

「脳は単純か複雑か？脳のよう な人工知能は作れるのか？」

ミーティング資料

2015-11-12

産業技術総合研究所

人工知能研究センター

脳型人工知能研究チーム

一杉裕志

概要

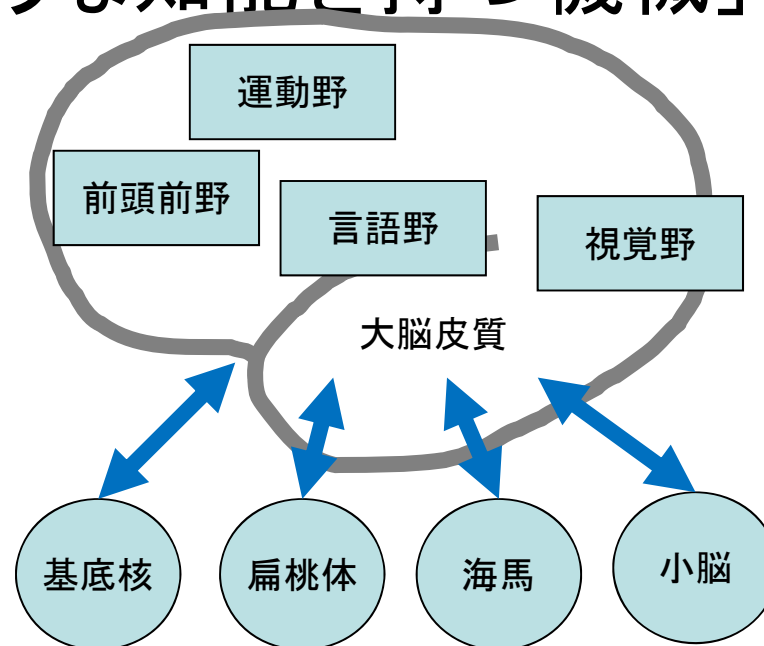
- 脳は複雑で人間に理解することは不可能と信じる人は多い。しかしひとまず「脳は単純」と仮定することで有益な結果が得られることもある。脳の機械論、斉一説、適応の度合い、振る舞いの合理性、モジュラリティ等の観点で、脳がどの程度単純でどの程度複雑かについて議論したい。

忌憚ないご意見、よろしく願います。

私の研究の長期的目標

- 脳全体のアーキテクチャを模倣して「人間のようない知能を持つ機械」を作る。

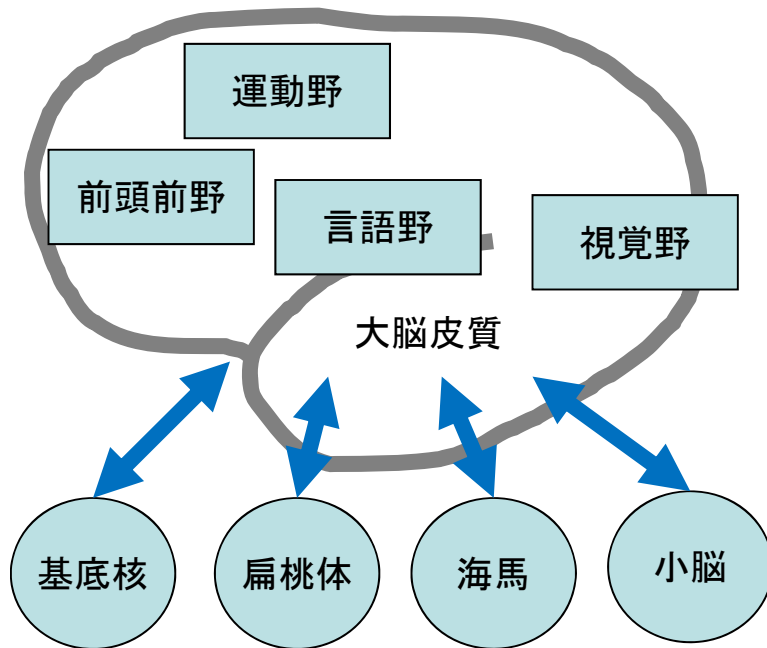
20～30年後？



脳のリバーエンジニアリング

脳の各器官のモデル

脳を構成する主要な要素



脳の各器官の機械学習装置としてのモデル

大脳皮質: SOM、ICA、ベイジアンネットワーク

大脳基底核、扁桃体: 強化学習

小脳: パーセプトロン、リキッドステートマシン

海馬: 自己連想ネットワーク

主な領野の情報処理装置としての役割

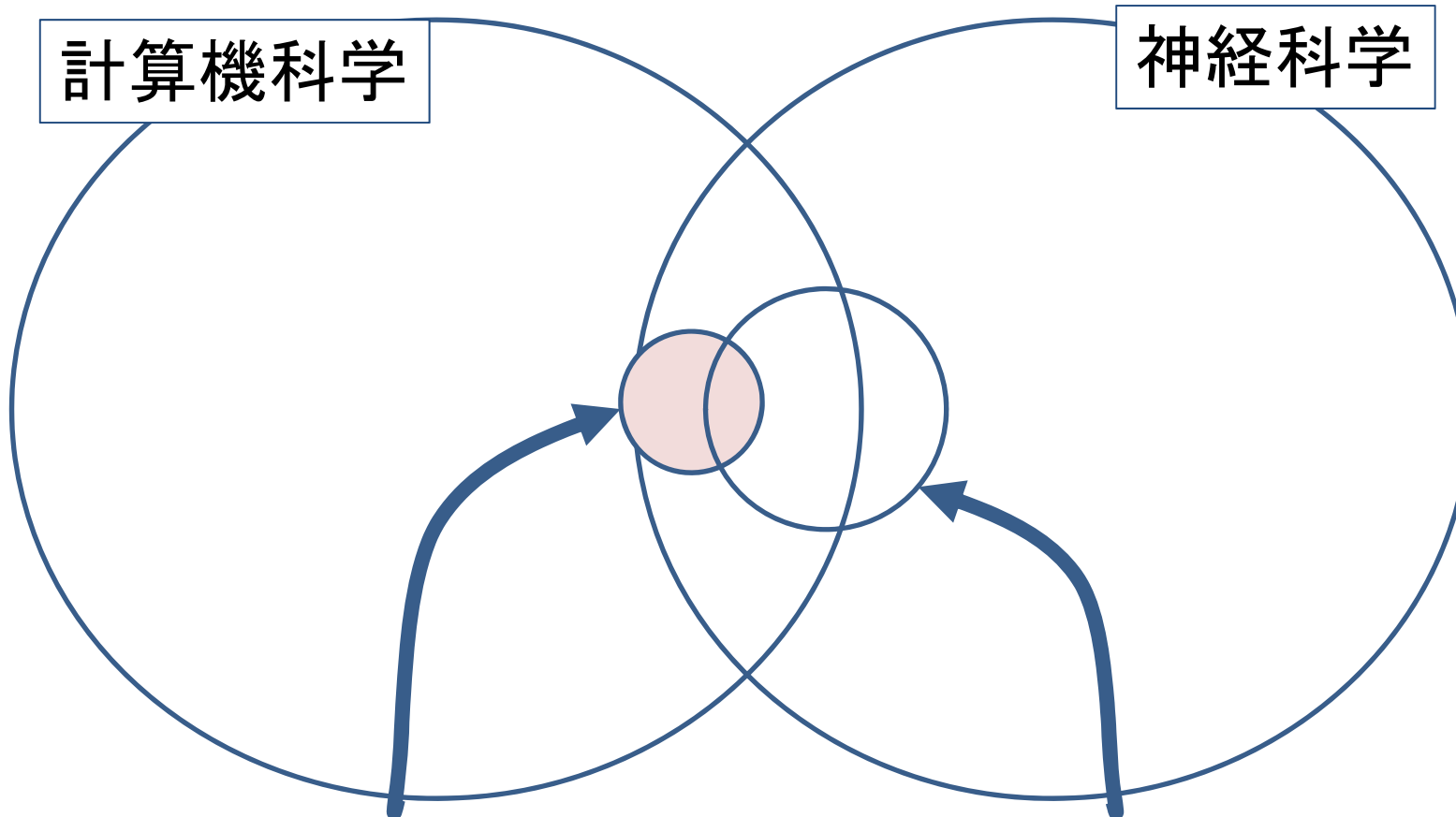
視覚野: deep learning

運動野: 階層型強化学習

前頭前野: 状態遷移機械?

言語野: チャートパーサ?

脳の知能に関する主要な器官の計算論的モデルは不完全ながら出そろってきている。これらの器官の間の連携のモデルを考えることで、脳全体の機能の再現に挑戦すべき時期に来ている。



計算機科学

神経科学

脳に学んで
人間のような知能を作る
という研究テーマ

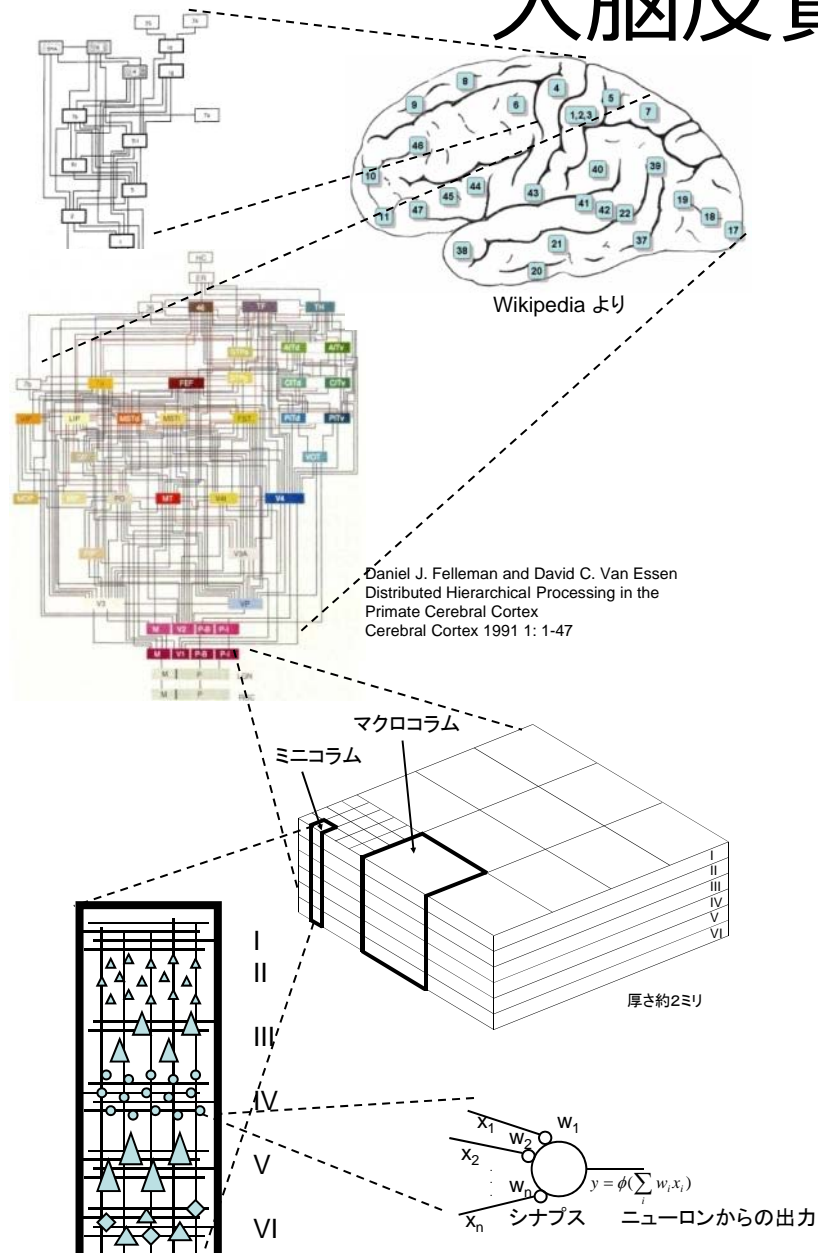
計算論的神経科学

脳を情報処理装置に見立てて
理解する学問

自己紹介

- 1990年東京工業大学大学院情報科学専攻修士課程修了、米澤研究室
- 1993年東京大学大学院情報科学専攻博士課程修了、博士(理学)
- 同年電子技術総合研究所入所
 - 並列言語、拡張可能言語、オブジェクト指向言語のモジュール機構、スクリプト言語等を研究
- 2001年より産業技術総合研究所に改組
- 2005年より計算論的神経科学を研究
- 2015年より人工知能研究センターに配属

大脳皮質の不思議さ



- **脳の様々な高次機能**
(認識、意思決定、運動制御、思考、推論、言語理解など) が、
たった50個程度の領野のネットワークで実現されている。

大脳皮質の動作原理解明が最大の課題

近似確率伝搬アルゴリズム [Ichisugi 2007]

$$l_{XY}^{t+1} = z_Y^t + W_{XY} o_Y^t$$

$$o_X^{t+1} = \prod_{Y \in \text{children}(X)} l_{XY}^{t+1}$$

$$k_{UX}^{t+1} = W_{UX}^T b_U$$

$$p_X^{t+1} = \sum_{U \in \text{parents}(X)} k_{UX}^{t+1}$$

$$r_X^{t+1} = o_X^{t+1} \otimes p_X^{t+1}$$

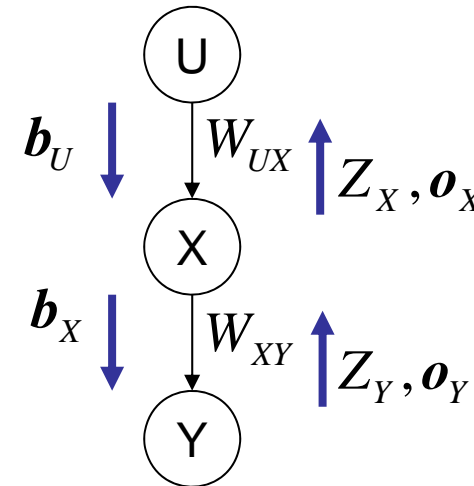
$$Z_X^{t+1} = \sum_i (r_X^{t+1})_i \quad (= \|r_X^{t+1}\|_1 = o_X^{t+1} \bullet p_X^{t+1})$$

$$z_X^{t+1} = (Z_X^{t+1}, Z_X^{t+1}, \dots, Z_X^{t+1})^T$$

$$b_X^{t+1} = (1/Z_X^{t+1}) r_X^{t+1}$$

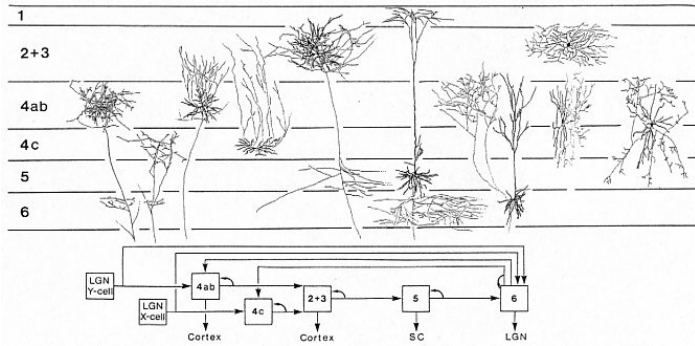
ただし、 $\mathbf{x} \otimes \mathbf{y} = (x_1 y_1, x_2 y_2, \dots, x_n y_n)^T$

Pearl のアルゴリズム [Pearl 1988] をいくつかの仮定のもとで近似。

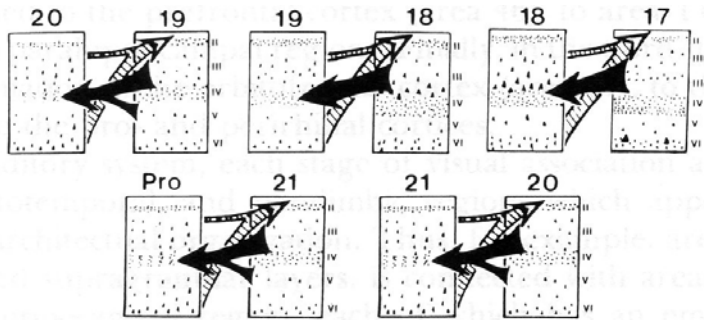


- ・神経回路で実現可能
- ・大規模化可能な計算量・記憶量

コラム構造・6層構造との一致



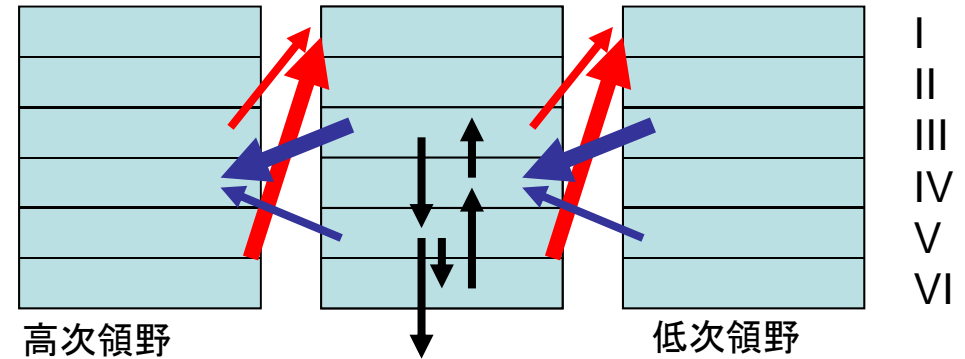
[Gilbert 1983]



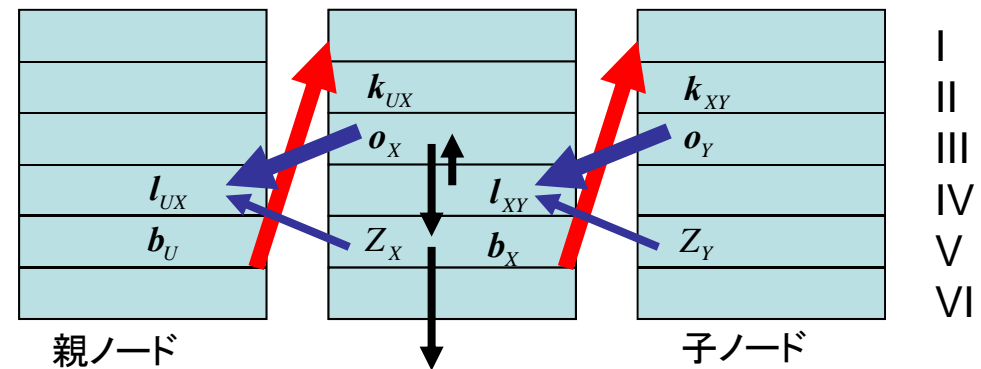
[Pandya and Yeterian 1985]

Pandya, D.N. and Yeterian, E.H., Architecture and connections of cortical association areas. In: Peters A, Jones EG, eds. Cerebral Cortex (Vol. 4): Association and Auditory Cortices. New York: Plenum Press, 3-61, 1985.

Gilbert, C.D., Microcircuitry of the visual-cortex, Annual review of neuroscience, 6: 217-247, 1983.



解剖学的構造



モデル

我々の作業仮説:「ベイジアンネットで 大脳皮質の高次機能が再現できる」

↑
抽象度
抽象

抽象化レベル	構成要素	要素数
領野	視覚野、運動野、 言語野、...	100万ノードのベ イジアンネット
ベイジアンネット	確率変数	1種類
神経回路		1000億のニュー ロン
ニューロン	興奮性、抑制 性、...	10~100種類
タンパク質等の分 子		1万以上
DNA	塩基 A,T,G,C (ある いはアミノ酸)	4 (あるいは 20)

知能の再現
に必要な
抽象レベル

病気の治療や
薬の開発に
必要な
抽象レベル

参考：コンピュータの抽象レイヤ

抽象度 ↑

抽象化レベル	構成要素	要素数
ソフトウェア		無数のアプリケーション
アプリケーション インターフェース	open, read, write, fork, ...	少数の関数・システムコール
オペレーティング システム		1000万行以上の ソースコード
命令セット	ADD, SUB, JUMP, ...	100程度
CPUアーキテクチャ		10億個以上のトラ ンジスタ
論理素子	AND, OR, NOT	3
デバイス		膨大な数の分子と 電子
原子	Si, P, B, ...	数十種類

複雑だが、
それぞれ
独立に
開発可能

単純、
万能、
安定

知能は分子レベル・神経回路レベル とは独立に実現可能であろう

↑
抽象度
抽象度

抽象化レベル	構成要素	要素数
領野	視覚野、運動野、言語野、...	100万ノードのベイジアンネット
ベイジアンネット	確率変数	1種類
神経回路		1000億のニューロン
ニューロン	興奮性、抑制性、...	10~100種類
タンパク質等の分子		1万以上
DNA	塩基 A,T,G,C (あるいはアミノ酸)	4 (あるいは 20)

複雑だが、それぞれ独立に進化？

知能の再現に必要な抽象レベル

単純、万能、安定？

確率モデルを用いる 「いまどきの人工知能技術」の例

- クラス分類: ナイーブベイズ
- トピック分類: 潜在的ディリクレ配分法(LDA)
- 音声認識: 隠れマルコフモデル(HMM)
- 言語モデル: n-gram
- ロボットの位置推定・地図推定: ダイナミックベイジアンネットワーク
- 構文解析: 確率文脈自由文法

- **大脳皮質は、おそらくこれらに似たものを包含する巨大な確率モデル**
 - 注: ベイジアンネットワークは確率モデルの一種

よく聞かれる疑問：
脳の高次機能は人間に作れる程度に
単純なのか？

脳は単純か複雑か？

単純	複雑
機械論	生氣論
斉一説	非斉一説
適応説	非適応説
合理的説	非合理的説
Modular 説	Messy 説
理解可能論	不可知論

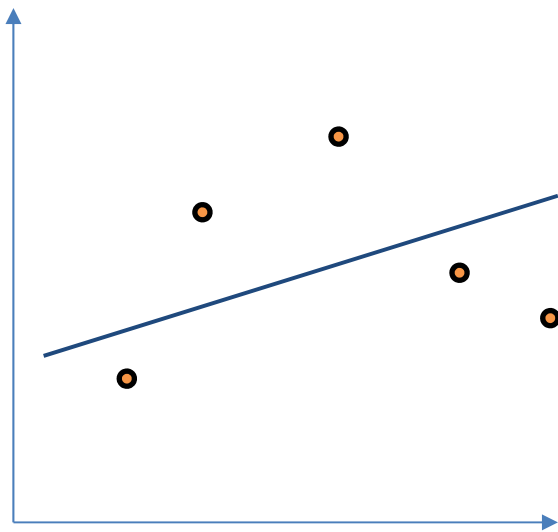
呼び方は一部
私が適当に付
けました

脳は単純か複雑か？

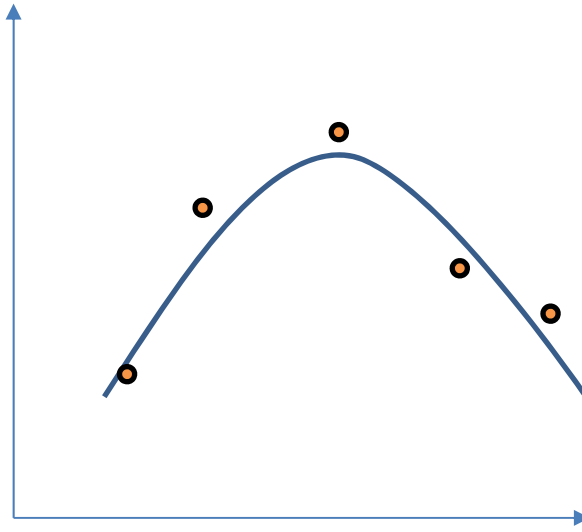
単純	複雑
機械論	生氣論
斉一説	非斉一説
適応説	非適応説
合理的説	非合理的説
Modular 説	Messy 説
理解可能論	不可知論

- いずれも科学的な仮説。
(決して科学 vs. オカルトではない。)
- 前者は実験データを完全に説明しきれないが、
後者は**ほぼ万能になんでも説明できる**。

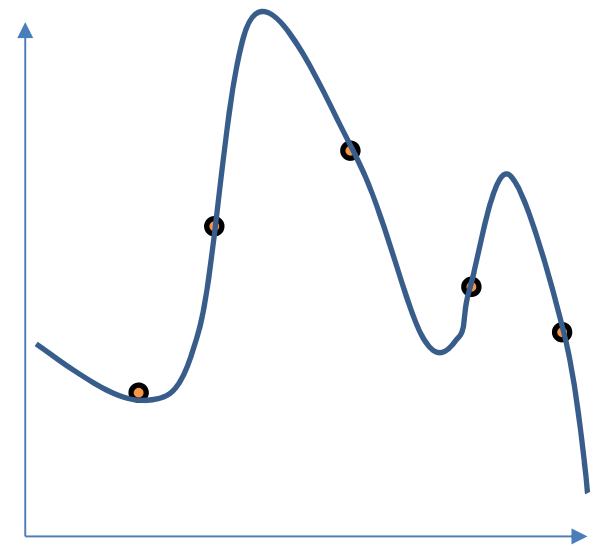
モデルの複雑さと当てはまりの良さ



$$y = ax + b$$

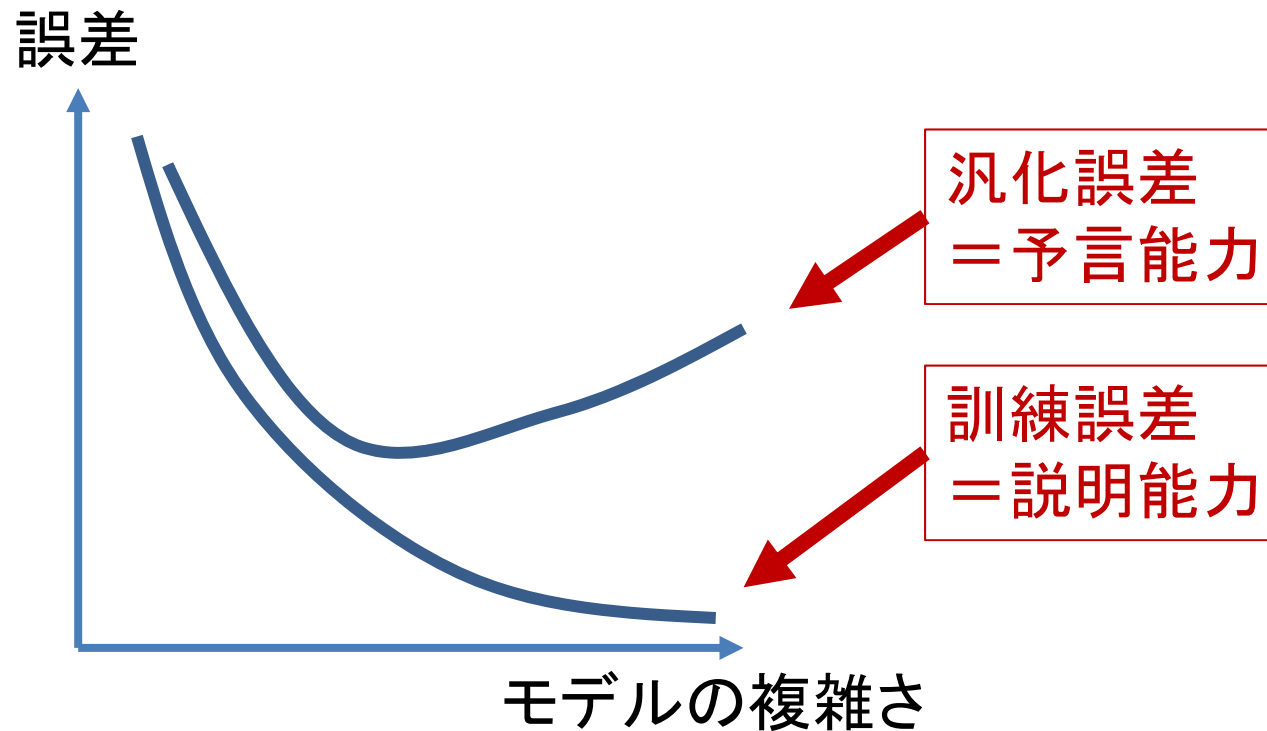


$$y = ax^2 + bx + c$$



$$y = ax^5 + bx^4 + cx^3 + dx^2 + ex + f$$

モデルの複雑さと汎化誤差



複雑すぎる仮説・理論は、実験事実の**説明能力は高いが**、未知の現象の**予言能力は低い**。

モデル選択

- データの説明に多少不都合があっても、単純なモデルを採用したほうが**有用な可能性が高い**。複雑なモデルは過適合しがち。
- 自然科学の理論・モデルにも当てはまる。**多少説明能力が悪くても単純なモデルの方が有用な可能性が高い**。

脳は単純か複雑か？

単純	複雑
機械論	生氣論
斉一説	非斉一説
適応説	非適応説
合理的説	非合理的説
Modular 説	Messy 説
理解可能論	不可知論

- 前者は実験データを完全に説明しきれないが、後者は**ほぼ万能になんでも説明できる**。しかし予言能力が弱い。前者の立場で「**ひとまず**」考えてみるのが重要。
- 「モデル選択」の考え方が身についていない人は後者が正しいと考えがち。

機械論 vs 生氣論

- 機械論：
 - 「天候や地形、生命などの諸々の自然現象を、心や精神や意志、靈魂などの概念を用いずに、その部分或いは要素の決定論的な因果関係のみ、特に古典力学的な因果連鎖のみで、解釈でき全体の振る舞いも予測できる、とする立場」
(Wikipedia「機械論」)
- 生氣論：
 - 「生命現象には物理学及び化学の法則だけでは説明できない独特の原理があるとする説」
(Wikipedia「生氣論」)

機械論（生物は他の自然現象と同じ） vs. 生氣論（生物は他と違う）の歴史

- 紀元前4世紀 アリストテレスの自然発生説
- 17世紀 大型生物の自然発生説は下火に
- 19世紀初め 有機化合物という概念の提唱
- 1828年 有機化合物の合成
- 1859年 ダーウィン「種の起源」
- 1861年 パスツールの実験 微生物の自然発生を否定
- 1944年 シュレディンガー「生物とは何か」
- 20世紀後半 分子生物学の隆盛
- 20世紀末 計算論的神経科学の発展

脳の機械論と生气論

- 共通認識:「機械論的でない知能の原理」に関する手がかりは、神経科学でも計算機科学でも何1つ見つかっていない。
- 脳の機械論者の主張:
 - 「だから知能は機械で実現可能。」
- 脳の生气論者の主張:
 - 「だから知能を人工的に実現する見込みはまったく立っていない。」
 - 「真の人工知能がいまだに実現できていない。これは機械論が間違ってる証拠。」

ひとまず機械論

- 「脳は未知の原理で動いている」という生氣論は、ノイズや実験計画の不備などで説明されるべき**あらゆる不可解な現象を説明できる**。しかし何も生み出していない。
- ひとまず脳の情報処理原理を、既存の理論の範囲で理解するよう努力するべき。
 - 情報理論・機械学習理論・確率論は十分に一般性の高い枠組み。そう簡単に拡張が必要になるとは個人的には思えない。

生氣論者も ひとまず機械論で研究すべき

- クラークの第2法則

- 「可能性の限界を測る唯一の方法は、不可能であるとされることまでやってみることである。」

- (クラークの三法則 – Wikipedia)

- 機械論で脳の高次機能を1つ1つ実現していくべき。もし生氣論が正しければ、機械論で実現できない「何か」が浮き彫りになるはず。それが生氣論の証明になる。

斉一説 vs. 「非斉一説」

- 斉一説：
 - 同じ法則が違う場所、違う時間でも成り立っている。
- 非斉一説：
 - 場所ごと、時代ごとに違う法則に支配される。

地質学における斉一説

- 「斉一説 - Wikipedia」

<http://ja.wikipedia.org/wiki/%E6%96%89%E4%B8%80%E8%AA%AC>

- 「自然において、過去に作用した過程は現在観察されている過程と同じだろう、と想定する考え方。「現在は過去を解く鍵」という表現で知られる近代地質学の基礎となった地球観。**天変地異説に対立する説**として登場した。」
- 「斉一性原理を仮定せず過去を解釈しようとするなら、ある現象が生じたのはその当時の自然法則が現在と異なっていたためとする**安易な説明**につながりやすい。斉一説は**ひとまず過去も現在と諸条件が同じだと仮定**することで、現在生じている過程である現象が説明できるなら、その説明を採用するとする近代地質学の基本的な考え方である。」

脳の斉一説と非斉一説

- 斉一説：
 - マウスの脳とヒトの脳は同じ。(同じ哺乳類)
 - ゼブラフィッシュの脳とヒトの脳は同じ。(同じ脊椎動物)
 - 視覚野も言語野も運動野も動作原理は同じ。
(同じ大脳新皮質)
- 非斉一説：
 - ヒトだけが持つ高次機能がいろいろある。ヒトの脳は他の類人猿の脳とは違いがあるはずだ。
 - 領野ごとに解剖学的構造は微妙に違う。情報処理も違うのではないか。
- これら各主張は脳の専門家の間でもよく議論になる。

ひとまず斉一説

- 「脳のアーキテクチャは種ごとに違う」、「大脳皮質の原理は領野ごとに違う」と仮定すると、いろいろな実験データが**安易に説明**できてしまう。
- 斉一説を**ひとまず仮定**すべき。
- ただし「非斉一説」の可能性もつねに念頭に置くべき。
 - 例えば、チョウチンアンコウの発光器官はもとはヒレだが、もはや泳ぐための器官ではない。
同様に、ヒト高次機能に関与する領野は他の領野と多少違う可能性がある。

「適応説」vs.「非適応説」

- 「適応説」:
 - 生物は子孫を確実に残せるよう最適化されている。
 - 安易な適応説の主張は「適応主義」という呼び名で批判される。
「鼻はメガネをかけるためにある」
- 「非適応説」:
 - 「適応途中説」:
人類の脳は進化の途中であり最適化しきれていない。
 - 「偶然説」:
いろいろな形質は単なる「進化上の偶然」。

ひとまず適応説

- 偶然説はどんな形質も説明できる万能理論。しかし、何も生み出さない。
- ただし「適応途中説」は常に念頭に置く必要がある。
 - 例：人類が文字を使うようになったのはごく最近。脳は文字の処理には最適化されていない。（文字の方が脳に処理しやすいよう最適化されている。）

「合理的説」vs.「非合理的説」

- 合理的説: 生物は子孫を残すために合理的に振る舞う。
 - 生物学者は生物は合理的だと考える人が多い気がする。
 - 生物の一見不合理な振る舞いにも合理的解釈を与えようと努力する。
 - 例: 利他的行動に対する血縁淘汰説
 - 合理的解釈がない場合、なぜ不合理な振る舞いをする生物が淘汰されずに生き残ったのかを説明する必要が生じる。
- 非合理的説: 生物の振る舞いは非合理的である。
 - 認知科学者は生物(人間)は非合理的だと考える人が多い気がする。
 - 人間の不合理な振る舞いを見つけようと努力する。
 - 人間の一眼不合理に見える振る舞いに合理的解釈を与えようとする試みは比較的少ない。
 - 合理的解釈の証明は難しく、たぶん論文になりにくい。

ひとまず「合理的説」、しかし、 真に非合理的な振る舞いにこそ期待

- 人間の振る舞いは複雑だが、少なくとも近似的には合理的に振る舞うと考えてよいだろう。
 - 100円と1万円を選ばせると100円を取る人はいない！
- 「合理的説」を前提にしなければ、脳のアルゴリズムを推定しようがない。
- 一方で、**不合理な振る舞いの方が、脳のアルゴリズムの候補を絞り込む制約条件としては価値が高い。**
 - 残念ながら、生物の一見不合理に見える振る舞いも、よく考えると合理的な説明が付きそうな場合が多い。
 - それだと脳のアルゴリズムを推定するヒントとしては役立たない。
 - **真に非合理的な振る舞いをご存知の研究者の方はお知らせください。**

「Modular 説」 vs. 「Messy 説」

- Modular 説
 - 生物の体も脳も、それなりにモジュール化され、機能が分担されている。
- Messy 説
 - 進化はいきあたりばったり。たまたまうまく動けばそれがずっと使われる。
 - 人工物は設計者の都合でモジュール分割されるが、生物には設計者はいない。

脳の modular 説と messy 説

- Modular 説:
 - 大脳皮質、海馬、基底核、小脳はそれぞれ違う情報処理を行っている。
 - 領野ごとに受け持つ機能が違う。
(大脳皮質の機能局在説)
- Messy 説:
 - 脳の機能は簡単にはモジュールに分けられない。1つの高次機能の実現には複数の器官が密接に関わっている。
 - したがって脳の理解は一筋縄ではいかない。
- どちらもそれなりに正しい。

脳はそこそこ modular

- 環境が不変ならば生物は messy なのが最適。でも環境が変化するときには modular じゃないと進化しにくいのでは。
 - Messy 説によれば脳だけでなく内臓も messy はず。しかし、心臓はポンプ、肺はガス交換、と言うふうに modular になってる。
- 脳は1つの目的関数を最適化するために複数の器官がシームレスに協調動作するシステム。modular だとしても、システム全体を理解することが大事。

「理解可能論」vs.「不可知論」

- 脳の理解可能論：
 - 「脳は人間に理解可能なはず。」
← 自然科学の大前提となる作業仮説
 - 「脳と同じものが人工的に作れるはず。」
← 工学の大前提となる作業仮説
- 脳の不可知論：
 - 「脳は人間に理解できないかもしれない。」
 - 医学系の人には安易に「脳を理解した」と考えることを戒める。
 - 理解したつもりの安易な治療は大変危険！
前例：ロボットミー、薬物治療の副作用

「真の人工知能はできない」と 断言する専門家は多い

- よくよく聞いてみると「私がいままでやっていた研究の延長ではできません。」と言っているだけだったりする。
- クラークの第1法則
 - 「高名だが年配の科学者が可能であると言った場合、その主張はほぼ間違いない。また不可能であると言った場合には、その主張はまず間違っている。」

(クラークの三法則 – Wikipedia)

ひとまず「理解可能論」

- 理学・工学系の人には「脳は理解可能・実現可能」という作業仮説を信じて研究すべき。
 - 「できないかもなあ」と思いながら研究しているのは、できるはずのものも、できない。

脳は単純か複雑か？まとめ

ひとまず単純と仮定	ただし…
機械論	生氣論の可能性も0ではない
斉一説	非斉一説も念頭に
適応説	適応途中説も念頭に
合理的説	不合理な振る舞いはよいヒント
Modular 説	システム全体の理解が大事
理解可能論	人体への適用は慎重に