

2018/3/30
三菱電機先端総合技術研究所

人工知能の将来展望 ～真のAIの実現にむけて～

玉川大学脳科学研究所／工学部
大森隆司

自己紹介



- 大森隆司(おおもり たかし)
 - 1980 東京大学 大学院工学系研究科修了, 81年博士中退
 - 1981 東京大学 工学部助手
 - 1987 東京農工大学 工学部講師, 88年 助教授, 98年 教授
 - 2000年 北海道大学 大学院工学研究科 教授
 - 2006年 玉川大学 工学部／脳科学研究所 教授
 - 2015-2016年 日本認知科学会 会長
 - 現在 玉川大学学術研究所所長
日本神経回路学会 会長(2017 -)
- 専門: 認知科学, 人工知能, 脳科学

興味: ヒトという知能を持つ生物



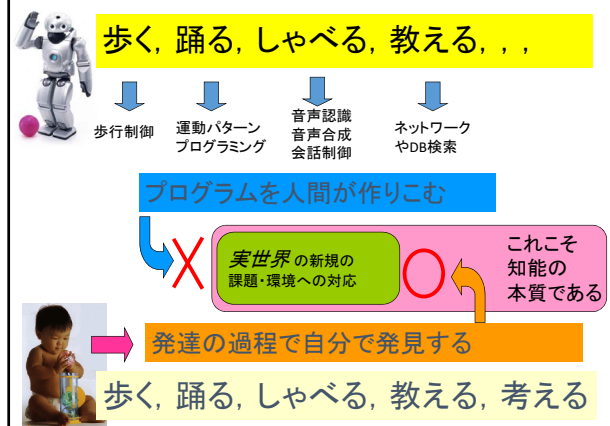
石器を使う親を対ごしのそきこむ
野生チンパンジーの子ども

人間の知的な活動

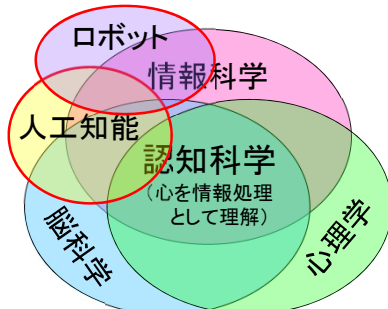
- 道具をつかう
- 言葉をはなす
- 多様な学習をする
- 社会的な活動をする

個別の能力の背後にある
知の基盤となる情報処理
能力とは何なのか?

ヒトの観察だけでは限界がある



認知科学 = こころの情報科学



本日のストーリー

- 人工知能の歴史
 - 記号計算主義
 - ニューラルネット
 - Deep Learningの位置づけ
 - Beyond Deep Learning
- 価値ベースの認知アーキテクチャ
 - 感情: 価値計算システム仮説
 - 推論: 直観的推論と論理的推論と価値
 - 記憶: エピソード記憶+価値 → 意思決定
 - 統合アーキテクチャ?
- まとめ

過去
↓
現在
↓
未来

人工知能の歴史(黎明期)

- コンピュータ前：推論の哲学, 数理論理学, 不完全定理,
チューリング機械 → 推論は機械化できる
- コンピュータの発明 → 人工知能という学問の確立(1956)
- サイバネティクス: 神経ネットワーク, Perceptronの発明 → 学習の発見
- ゲームAI: チェッカー, チェス → ベンチマークテスト化
- 記号的推論の実現, 数学定理の証明(1955)
- 「AI」の誕生, ダートマス会議(1956)
- 第一次ブーム(1956-1974), GPS(1957),
英語対話→ELIZA(人工無脳),
積み木世界→SHRDLU(1971, 英語で対話+計画立案+実効)
- 1974～: ニューラルネットの冬の時代, 論理プログラミング・常識推論

限界：計算パワー, 組み合わせ爆発, 知識の不足, フレーム問題

思想：記号計算主義

- 「我々の思考, 特に問題解決に至る有効な思考(アルゴリズム)は記号列の操作として表現される」という考え方
- 1950-1980年代に主流を占めた知能の考え方
- 全てのアルゴリズムは記号列として表現でき, その表現と操作の方法は何通りかが有名である.
 - ポストマシン, チューリングマシン,
チャーチの帰納的関数 等は互換である

推論の記号操作(形式操作)

- 推論： A と $(A \supset B) \rightarrow B$
 - この実現には「 A が共通」ということの検出が必要
 - 一旦 A が発見 → 実世界とは無関係に B が成立
- 公理系の無矛盾性と完全性
 - 公理と推論規則からは矛盾 $A \wedge \neg A$ が生じない
 - 対象世界で成立する全ての命題が導ける

ゲーデルの定理：どのような公理系からも導き出せない定理が存在するという証明。
その証明には, 有名なゲーデルナンバールが出てくる。

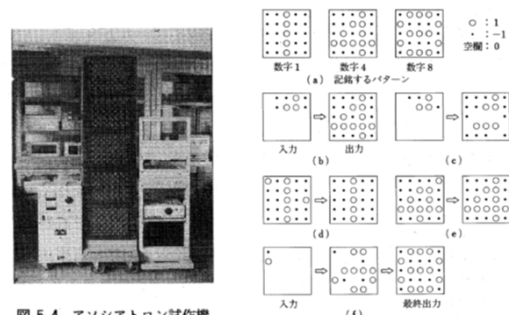
記号計算主義の破綻

- 知識の不足による問題
 - 常識的なことが表現できない
 - 対策: 知識をもっと増やせばいい
- 知識の増加による問題
 - 組合せ爆発による探索時間の爆発
 - フレーム問題: 関係のある知識はどれ?
- 記号であることによる問題
 - Symbol Grounding Problem
記号の組合せは何を意味している?
 - 知識はだれが与えるのか

知能研究(ニューラルネット系)

- 知能の研究
 - 広義の認知科学(心理, 情報, 脳)
 - 本格的な始まり: 1950年代, 計算機の登場以降
- 何回かの工学的ブーム
 - 第一次ニューラルネット(1960年代～1972)
 - 第一次人工知能(1970年代～1980年代前半, 記号処理)
 - 第二次ニューラルネット(1980年代前半～1990頃)
 - お休み, インターネット → 社会基盤の変化
(Google, Amazon, Microsoftなどの情報企業)
 - 第三次ニューラルネット(2010頃～)
→ 第二次人工知能 (～現在)

アソシアトロン (1972中野)



本日の話題: 福島邦彦

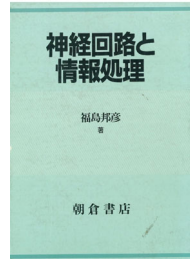


東京工科大学 メディア学部 教授

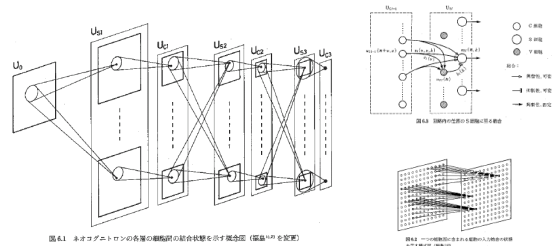
1958年京都大学工学部電子工学科卒。
NHKの放送科学基礎研究所、放送技術研究所などを経て、
1989年大阪大学基礎工学部生物工学科・教授。
1999年電気通信大学電気通信学部情報通信工学科・教授。
2001年より東京工科大学教授、工学博士。

脳における情報処理機構の解明のために、神経回路モデルを仲介とする合成的手法を用いて研究。特に、視覚系における情報処理や記憶・学習・自己組織化の機構の神経回路モデルの構成、ニューロコンピュータなどに興味。「ネオコグニトロン」、「選択的注意機構のモデル」などを提唱。最近では、active vision などの神経回路モデル構築に力を入れている。

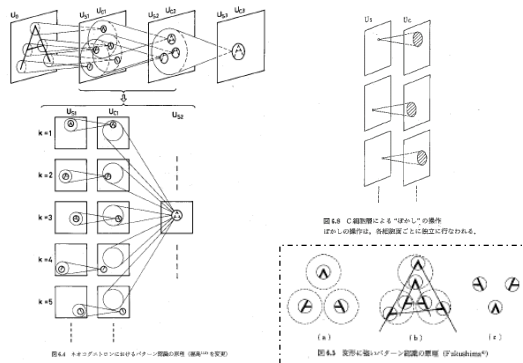
科学技術庁長官賞(ネオコグニトロンの研究)、電子情報通信学会業績賞、論文賞、IEEE Neural Networks Pioneer Award



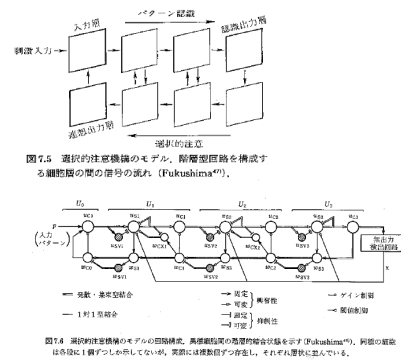
ネオコグニトロン



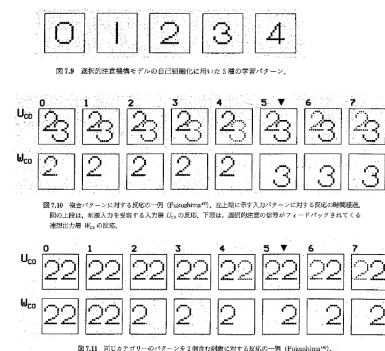
位置不変性の獲得 by C細胞



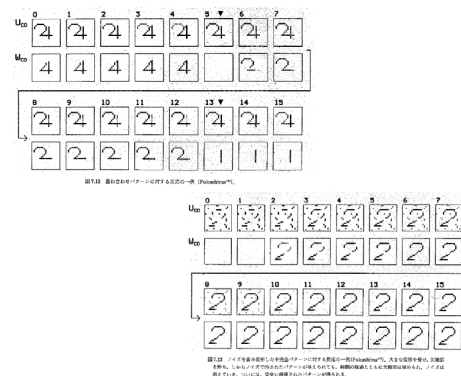
選択的注意のモデル



基本的な能力



多義的な図形や補完能力



ハイブリッド画像理解システム

「選択的注意のモデル」

- ニューラルネット
 - フレキシブルな画像想起
 - 学習の利点はあるが、
 - 思い通りに制御できない
 - 想起の精度がいま一つ
- 普通の記号的プログラム
 - 制御はプログラムで自由
 - 精度は十分高いが、
 - 人間が対象知識の記述
 - 柔軟性が不足する

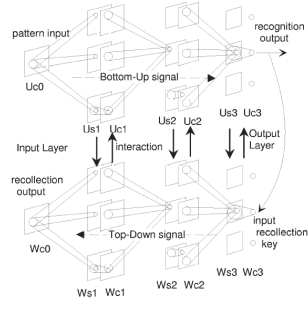
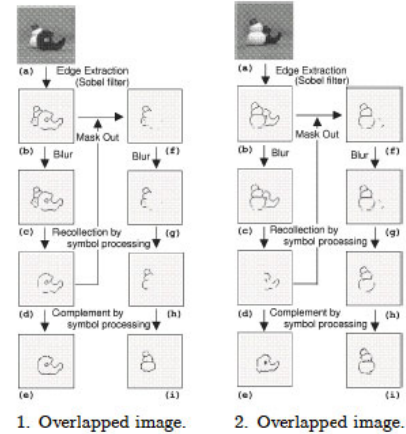


Fig.1: Structure of Selective Attention Model.

動作例 (3) 複数 物体



第二次ニューラルネットブーム

- 1980年台～1990くらい
 - バックプロパゲーション：階層ネットワークの学習
 - 自己組織化マップ：大量データの可視化
 - 階層 3～5層くらい、細胞数 数万程度
 - 学習データの収集に苦労
 - 身体性の重視

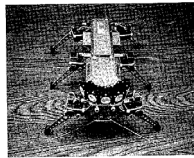


図 7.9 MIT の人工知能研究所で製作された六脚歩行ロボット Ghengis.

冬の時代ふたたび

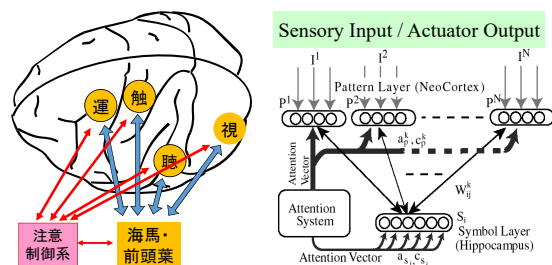
- ニューラルネットは限界 → 企業の撤退
 - 企業研究者 → 大学 or 分野転換

次に起きたのはインターネットブーム

- WEB → コンピューターは部品
- インターネットビジネス
- 巨大情報企業の出現
 - Google, Amazon, MicroSoft etc.
 - 政府支援を介さない、応用に直結した閉じた開発
 - 研究者の集積 → 強力な開発力
 - 他の企業へのインパクト

PATONモデル

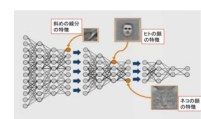
2000



- 概念記憶の表現システムとしての「連合野－海馬 記憶ネットワーク」
- 脳内の計算手続きの表現としての「内的な注意の系列」

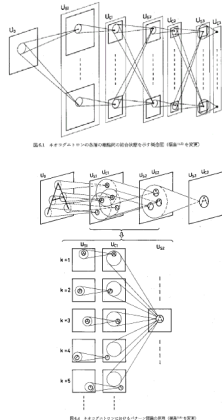
第三次ニューラルネットブーム

- 2011年くらい～現在
 - Deep Learning：バックプロパゲーションの延長
 - 大量のデータを学習（数十万サンプルは当たり前）
 - 各種の分類タスクで驚異的な成績 → インパクト
 - なぜ高性能なのか？、現時点では理論がない
 - 基本的に「統計的な関数近似」
- 階層ネットワークで認識機能はカバーできそう
- 行動学習には強化学習（DQN）



Deep Learning

- 脳視覚系のモデル化
 - ネオコグニトン
 - CNN(視覚的なDeep Learning)
- 選択的注意のモデル
 - (認識と想起の相互作用)
- 特徴：極めて強力な学習
- 展開
 - 多階層+大量のデータ→現実的な学習
 - 新しい機能 ex. LSTM (BPTにもRNNなどいろいろある)
 - Sigmoid関数による統計的な関数近似
 - + おそらく動的Kernel法のような原理
 - 限界：意味は理解しない...なぜ？



人工知能研究の現在

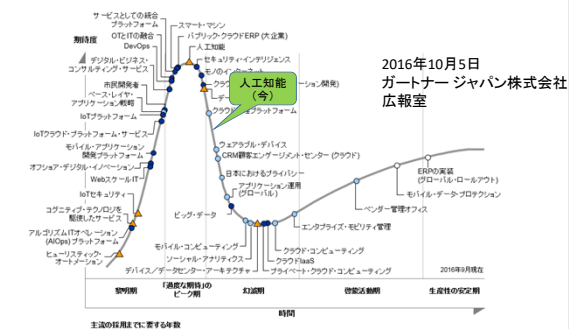
Deep Learning 全盛

- ブームは落ち着き、着実な進展に移行中 cf. 1980年代NNブーム
- 急速な技術発展, 新しい応用が次々 → 人材不足
- 米国の後追い：モデル, ツール, 応用 → 勝てない

未解決の課題

- パラメータ設定, データ数, 信頼性：研究が進行中
- 原理の理解：統計的な関数近似...の先は？
- フレーム問題：データの範囲内に機能限定して回避？
- Grounding問題は手付かず
- Beyond Deep Learning

ブームとは...



Curiosity-driven Exploration by Self-supervised Prediction

AIの学習アルゴリズムに人間でいう「好奇心」となるモデルを組み込みました。そしてスーパーマリオブラザーズやDoomをプレイさせてみたところ、画面からの情報以外に特にフィードバックがない状態でもまるで人と同じような上達を見せたと発表しました。



(a) learn to explore on Level-1



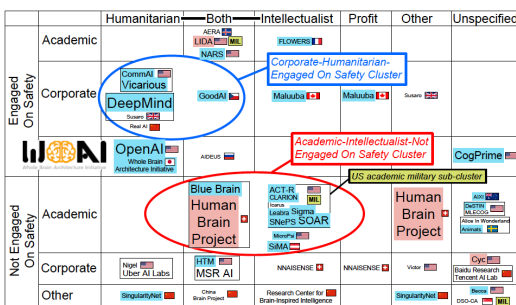
(b) explore faster on Level-2

<https://www.youtube.com/watch?v=58vJ3FHOyhUn3A>

第2回 AI&Cセミナー

ドワンゴ人工知能研究所 山川 宏氏のより借り物

Overview of the 45 identified AGI R&D projects characterized according to 7 attributes



第2回 AI&Cセミナー

ドワンゴ人工知能研究所 山川 宏氏のより借り物

神経科学分野でもAIとの接点が着目されつつある

Artificial Intelligence and Imagination:
Exploring the Frontiers of Knowledge
Demis Hassabis, PhD / University College London

人工知能の研究は驚異的な進歩を遂げています。神経科学は、AIアーキテクチャとアルゴリズムのインスピレーションと検証の両方において大きな役割を果たしています。この講義では、AIと神経科学の深いつながりと、想像力、記憶、計画などの分野での仕事の例を描いて、両分野がどのように役立つかを見ていきます。

NEUROSCIENCE 2017
THEME: TECHNIQUE
Artificial Intelligence
Demis Hassabis

Artificial intelligence (AI) research has been advancing at an incredible pace. Neuroscience plays a big role in both inspiring AI architectures and algorithms. This lecture will look at the deep connection between AI and neuroscience and how both fields are drawing on examples of work in areas such as imagination, memory, and planning. **CME**

第2回 AI&Cセミナー

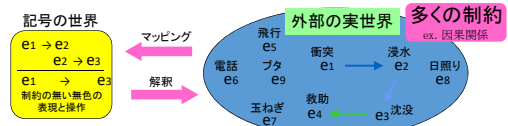
ドワンゴ人工知能研究所 山川 宏氏のより借り物

人工知能の限界



Symbol Grounding Problem (記号接地問題)

- Harnad S.: The symbol grounding problem, *Phisica D42*, p.335 (1990)
- シンボルの起源:「ある対象や行為, あるいは思考がある別のものを表すときに生じる」
 - 単語は記号, 言語は記号の操作による意味の表現
 - 指差しもまた一種の記号である
- 記号計算の記号は「ある別のものを表わしている」か? → No.
 - 結果に解釈を与えるのは別の機構・知識である
 - だから無意味な記号探索をする → フレーム問題



直観的な物理学や心理学をもたない現在のAI

(Lake, 2016)

a woman riding a horse on a dirt road

an airplane is parked on the tarmac at an airport

a group of people standing on top of a beach



Figure 6: Perceiving scenes without intuitive physics, intuitive psychology, compositionality, and causality. Image captions are generated by a deep neural network (Karpathy & Fei-Fei, 2015) using code from github.com/karpathy/neuraltalk2. Image credits: Gabriel Villena Fernandez (left), TVBS Taiwan / Agence France-Presse (middle) and AP Photo / Dave Martin (right). Similar examples using images from Reuters news can be found at twitter.com/interesting.jpg.

第2回 AI&Xセミナー

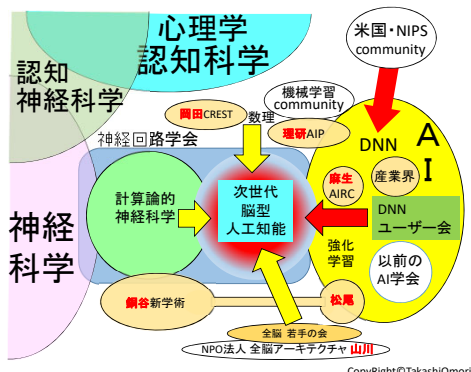
ドワンゴ人工知能研究所 山川 宏 氏のより借り物

「困ったときの脳だのみ」

- 神経細胞の解明 → Perceptron (学習の概念)
- 脳の階層構造 → 階層ネットワーク
- 脳の領野構造 ヒント 自己組織化マップ(SOM)
- ? ? バックプロパゲーション
- 視覚系の構造 ヒント ネオコグニトロン, CNN
- 動物行動学 ヒント 強化学習
- 脳の多層構造 ヒント Deep Learning ?

- 脳はニューラルネットの発展に直接には寄与していない。
- 方向性の示唆, ヒントという意味での寄与は多大
- 理由: 脳は知能のサンプル, 存在証明

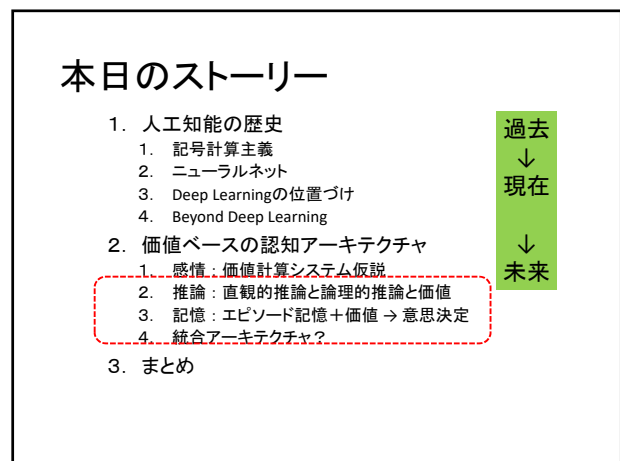
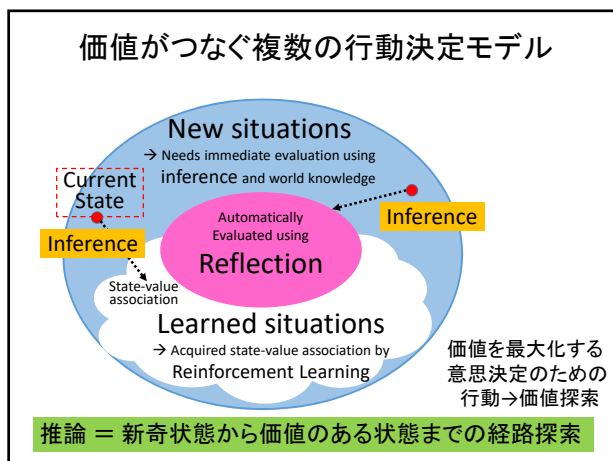
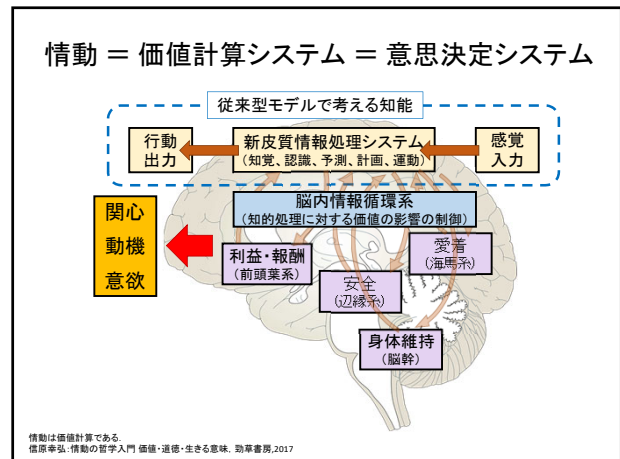
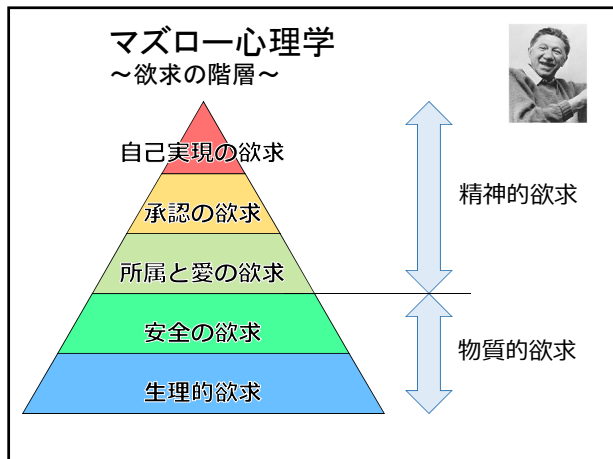
Beyond Deep Learning



本日のストーリー

1. 人工知能の歴史
 1. 記号計算主義
 2. ニューラルネット
 3. Deep Learningの位置づけ
 4. Beyond Deep Learning
2. 価値ベースの認知アーキテクチャ
 1. 感情: 価値計算システム仮説
 2. 推論: 直観的推論と論理的推論と価値
 3. 記憶: エピソード記憶+価値 → 意思決定
 4. 統合アーキテクチャ?
3. まとめ

過去
↓
現在
↓
未来

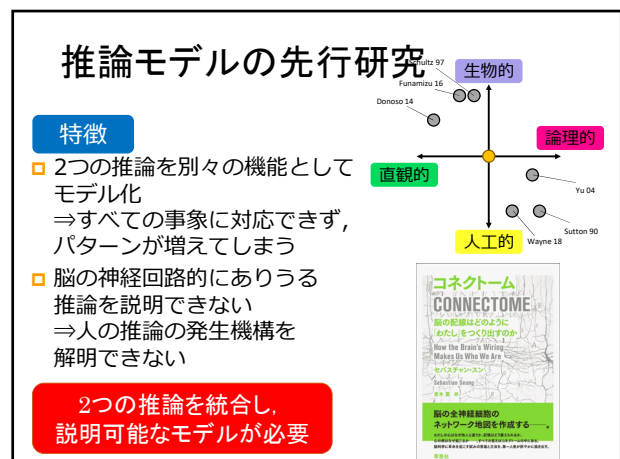


推論の二重過程と二重システム仮説 ～認知科学～

直観的推論	論理的推論
作業記憶は使わず	作業記憶が必要
無意識的, 自律的	意識的, メンタルシミュレーション
速い	遅い
バイアスに影響されやすい	規範的, 公平
文脈依存	抽象的
確率的, 分散的	論理的, シンボリック
暗黙知 (経験的確率) を利用	明示的な知識を利用
推論が浅い	深い推論が可能
進化的に古い	進化的に新しい

従来の推論モデルとはどのようなものであったか

[1] K. Evans, 2013.

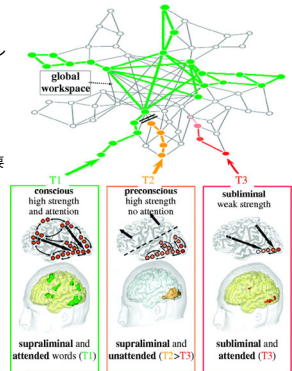


謎：脳におけるシンボル(離散情報処理)

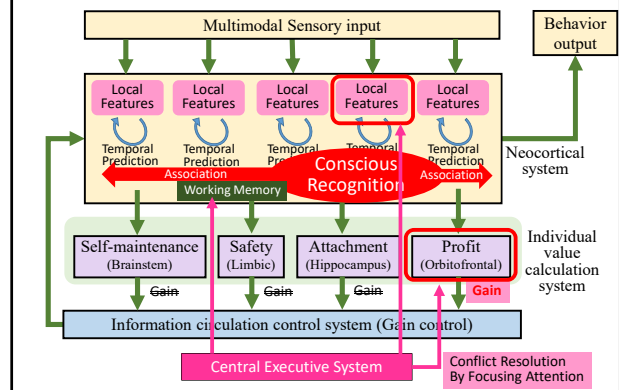
ありうるメカニズム

- System 1 と System 2 の二重プロセスモデル
→ メカニズムの説明になっていない
- 単体の細胞で、組み合わせて論理
⇒ うそつばい
- 脳全体で一時的に一つの概念
しかし、脳の活動はほぼ連続に見える
→ 全体システムとしてのアーキテクチャが必要

Global Workspace Theory
PHILOSOPHICAL TRANSACTIONS
OF THE ROYAL SOCIETY B
BIOLOGICAL SCIENCES
Levels of processing during non-conscious perception:
a critical review of 'visual masking'
Sid Kouider, Stanislas Dehaene, Published 29 May 2007.
Integrated Information Theory
(統合情報理論) **意識のモデル**



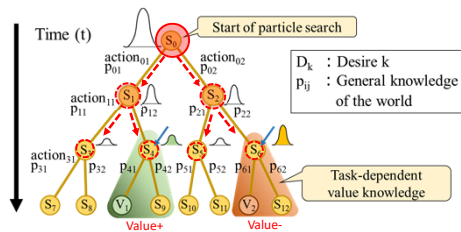
Computational Modeling of Intuitive/Symbolic Inference



直観的探索

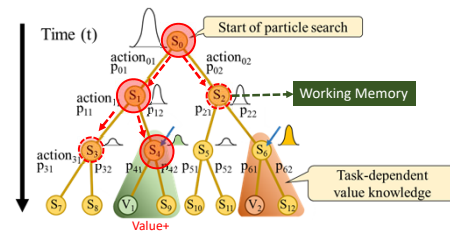
～粒子モデルによる局所確率的予測～

- 局所特徴からの予測 → 局所特徴の分布 → 部分評価
- 多数のParticleを確率的に発生して予測 → 確率的並行探索
- Particle はニューロンの発火? → 脳では合理的、媒体は脳波?
- Tree的な探索が必要な問題 → 予測が分岐 → 混乱・スパンが狭い



直観的推論 + 意識化 = シンボリック推論

- 粒子が確率的に分岐 → 予測の分布は多峰性
→ そのうち一つに注意 → 意識化 → 推論の継続
- 選択されなかった選択枝は作業記憶に保存
または、消えてしまう
- シンボリックの状態までの収束に時間 → 遅い



連想記憶を用いた推論モデル

連想記憶

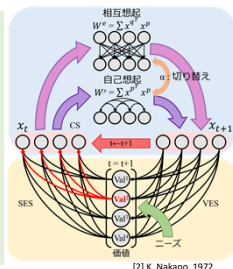
- 記憶パターンを分散的に貯蔵し、部分的な記憶情報を基に必要な記憶を読み出す

状態表現 $x^p = \{x_i^p = \{1, -1\} : i = 1 \dots N\}$
 $N = \text{例えば} 5,000$

直交条件 $\frac{1}{N} x^q x^p = \begin{cases} 1 : p = q \\ 0 : p \neq q \end{cases}$

連想行列 $W^e = \sum_q \sum_p Pr(q|p) x^q \frac{1}{N} x^p T$

連想計算 $x^c = W^e x^p$
 $= \sum_q Pr(q|p) x^q \frac{1}{N} x^p T x^p$
 $= \sum_q Pr(q|p) x^q$



アソシアトロン
中野 馨, 1972

直観的推論

～概要、および定式化～

- 局所特徴_t + 行為_t → 局所特徴_{t+1}を予測
- 神経興奮による、多方向への分岐
⇒ 確率的並列的な予測

連想的探索

$$W^e x^0 = \sum_q \sum_p Pr(q|0) x^q \frac{1}{N} x^0 T x^0$$

$$= \sum_q Pr(q|0) x^q \left(\frac{1}{N} x^0 T x^0 \right) = Pr(1|0) x^1 + Pr(2|0) x^2$$

連想の反復

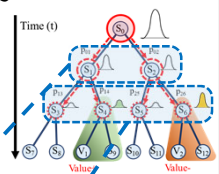
$$W^e (W^e x^0) = \sum_q \sum_p Pr(q|p) x^q \frac{1}{N} x^p T (Pr(1|0) x^1 + Pr(2|0) x^2)$$

$$= \sum_q \sum_{p=1} Pr(q|1) Pr(1|0) x^q \frac{1}{N} x^1 T x^1 + \sum_q \sum_{p=2} Pr(q|2) Pr(2|0) x^q \frac{1}{N} x^2 T x^2$$

$$= Pr(3|0) x^3 + Pr(4|0) x^4 + Pr(5|0) x^5 + Pr(6|0) x^6$$

$$Val(Wx^0) = Val^3 Pr(3|0) x^3 + Val^4 Pr(4|0) x^4 + Val^5 Pr(5|0) x^5 + Val^6 Pr(6|0) x^6$$

$$= Val^4 Pr(4|0) x^4 \Rightarrow \text{意思決定}$$

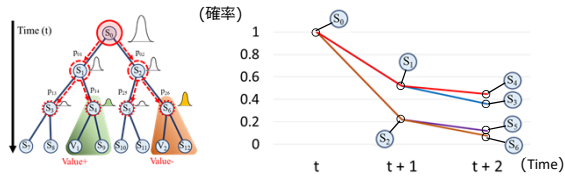


直観的推論結果

～相互想起モデルによる実装～

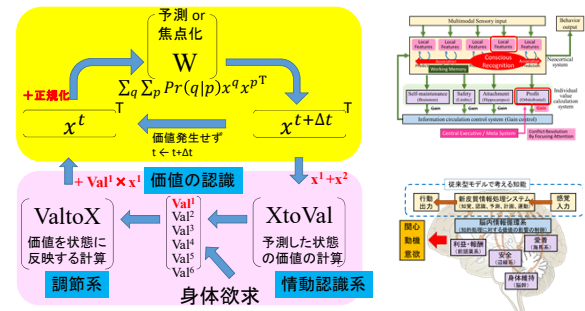
□ 状態ベクトル：N=5,000のランダムベクトル群
($S_i : i=0 \cdots 14$)

□ 現在状態(S_0)から過去の経験に基づいた
状態ベクトル毎の状態遷移の強度を計算



ベイズ的な確率推論に価値を紐づけることで
意思決定をすることが可能になった

論理的推論～連想モデルによる定式化～



論理的推論

～自己想起モデルによる定式化～

1. 価値の焦点化・予測：自己相関行列の作成

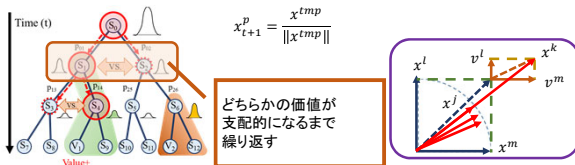
$$W^S = \sum_p x^p \frac{1}{N} x^p{}^T$$

2. 状態ベクトルの変調：状態ベクトルと価値ベクトルの合成

$$x^{tmp} = W^S (x_t^c + r W^{pp} x^c)$$

3. 更新状態ベクトルの正規化

$$x_{t+1}^p = \frac{x^{tmp}}{\|x^{tmp}\|}$$



論理的推論結果

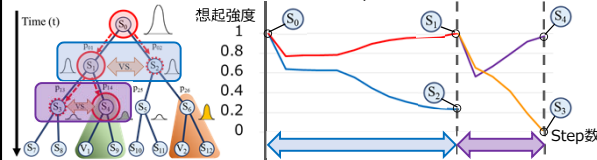
～自己想起モデルによる実装～

1Step目

□ 大きい価値が見いだせず、一方に絞れない
⇒相互抑制により、収束を促す

2Step目

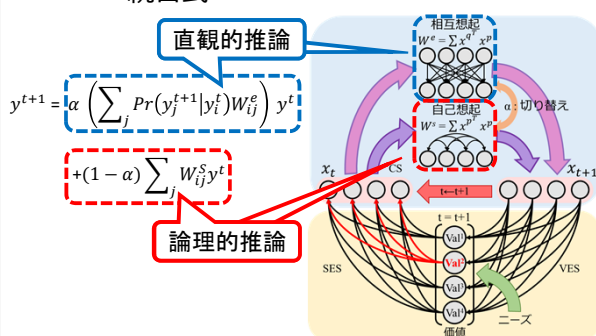
□ 価値が見いだせたことにより、高い価値に収束



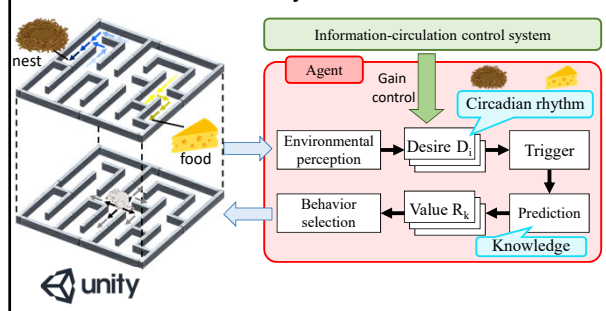
価値の大小にとられず推論をすることが可能

直観的推論・論理的推論

～統合式～



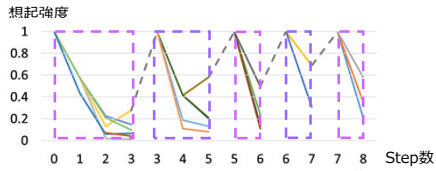
Real-time conflict resolution through value control ～Meta-system～



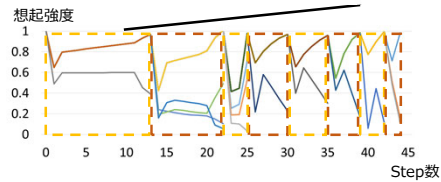
迷路課題結果

～推論機能ごとの特性比較～

直観的推論



論理的推論

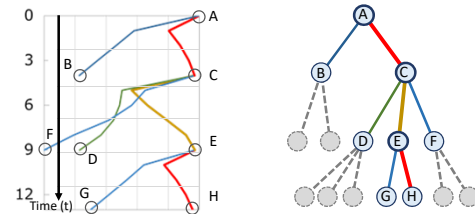


迷路課題結果

～推論結果の解釈～

- 横軸に想起強度, 縦軸に時刻
- 各点は迷路における位置を示す
- Tree探索の深さ優先探索のような結果

0 0.2 0.4 0.6 0.8 1 (想起強度)



統合アーキテクチャはあるのか？

