように飛ば (羽ばたか) ない
- 機械は泳ぐか? ----- No? (船は帆走・航行 sail?)
- 機械は走るか? ----- Yes?
- 機械は駆けるか? ----- No?
 - (2足歩行ロボットならYes?)
- では機械が「考える」とは? ----- Yes or No ??
 - 生物 (ここでは人間) を模倣すべきか?
 - Deep BlueやAlphaGoは確かに勝ったが知性があると思うか?
 - 哲学者サールの反論 (中国語の部屋)

12

考えたいこと (2) --- 評価基準

- 何ができたら知能や知的処理が実現できたと言えるか?
- 生物 (人間) の判断・決定を基準
 - アルゴリズム的ではなさそうな「発想」や「思いつき」もある
 - 他方、誤りもある。これもまねるか?
 - 文化 (国、民族、宗教) に依存する。人として知性・知能の共通部分を取り出さないと普遍的な評価はできない (認知科学)
- 可能な限り広く・深く推論 (合理性)
 - 既知の知識・データ、環境からの情報を元に、可能な限りの処理を行い、結果を得る
 - 評価基準は明確
 - できることはできなくてはいけない。求められるものは求められなくてはならない。
 - 必ずしも人間的ではない (人はときどき非合理的)

13

AI研究の特徴づけ – 各種AIの教科書の定義から –

前々ページの
「考える」を参
照
↓

**人間の行動
を基準**

**人間の内部思考
方法を基準**

<u>人間のように行動する</u> 人間が行う場合には知能を 必要とする機能を達成する 機械を創る技術	<u>合理的に行動する</u> 知的行動の自動化に関す る計算機科学の一分野
<u>人間のように考える</u> 計算機が考えるようにする。 真の意味で心を持った機械 を造る刺激的で新しい試み	<u>合理的に考える</u> 計算機モデルを用いた心 の機能の研究 認識、推論、行為を可能 にする計算の研究

評価基準 → **人間の決定を基準** (誤りもまねるかも) **合理性** (可能な範囲で**最善の解を求**
める理想的知能像) を基準

14

AI研究の特徴づけ – 各種AIの教科書の定義から –

前々ページの
「考える」を参
照
↓

**人間の行動
を基準**

**人間の内部思考
方法を基準**

<u>人間のように行動する</u> 「チューリングテスト」 チューリングテストに合格 する (cf. Eliza in emacs)	<u>合理的に行動する</u> 「合理的エージェント」 知識や信念から可能な限 りの合理的結論をえる
<u>人間のように考える</u> 「認知科学的アプローチ」 人間の記憶や思考の過程を 考える	<u>合理的に考える</u> 「形式的な推論や思考の 規則」論理学

評価基準 → **人間の決定を基準** (人間の誤りもまねるかも) **合理性** (可能な範囲で**最善の解を求**
める理想的知能像) を基準

15

合理的に行動する

- 合理的エージェントアプローチ：
 - 保持している情報（知識、時間、使用できる資源など）に基づき、常に最善の決定を下す。この意味で知的。
 - 調べれば分かるけど、知らなかったことにしよう（＝非合理的）
 - 人間の知性と機械（計算機）にできる知性を分離した
 - 情報（知識）が不完全で計算資源も限りあるだろう
 - （可能な限り、限りある）合理性 → 限定合理性
 - 「一部は論理的推論。でもそれがすべてではない」という（広い）アプローチ（人間の思考過程を真似ても真似なくてもよい）
 - たとえば焼けどのとき手を引っ込める行為（考えてはいない。反射的行為。このような行為の実現方法は？）
 - 人間の行動を判断基準にするよりは明確
 - 計算機科学（つまり本授業）の中心的枠組み

16

人工知能とは（現代の定義）

- 人工知能とは、知的に行動する（プログラムとしての）合理的エージェント (computational agents) の実現とその行動の解析。ここでエージェントとは与えられた環境で自律的に動作する主体とする。
- ここで知的とは：
 - 与えられたゴールと現在の状態に応じて適切な行動を自らとる（自律性）
 - 環境やゴールの変化に柔軟に対応できる（適応）
 - 経験や例から学ぶことができる（学習）
 - 認知あるいは計算資源の限界を考慮して、その範囲内で適切な判断を下す（合理性）

17

参考（近年のAIの定義例。研究の進展とともに）

Purpose of Artificial Intelligence

- The modern definition of artificial intelligence is "the study and design of intelligent agents," where an intelligent agent is:
 - A system that perceives its environment and takes actions which maximizes its chances of success.
 - "Success" is ambiguous.
 - ➔ A system that maximize expected utility, where the specific utility information (objective) is supplied by the human designers of the system.
 - Utility does not fix and cannot be supplied by the designer in many cases.
 - ➔ A system that may be uncertain about the true objectives of the humans on whose behalf it operates, and therefore, must learn what to maximize and must function appropriately even while uncertain about the objective.

18

参考（近年のAIの定義例。時代というか進展とともに）

Purpose of Artificial Intelligence

- An agent acts intelligently if:
 - Its actions are appropriate for its goals and circumstances
 - It is flexible to changing environments and goals
 - It learns from experience
 - It makes appropriate choices given perceptual and computational limitations

19

ただし

- 四枠どれもAIの研究であることに注意。AIは広く知性を科学的に解明したいという目的がある。
- 人間も…
 - あるときは合理的、別のときは非合理的に。
あるときは論理的、あるときは感覚的（直感的）に
 - それぞれの枠のアプローチがきれいに分かれている訳ではない。
 - 青枠の「人間のように考える」は機械で意識とか人間性 (humanity) を実現したい、科学的に解明したい、という究極の目標がある
- 強いAIと弱いAIという分類（哲学者：サール, Searle）
 - 強：心（意識や自我）を実現する（アルゴリズムで可能という立場）
 - サールは脳は機械（のようなもの）と考えているが、強いAIは否定的。
 - 弱：機械は心・意識は持たない。そこまで追求しない（できない）
 - 現状では弱いAIの一部があるだけ。

20

21 歴史

人工知能の歴史 — 人工知能以前

- 学問としての基本的な問いかけは2000年以上前
 - 哲学の分野で議論
- プログラムのできる電子的コンピュータの誕生
 - Corad Zuseがリレー式の2進法計算機Z3を完成 (1941)
- Maurice Wilkesらが最初のプログラム格納式電子計算機EDSACを完成(1949)。ただし真空管式。
- McCulloch & Pitts 神経細胞の数理的モデル (1943)
 - ニューラルネットの始まり
- Chessのプログラム (Claude Shannon 1950, Alan Turing 1953)
 - Shannon は情報理論、Turing はチューリングマシン
- ニューラルネット (Marvin Minsky とDean Edmonds、1951)
 - パーセプトロン。しかしこのあとしばらくブランクがある

22

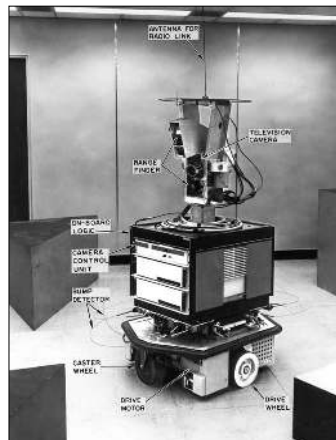
人工知能の歴史 — 初期 (1950-60年代)

- Artificial Intelligence (人工知能) という言葉は、Dartmouthで開かれたワークショップ (1956) にJohn McCarthyによって初めて使われた。
- 50年代～60年代後半：**基本的原理の創出の時代**
 - 一般問題解決器 (Herbert Simon)
 - 問題をサブゴールの列に分解し、探索により答えを探す
 - 定理証明 (Theorem proving)
 - ゲーム理論
 - Formal calculus (論理学、数学基礎論)

23

人工知能の歴史 — 初期 (1950-60年代)

- John McCarthyはMIT でLISP (list processor) 言語を作る (1958)
 - AI言語の王者と言われていた。その後Prologも利用。
 - 現在はJavaなどが多く利用される。
- MinskyがStanford Research Institute (SRI) を創設。
- Shakey ロボット ('69-71ごろ)
 - 最初の自走式で、当時の人工知能の技術で制御されたロボット
 - LISP。視覚 (カメラ) あり



24

人工知能の歴史 (1960年代後半—80年代前半)

- **知識に基づくシステム (Knowledge-is-Power period)**
- エキスパートシステム (Expert systems) : 特定の問題領域に絞込みAIを適用し実用域に達しよう
 - 専門知識と推論を計算機上に表現 (ルール、フレーム表現)
 - 医療診断 — 血液感染症 (MYCIN), 分子構造解析 (Dendral)
- ルール (production rule) による知識表現が大多数
 - 例) If: エンジンがかからず and ヘッドライトが暗い
Then: バッテリーが不良である (0.8)
- その他: フレーム、セマンティックネットワーク、論理式
- 知識工学
 - 知識エンジニア (Knowledge engineer) という職があった
 - 専門家にインタビューして知識を計算機上に表現する

25

人工知能の歴史 (1980年代)

- **AIが産業となった時代**
 - 最初に成功した商用エキスパートシステム R1
 - DECによる計算機システム構成助言システム（年間4千万ドルの経費節減）
 - 研究プロジェクト（主目的は並列マシン？）
 - 日本：第五世代プロジェクト（1982 - 1992）、
 - 米国：MCC（Microelectronics and Computer Technology Corp.）
 - LISP Machine（LISP専用マシン）
 - Symbolics社、XeroX社、Texas Instruments社などから発売
 - エキスパートシステム構築用のツール
 - Intellicorp社、Tecknowledge社、Carnegie Group社、Inference社など。
 - 1988年には20億ドルの市場規模

26

人工知能の歴史 (1980年後半～90年代)

- **Neural Network (NN)の復活、遺伝アルゴリズム、Bio-inspiredアルゴリズム研究の出現 (85年～)**
 - 一方で単にルールを蓄積するだけでは限界
 - 専門家からの知識獲得も困難（知識獲得ボトルネック）
 - 新しい打開手段としての期待。
- **実世界事例をベースとする研究 (90年代ごろから)**
 - 応用範囲の拡大、特にNNなどの機械学習の一部が有用
 - 音声処理、画像処理、
 - システム間や実環境（実世界）とのインタラクション
 - 自律エージェント（インタラクションを介して自律判断）
 - マルチエージェントへの着目（⇐ インターネットの普及）

27

人工知能の歴史 (2000年～)

- **機械学習、特に統計に基づく学習が普及**
 - インターネットの普及やセンサネットワーク (IoT)などによる大量のデータの存在 --- ビックデータ
 - ビックデータに基づく回帰分析、クラスタリング、NN
 - 2015年ぐらいまで日本では人工知能・機械学習という言葉は禁句
 - NN (CNN, Deep Learning, Deep Q-Networkなど) や各種機械学習のパッケージが普及
 - 多くの試みと実用化が進む
 - 実世界やネットワークを介したエージェント間の相互作用と相互調整のための社会学習
 - Social Networking Services、複雑系、複雑ネットワーク
 - エージェントの競争や協調（マルチエージェントシステム研究）

28

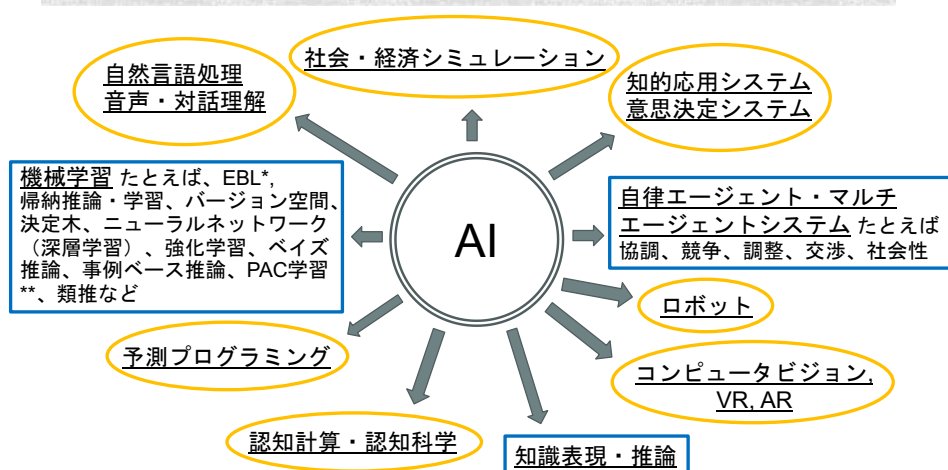
人工知能の歴史 (まとめ)

- 人工知能は独立した分野であり、細分化と同時に他の分野とのかかわりが強い。たとえば、
 - 技術の細分化：ニューラルネットワーク、強化学習、知識表現、論理、遺伝アルゴリズム、人工生命、定理証明、数式処理、ファジーなどは切り離されて独立して研究をされている
 - 分野の細分化：自然言語処理、画像理解、音声認識・理解、知能ロボット、仮想現実感、そのほか計算機・機械の知的な処理・制御も人の知的行動の一部を切り出したもの（だが人工知能とは明示しないことも）
 - 周辺領域との融合：人文科学分野（社会科学、経済学、行動学）人間知能を学ぶという意味で、これらの知見は広い観点から人工知能の1分野でもあり、逆に計算機科学の成果が流入もしている
- 人工知能の成果はアルゴリズム・ツールとして切り出され、計算機科学の各種方面で活用されている。

29

AIの研究分野

オレンジはAIの応用・適用分野
青は理論、基礎・要素技術。



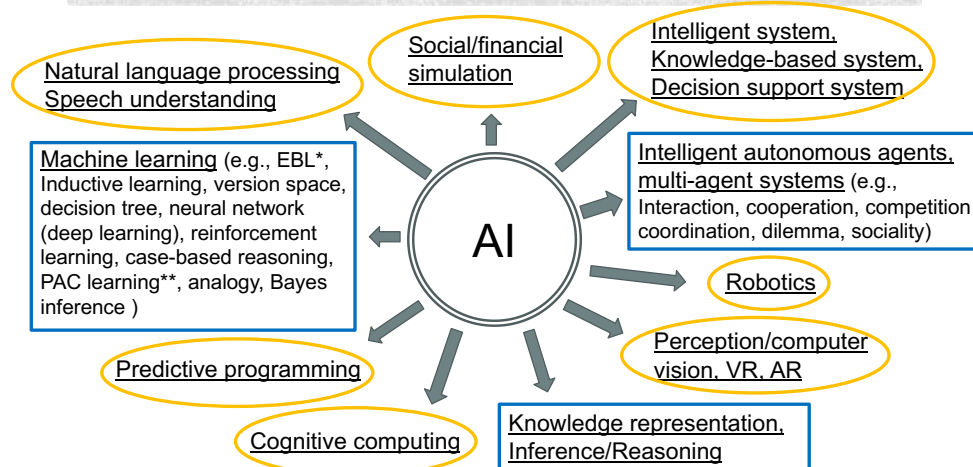
他にプランニング、不確実性推論などもある。

*: Explanation-Based Learning, **Probably Approximately Correct learning

30

AI research areas

Orange circles: application areas.
Blue rectangles: theory and basic technology that may be used in AI applications.



Other AI research topics: planning, optimization, negotiation, and more.

*: Explanation-Based Learning, **Probably Approximately Correct learning

31

レポート

- Waseda Moodleにレポートがあります。
- 締切はWaseda Moodleの設定に従ってください。
- ここでは自分の考えや意見を述べてください。正解はありません。考えて、その理由を書いてみてください

32

33 エージェント

Act (agere) ⇒ Agent (行動する（者）、動作主)

エージェント



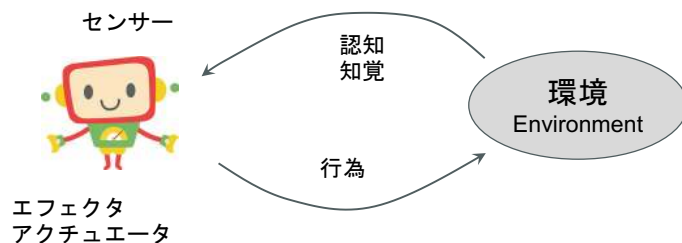
34

人工知能の目的（一部再掲）

- Artificial Intelligence とは知的かつ合理的に行動する計算主体としてのエージェント (computational agents) のあるべき性質やその原理、機能を解明し、これを実現（実装）すること
- エージェントは、環境で動作するもの。エージェントが知的とは（たとえば）：
 - 与えられたゴールや周囲の状況に対して適切に行動する
 - 環境やゴールの 変化に柔軟に対応できる
 - 経験から学ぶ
 - 認知の限界や計算資源の限界に応じて適切に判断・行動する
 - 「環境」は実世界、仮想世界であり、エージェントの周辺にある状態・状況全般。行動を決定し、また行動により変化しうる。

35

エージェント (Agent)



- 定義: エージェントとは、センサーを通して環境を認知し、エフェクタにより環境に対して自律的に合理的な行為を行う実体 (entity)
- 実体: 物理的実体のあるもの (機械、ロボット、人間など) だけでなく仮想的なソフトウェア・プログラムも
 - ソフトウェアエージェント: 基本的にはソフトウェアのみで構成されたエージェント

37

エージェントの例: 人

- センサ:
 - 目 (視覚), 耳 (聴覚), 皮膚 (触覚), 舌 (味覚), 鼻 (嗅覚), 神経 (反射、伝達)
- 認知:
 - 基本レベルでは神経細胞を流れる電気信号
 - 耳や目などで前処理が行われ- 映像内の物体認識 (場所、色、形など)、音響認知 (高低、強さ、方向など) ……
- エフェクター: 手、足、目、声帯、舌など……
- 行為: 手を上げる、向きを変える、歩く、走る、持つ、話すなど……
- エージェントのモデル化では、合理的行動を仮定。

38

(人工) エージェント

- ロボット
 - 認知：
 - 超音波（障害物）、接触、距離計、カメラなど
 - 行為：
 - 移動、方向転換 など
- ソフトウェアエージェント
 - プログラム
 - 認知：
 - プログラムへの入力 （数値、文字、音、イメージなど）
 - 行為：
 - プログラムからの出力

39

エージェント：他の例

- 組織（の代理, delegates）
 - Microsoft, Apple, Government of Japan, Waseda University, ...
 - 組織のdelegatesとして
- 人間の専門家
 - 教師、医者、トレーダー、エンジニア、旅行代理店、農家、etc...
 - 合理的行動や判断を模倣
- コンピュータ/デバイス
 - サーモスタット、ユーザインタフェース（例）、飛行機の制御プログラム、ネットワークコントローラ、ゲーム、ナビゲータ、車の制御、医療診断システム、Google Car, Mars rover,...
- 生き物
 - 犬、魚、鳥、昆虫、細菌、ウイルス（物理学でよくある）...

40

AI研究のゴール

- 科学的（学術的）な目標：人工的なシステムに知的な行動を実現させるための原理を解明し、それを理解すること
- 自然界のエージェントの知性の解明と活用。知的な人工システムの仕組みを実現するための解明。
- 知的システムの実現のための理論的・形式的なモデルを作り、それを実験・シミュレーションで確かめること
- 知識や知能が必要なタスクを計算機システム上にデザイン、実装し、それを実験的に評価すること
- 工学的な目標：知的な人工システムをデザインし、実装すること

41

エージェントの合理性

- 合理性：わかっていることから予測できる最大限の成功を考えると（前出）
 - 「信号が赤に変わりそうだったが、車がないので渡ろうとした。そうしたら、そこに鉄骨が落ちてきた」⇒知り得ないので不合理とはいえない。
- 理想合理的エージェントとは、
 - 可能な知覚列（知覚の履歴）に対し、エージェントが持つ知識（情報）を使って最大の性能（効果）を引き出すことが可能な動作を選択するもの
- 合理性には必要な知識を取得するという行為も含む
 - 調べられるのに「知らないことにしよう」は非合理（人間社会と同じ）。
- 最大の性能（効果）→ 何らかの基準が必要（performance measure）
 - 性能の基準：ゴールが達成されたか、速度、使用した計算資源と処理能力、環境への影響など。それぞれの問題に依存。

42

自律性と自律エージェント

- システムが自律的であるとは、その行為（行動、タスク）を環境や自らの経験から決定できること。
- もし外部から予め決められた命令のみに従って行為を決定する場合、自律的とは言わない。
- 自律エージェントのためにさらに必要なもの
 - 経験に基づいた学習，知覚
 - ただし経験を持たない段階でも、ある程度適切な行動を行うための事前知識は必要（生物の本能のようなもの）
- 本能に基づいて盲目的に行動するだけでなく、環境の変化に対して適応する。

43

エージェント：注意

- ただし必ずしも「知的」を前面に出さない研究もある。特にシミュレーション。たとえば、
 - 人流シミュレーション（エージェントは人）
 - 株式マーケットのシミュレーション
 - 鳥や昆虫の行動
- ただしそれでも自律性や限定的な合理性はありそう。
- 個々が単純な動作のみのエージェントであっても全体として見ると何らかの知的な行動や現象が垣間見られることがある。（創発: emergence）
 - 渡り鳥の編隊の形（V字型）
 - イワシの集団での群れ、など
 - どちらも自律的な最適判断によるものをベースとしている

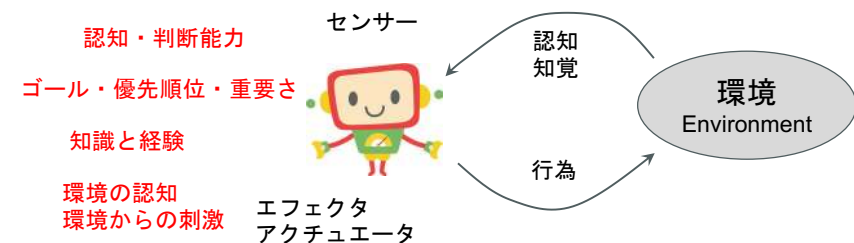
44

エージェント：注意

- もともと計算機科学の分野でエージェントという言葉は人工知能の分野で使われていたが（1988年頃から）、その後徐々に（1992-3年ごろから）、
「自律した（独立した）計算主体・プログラム」
として計算機科学の多くの分野で広く使われるようになった。

45

エージェント実現のプログラム



- エージェントが知的な行動をするためには、予め（あるいは後付けで）与えるべきものがある。
- これをプログラムとして実現しないといけない

47

エージェントが備えるべき能力(1)

- 認知・判断能力
 - 可能な行動（行為）のモデル（プログラム？）
 - 環境の認知と環境の内部表現（モデル）
 - 環境に自分の行為が与える作用・影響（モデルの変化）⇒ 課題：計算内に環境を表現（モデル化）する必要がある
- ゴール・優先順位・重要性
 - 希望する状態とその重要性
 - 上記2つを合わせて、希望状態に達する可能性のある行為を選択する（行為の優先順位も決まる）⇒ 課題：希望状態を与える or 学習する必要がある
課題：希望状態へ達する行為の計画（プランニング）

48

エージェントが備えるべき能力(2)

- 知識と経験
 - 行為と結果の知識（単純に影響を見るだけではなく、一連の行為を遂行して望む結果が得られるとき）
 - 過去の行為（の組合せ）と環境の変化を記憶。これにより似たような状況では直ちに対応できる。⇒ 課題：「選択可能な行為と想定される結果」（行動モデル）や「望ましい結果の記録」が必要（メモリ・経験）。
- 環境の観測・環境からの刺激
 - 行為の結果を確認する。
 - 環境からの刺激（たとえば想定外の結果、回避すべき状態など）を認識。⇒ 課題：環境の観測と内部モデルとの照合が必要

49

- これらをどうやって実現するか。。。。

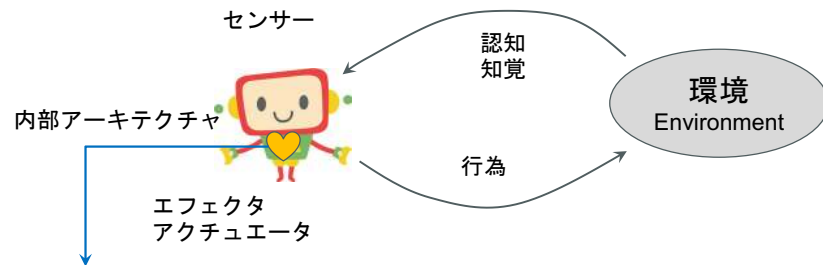
50

このマークは少し進んだ内容を示す。飛ばすこともあります。

51 付録

これまで考えられてきたエージェントのデザイン（アーキテクチャ）
付録は時間があればやりますが、基本的には自習です。
付録のスライドの左上には三角のマークがあります。

エージェント (Agent)



- 知的な制御プログラムに必要な構造や機能は？
 - エージェントの知覚（の列）から動作へのマッピングを実現するプログラムを設計すること
 - プログラムは計算機で動作する（有限の資源）

52

エージェントアーキテクチャ

Russell and Norvigの本からの例

- 単純反射エージェント (Simple reflex agents)**
 - 知覚と行為の対応表 (lookup table) や if-then 形式ルール (プロダクションルールと呼ぶ) などにより行動を選択。
 - 状態を持たず、過去の環境、知覚、行動などを記録するメモリはない。
- 記憶を持つ反射エージェント (Agents with memory)**
 - 1. に加え内部メモリを持ち、過去の環境や知覚列などを記憶でき、それに基づいて行動を選択。
- ゴール主導エージェント (Agents with goals)**
 - 2. に加え、ゴールとなる環境の状態を持つ。
 - その状態を考慮に入れて（通常はゴールの達成あるいは近づくために）行動を選択。
- 効用に基づくエージェント (Utility-based agents)**
 - 2. に加え、各状態に効用 (utility) と呼ばれるエージェントの望ましさの度合いを決める関数があり、その度合いを高めるように行動を決定。
 - ゴールの達成だけでなく、たとえば効率や安全性なども考慮に入れられる。

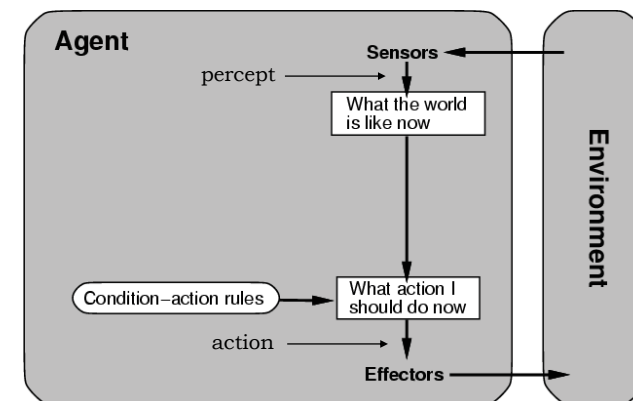
54

単純反射エージェント

- 知覚した内容とそれに対する行動の対応表（ルール）
 - 知覚 (percept) とはセンサを通して得た入力（センサの出力）
- 動物の反射的な単純行動に対応。
- 問題点
 - 非常に大きな対応表となる（たとえばチェスは、おおよそ 10^{120} の状態がある）。
 - 知覚の対応だけで、次の新状態に対しての行動は取れない。
 - たとえば、知覚→行動→新たな状態→行動とはいかない（新しい状態を知覚をしないと次の行動が起こせない）。
 - 環境の変化に対しては固定的（テーブルは固定）
 - 学習してテーブルを書き換えるには情報が少ない
 - ループが起こりうる（古い状態を覚えるメモリがない）。

55

単純反射エージェント



56

単純反射エージェントの基本動作

Function simple-reflex-agent (*percept*) returns *action*
static: *table* or *rules*; // a set of condition-action rules

```
state ← Interpret-Input (percept);  
rule ← Rule-Match (state, rules);  
action ← Rule-Action-Part (rule);  
returns action;
```

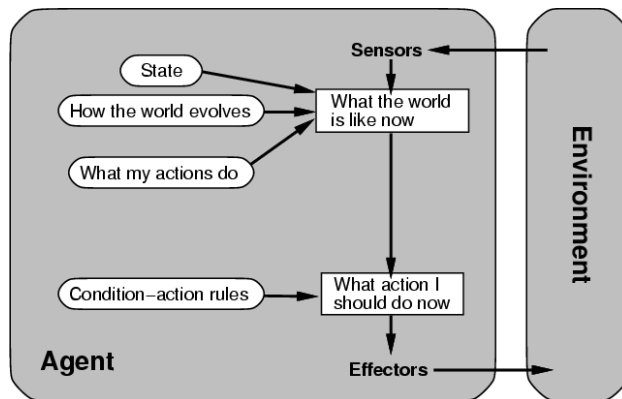
57

記憶を持つ反射エージェント

- エージェント内に現在の環境の**(内部) モデル**や過去に知覚した内容などを記憶する。
 - 内部モデルは「環境はこうである」とエージェントが考える環境
 - なお時間的推移を考えれば、記憶された状態には**過去の履歴**（過去に知覚した内容や環境）も含まれる。現在は過去の積み重ね。
- 外部から与えるとよいもの：下記のような記述
 1. **エージェントの行為が環境をどのように変えるか**
 2. **エージェントの行為とは独立して環境は変化するか**
 - 1. は状態遷移のモデル。たとえば地図を記憶し、学内の道案内を考えてみよ
 - 2. は確率的なモデル（何時、どのように変化するか分からないので）
- これにより行為によってもたらされる環境の変化に内部のモデルを追従させて、少し先の行為まで推論できる。

58

記憶を持つ反射エージェントのアーキテクチャ



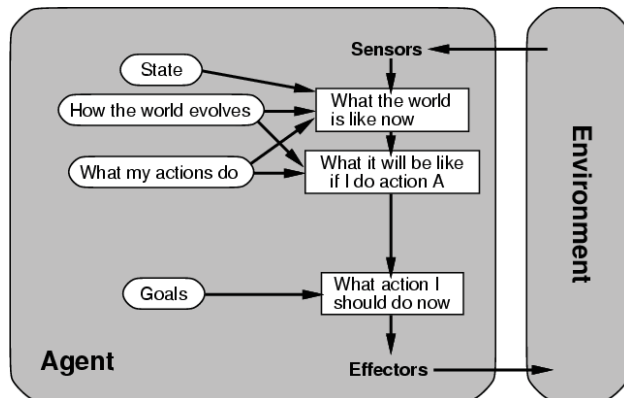
59

ゴール主導エージェント

- **ゴール**を持ち、その達成のために行為を選択する。
 - ゴールはエージェントが望む状態を表す。
- 単にメモリにそのときの状態を持つだけではなく、ゴールを持つことで、どの状態に変わることがより適切かを決定できる。
 - これによりある状態で可能な行為が複数あり（これまでは基本的に状態に対し単一行為が対応）、そのうち適切なものを選択する。
 - 授業に遅刻しそうなのにラーメンは食べに行かない（ゴールによるが）
 - **反射的** (reactive) に対し**熟考的** (Deliberative)とも言われる。
- さらに内部モデルに基づき数手先の状態を考え、ゴールに達する道筋を作る。
 - このように考えて作られた「ゴールに向けた行為の列」を**プラン**と呼び、それを生成する過程を**プランニング**と呼ぶ。

60

ゴール主導エージェント



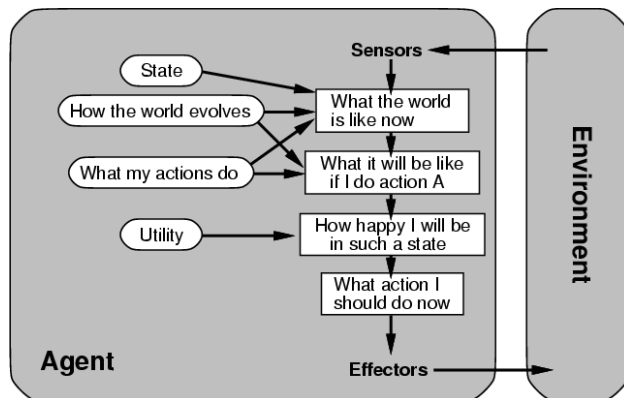
61

効用に基づくエージェント

- 仮にゴールが与えられたとしても、それを達成する方法が複数ある場合。または、単一指標でない場合。
 - たとえば、早く達成したい（迅速性）、確実に達成したい（成功率）、安全に達成したい（安全性）など
 - 安全性と迅速性など時には同時には達成できないものもある。
- 達成するにあたり、何が大切なのかを基準に行為を選択しながらゴールを達成する。
 - 効用関数 (utility function)**
 $U: \text{State} \rightarrow \text{数値}$
 を用意し、エージェントにとって何が望ましいかを数値で表す。
 - ゴールに近づくことも効用関数に含めることはできるので、ゴール主導を包含するエージェントと言える。

62

効用に基づくエージェント



63

エージェントが存在する環境の性質

- 環境の性質にエージェントの能力は左右される
- アクセス可能性 (accessible/inaccessible, observable) または (不) 可視性 (visible/invisible, partially observable)
 - エージェントが行為の決定に必要な情報をセンサにより完全に取得できる（観測可能）であるとき **アクセス可能** という。
 - 逆に、必要な情報に部分的にアクセスできないときもある。
- 決定的・非決定的 (Deterministic/Nondeterministic)
 - 環境の次の状態が、現在の環境の状態とエージェントの行為によって完全に決まるとき、環境は決定的であるという。
 - アクセス可能で決定的な環境では、エージェントは不確実性を考慮しなくてもよい。
 - 一方、実際には非決定的な環境でもエージェントが取得できない環境の情報があれば、決定的に見えるかも（決定的に見えても、非決定的な環境として扱う）。

64

エージェントが存在する環境の性質

- 静的環境・動的環境 (Static/Dynamic)
 - 静的環境：エージェントが行為を起こさなければ変化しない環境
 - 行為を選択している過程で環境は変わらないので扱いやすい。
 - 動的な環境では、環境からの情報を取得せずに熟考をすると、その結果が無意味になることもある。
- 離散的・連続的 (Discrete/Continuous)
 - 知覚した内容や行為が明確に離散値で表現できる場合、離散的であるといい、連続値で表現するときは連続的という。たとえば、チェスは離散的、自動ドライビングシステムでは、速度やハンドルの角度などが連続量となる。
 - 離散値と連続値の融合もある
 - 連続値でも計算機の内部では、数値を細かく分けて離散的に表わすことが多い