

大脳皮質って何？ 大脳皮質についての基礎知識

第37回 全脳アーキテクチャ勉強会

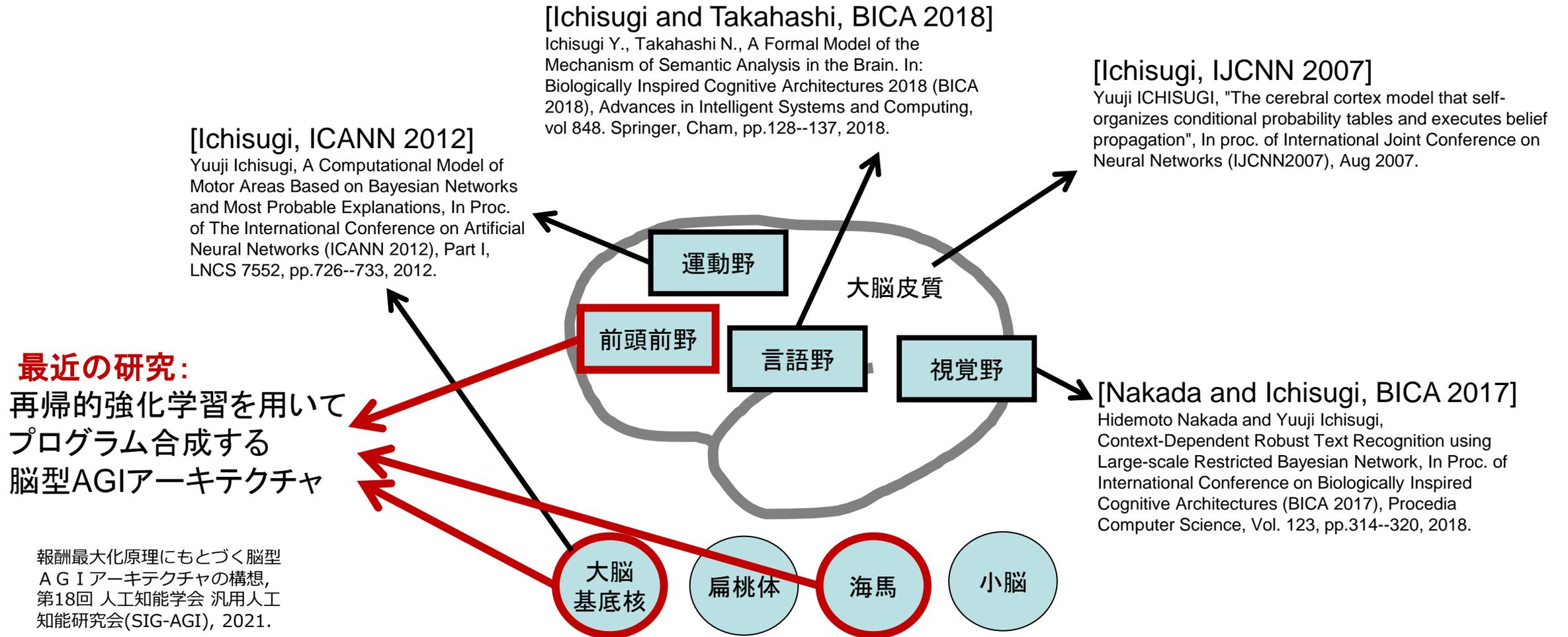
一杉裕志（産総研）

2022-02-17

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
- 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
(きょうの話は主に 2015年ごろの自分の知識)

私の研究の長期的目標： 脳全体のアーキテクチャに学んだAGIの実現



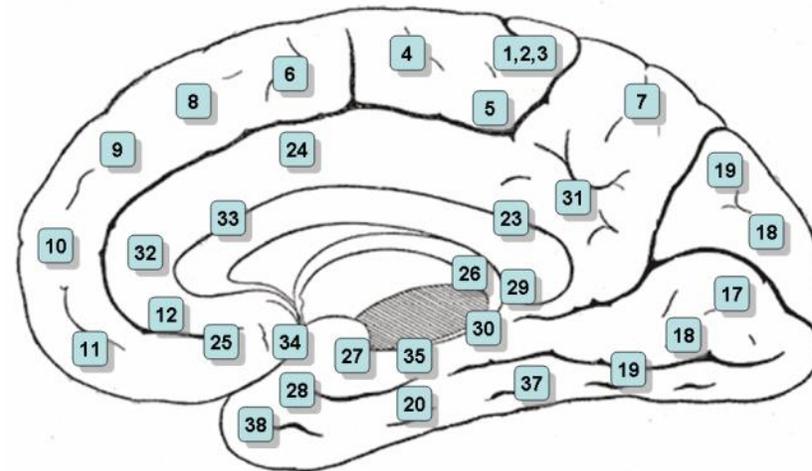
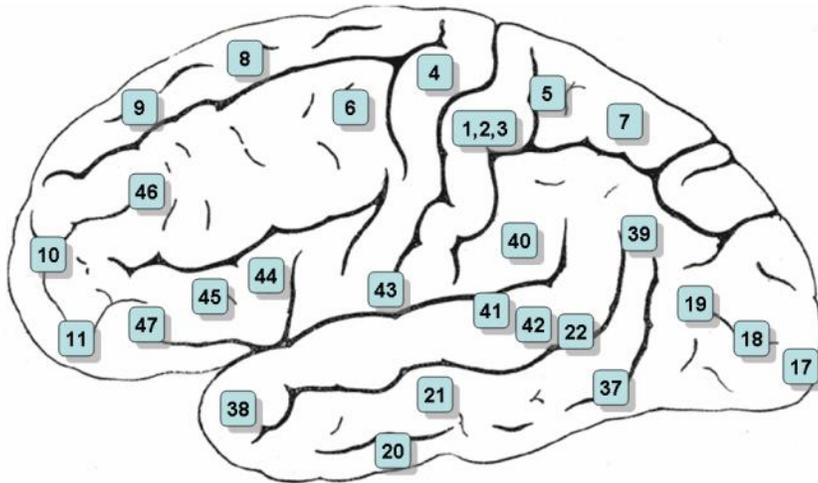
今日の話の構成

- 大脳皮質の基礎知識（25分）
 - 2015年前後に使った講演資料をベースにした内容
- 大脳皮質の計算論的モデル（10分）
 - SOM、スパースコーディング、ベイジアンネット
- 私の最近の研究（脳型AGIアーキテクチャ）との関係（5分）
- 質疑（5分）

大脳皮質に関する 神経科学的知見

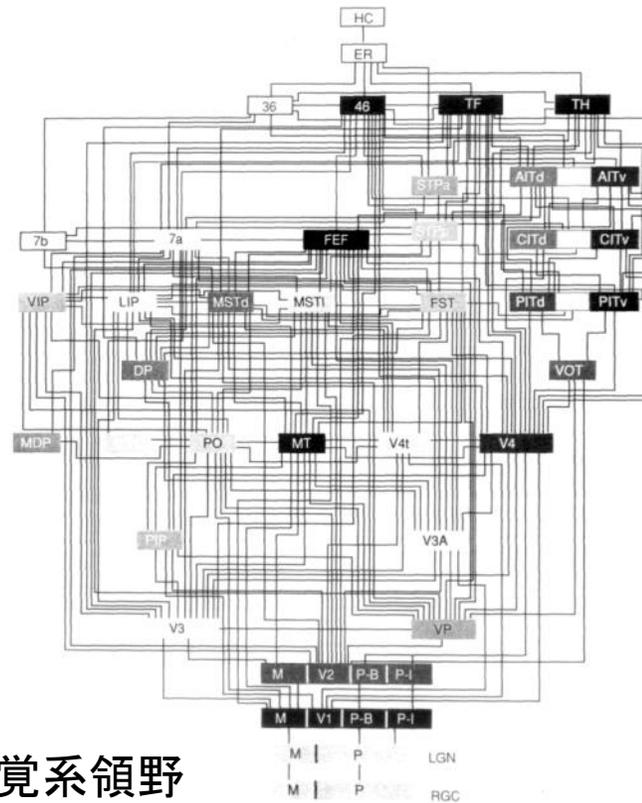
大脳皮質

- 脳の中でも知能をつかさどる重要な部分
 - 視覚野、言語野、運動野、前頭前野、 . . .



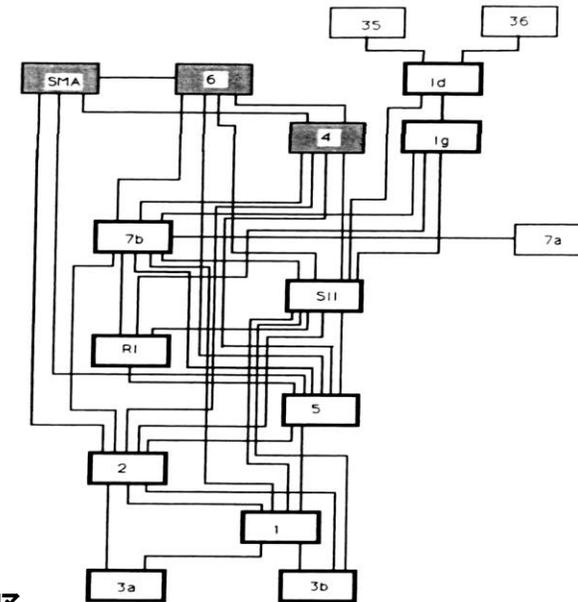
大脳皮質の領野

- 各領野の機能、接続構造はかなり明らかになりつつある



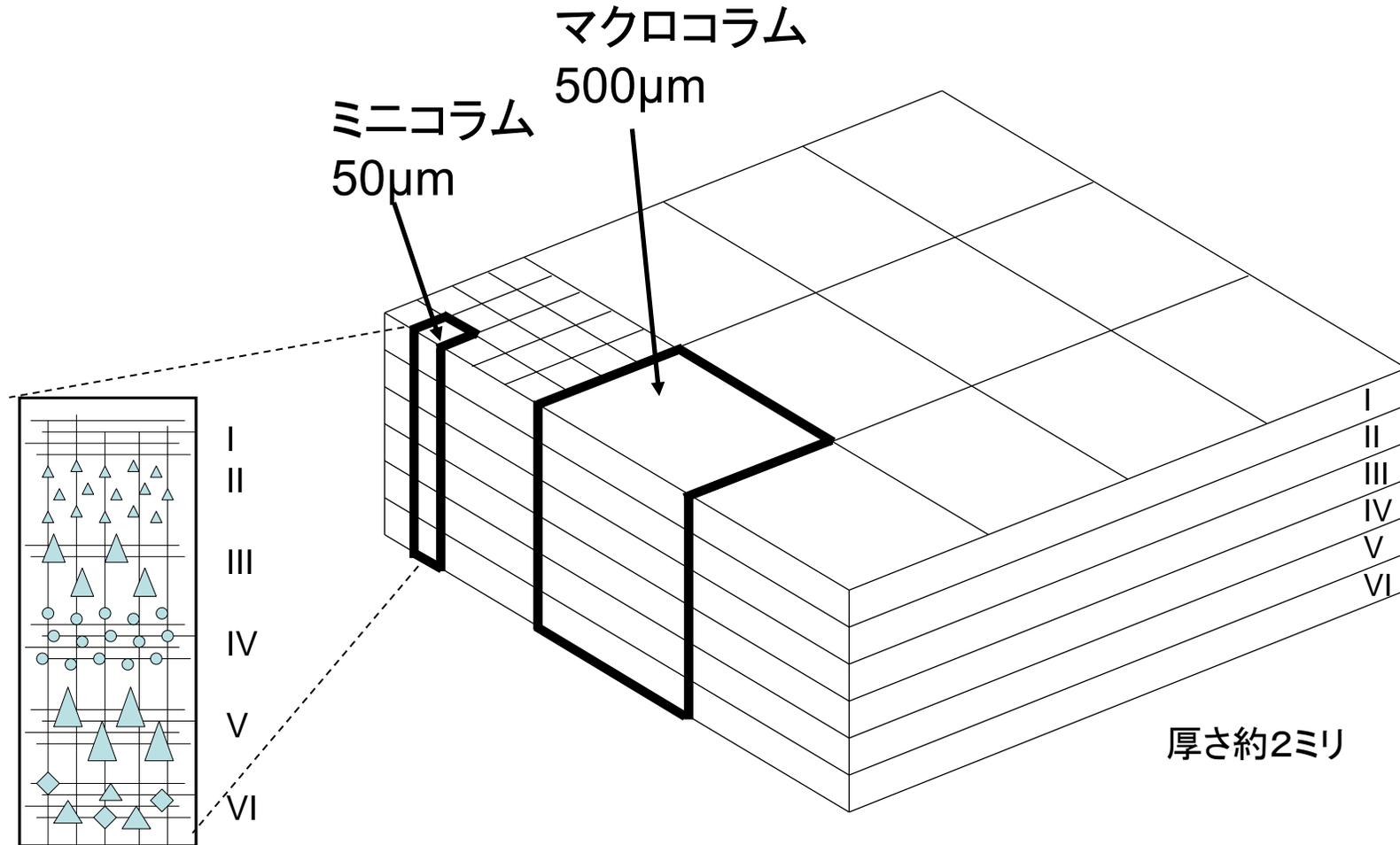
視覚系領野

Daniel J. Felleman and David C. Van Essen
Distributed Hierarchical Processing in the Primate Cerebral Cortex
Cerebral Cortex 1991 1: 1-47



運動野

大脳皮質のコラム構造の模式図



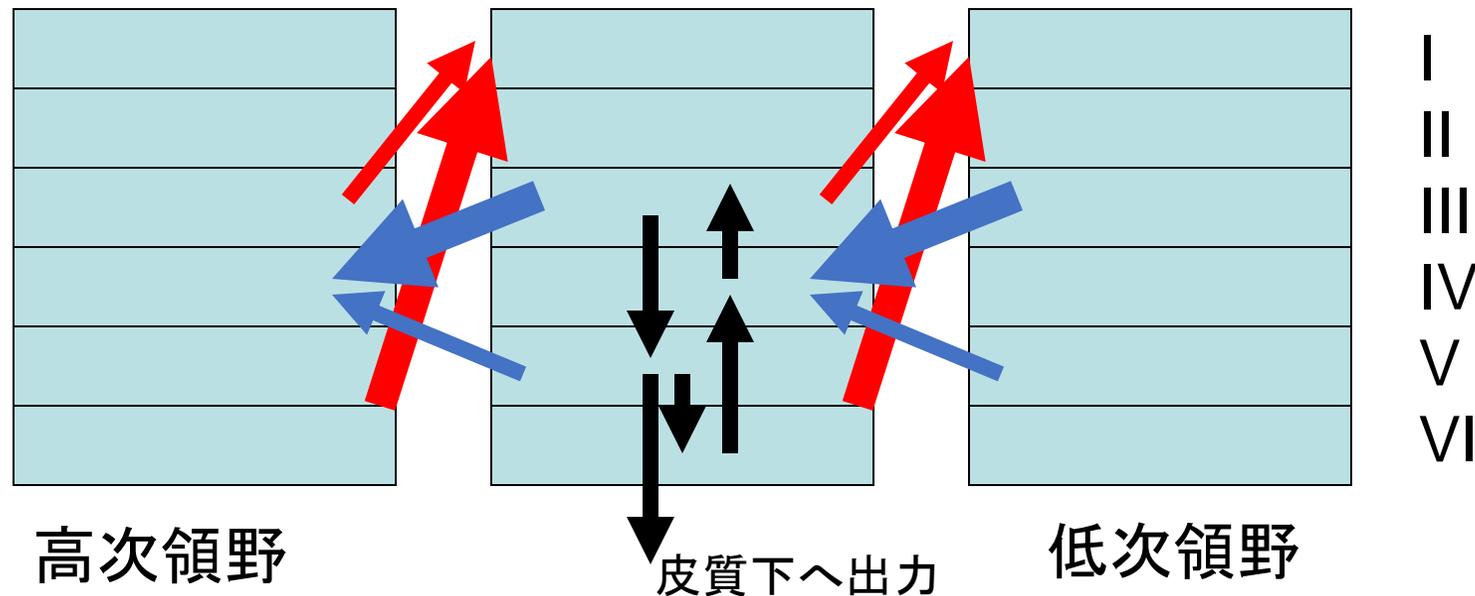
IT野の顔の向きに応答するコラム [Wang et al. 1996]

Wang G, Tanaka K and Tanifuji M,
Optical imaging of functional organization in the
monkey inferotemporal cortex
SCIENCE 272 (5268): 1665-1668 JUN 14 1996.

図は下記論文より
Keiji Tanaka
Columns for Complex Visual Object Features in the Inferotemporal Cortex:
Clustering of Cells with Similar but Slightly Different Stimulus Selectivities
Cerebral Cortex, Vol. 13, No. 1, 90-99, January 2003

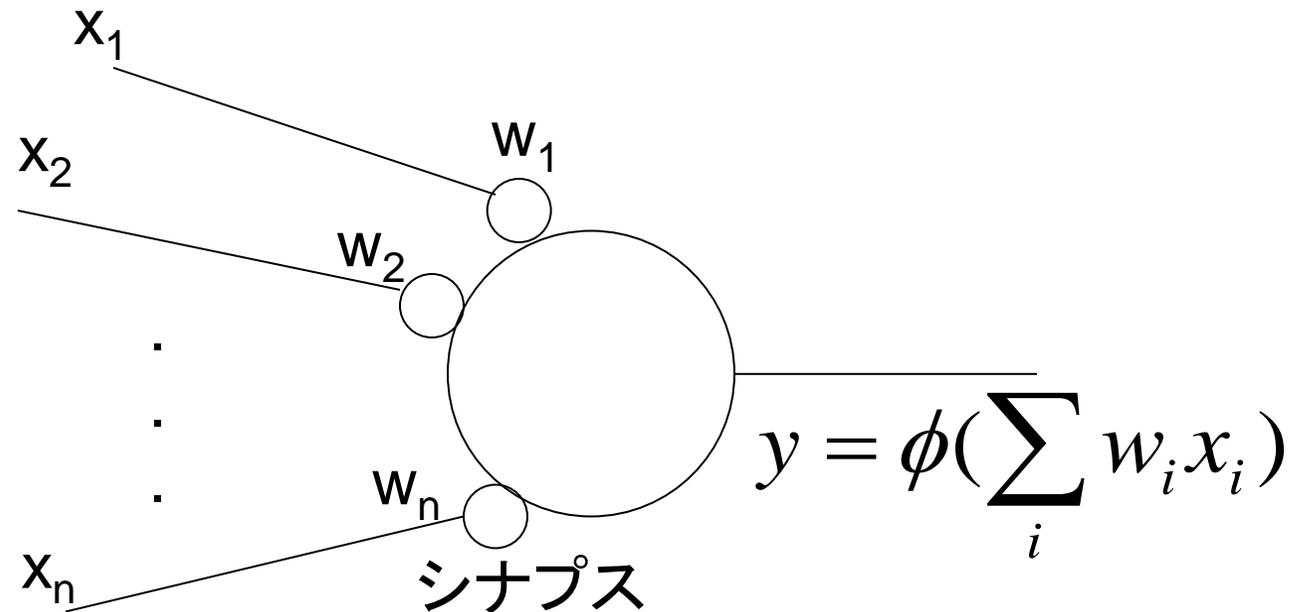
領野間・層間の結合 [Pandya and Yeterian 1985] [Gilbert 1983]

- 情報処理の途中結果の3層の情報が上位領野に送られ、最終結果の5層の情報は下位領野に戻る



ニューロン（神経細胞）

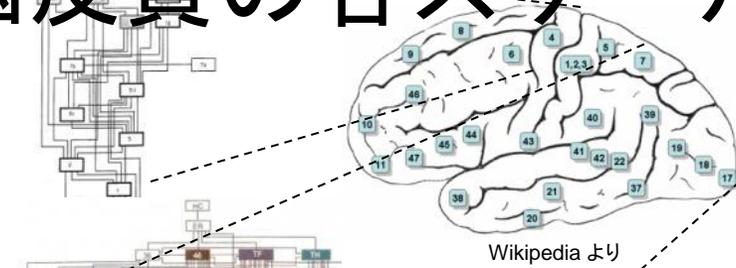
- ヒトの脳で1000億個
- 内積計算のような簡単な演算しか行えない
- シナプスが w という重みを学習



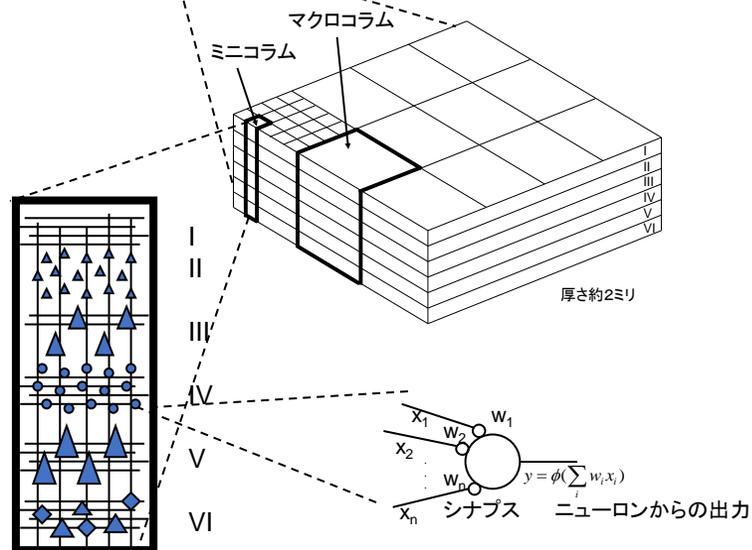
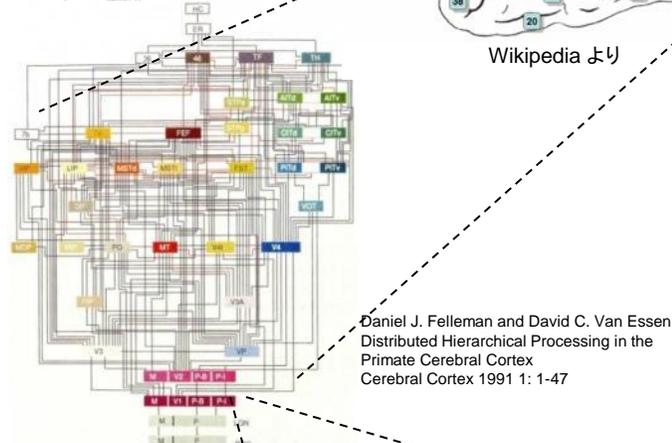
ニューロンへの入力

ニューロンからの出力

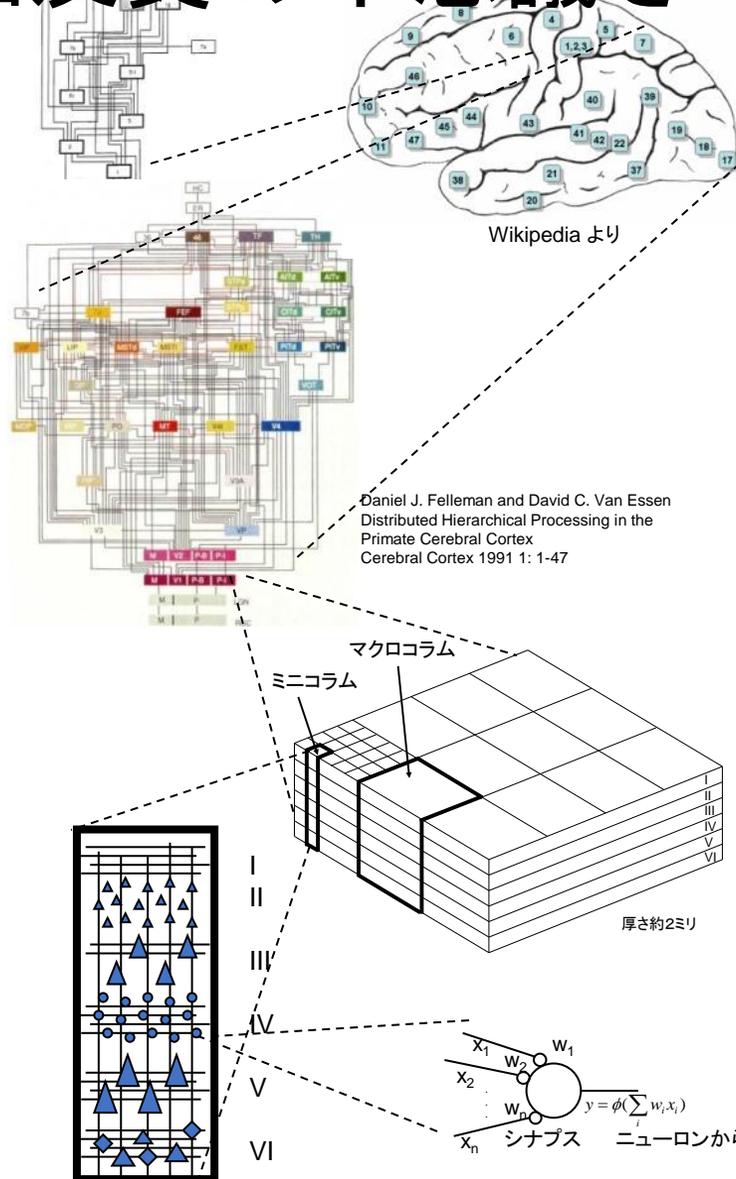
大脳皮質の各スケールでの構造



- 領野 約 50 個
- マクロコラム 約 100 万個
- ミニコラム 約 1 億個
- ニューロン 約 100 億個
- シナプス 約 100 兆個



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

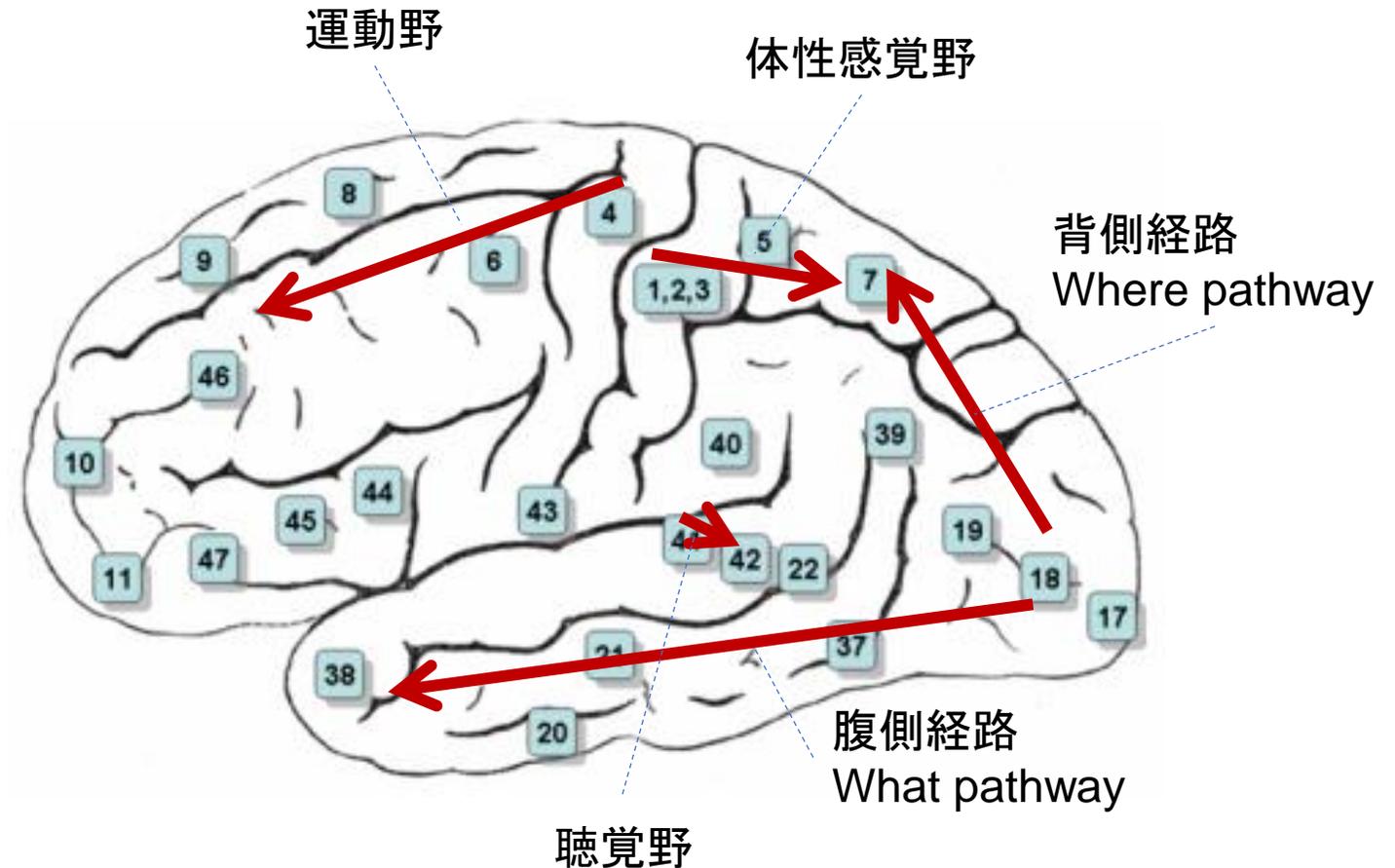
大脳皮質の動作原理がわかれば
脳のアーキテクチャ解明の
重要な手掛かりに

大脳皮質の領野

大脳皮質と deep learning の 大きな類似点

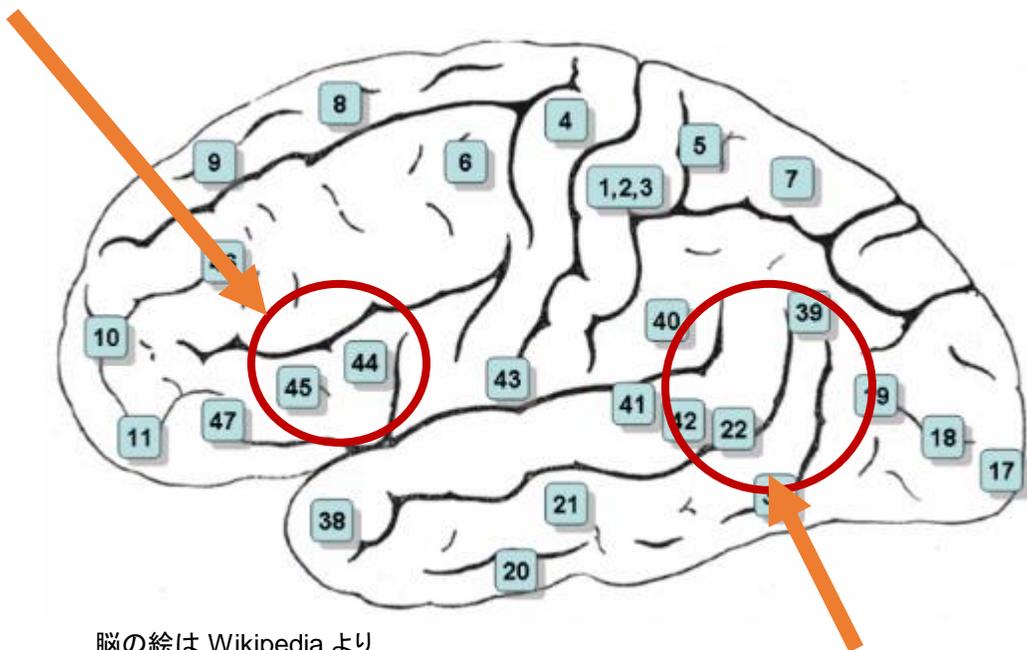
- 領野の階層構造

- 上の層ほど多くの特徴量（感覚入力）を統合
- 上の層ほど複雑な情報を表現
- 上の層ほど抽象度が高い情報を表現



言語野

ブローカー野(44・45野)



脳の絵は Wikipedia より

ウェルニッケ野
(一次聴覚野と角回の近く)

• ブローカー野：文法処理に関与

- ブローカー失語：
 - 発話：電文体（機能語が省略される）
 - 理解：文法に強く依存した文の理解が困難

• ウェルニッケ野：言語音と概念の連合に関与

- ウェルニッケ失語：
 - 発話：流暢に話すが意味のない内容
 - 理解：障害されている

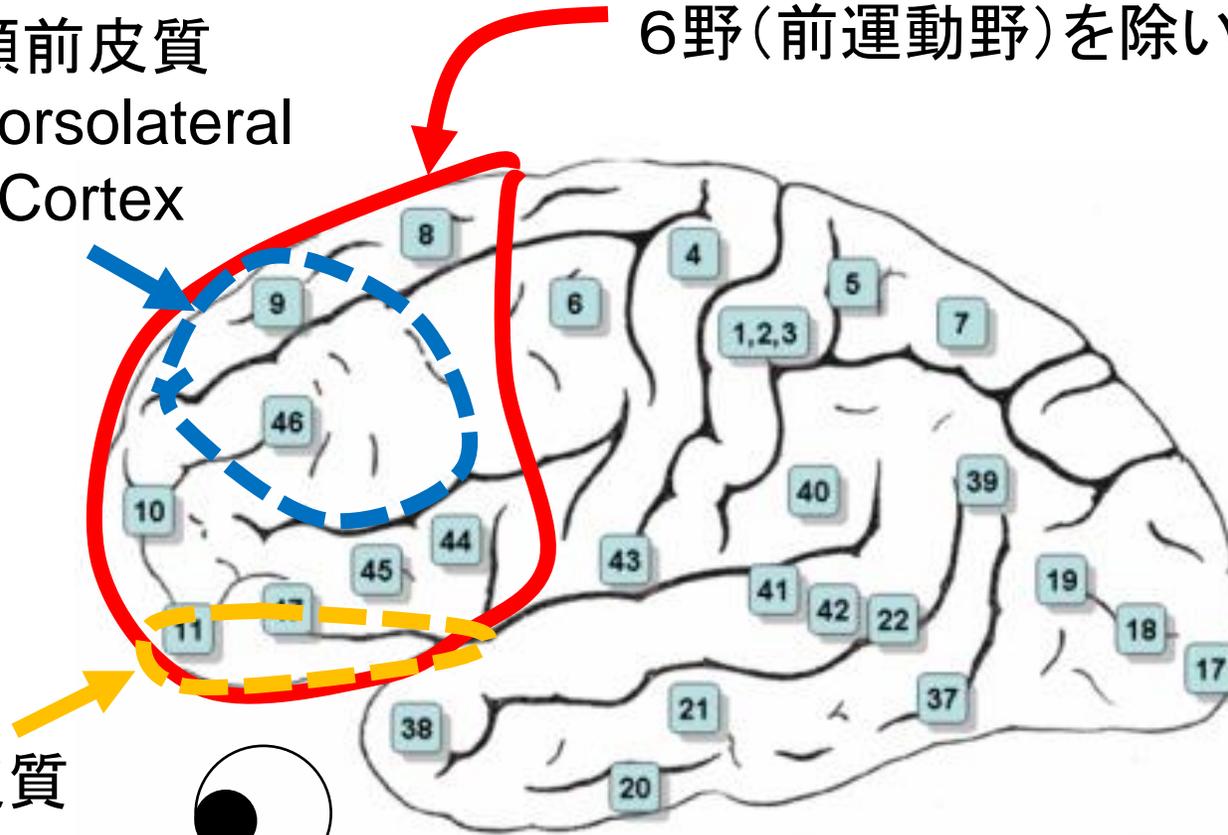
参考:カンデル神経科学 p.1336-1339、言語処理学事典 p.772-787

脳においては「発話」と「理解」というモジュール分割ではなく
「文法」と「意味」というモジュール分割がなされている

前頭前野 (PFC, Prefrontal cortex)

前頭葉のうち4野(一次運動野)と6野(前運動野)を除いたところ。

背外側前頭前皮質
DLPFC, Dorsolateral Prefrontal Cortex

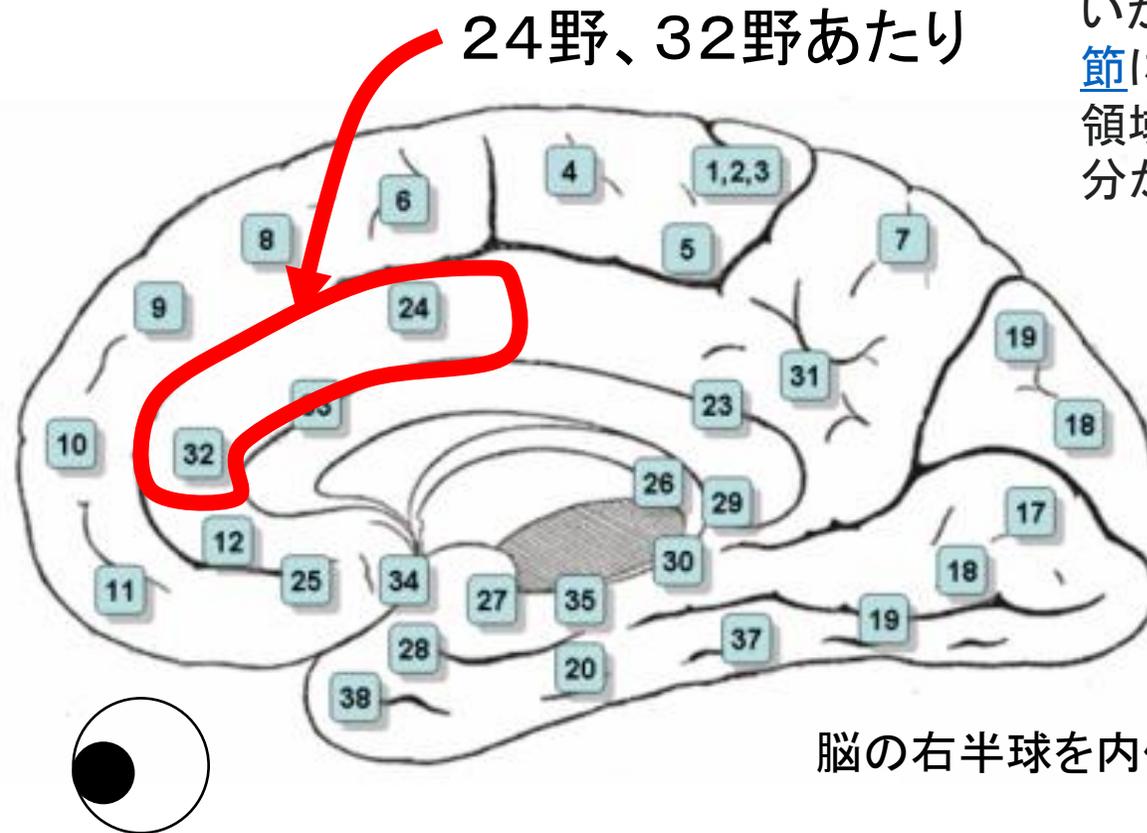


「前頭前野はヒトをヒトたらしめ、思考や創造性を担う脳の最初中枢であると考えられている。」

「前頭前野 - 脳科学辞典」
<http://bsd.neuroinf.jp/wiki/%E5%89%8D%E9%A0%AD%E5%89%8D%E9%87%8E>

眼窩前頭皮質
OFC,
Orbitofrontal cortex

前帯状皮質 (ACC, Anterior cingulate cortex)

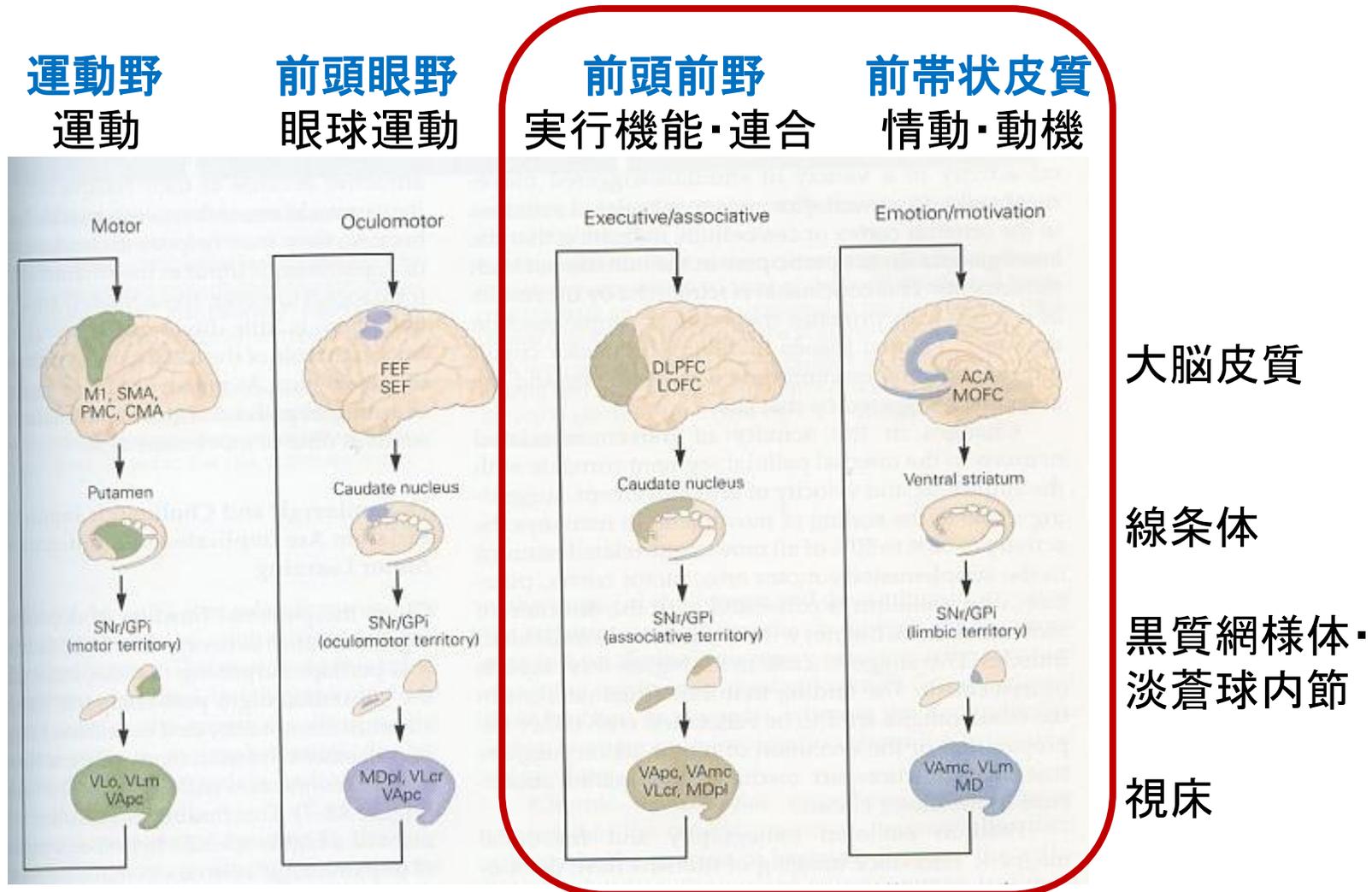


ACCは特にヒトにおいて、担う機能の違いから、行動モニタリングおよび行動調節に関わる領域、社会的認知に関わる領域、および情動に関わる領域に大きく分かれる。

「前帯状皮質 - 脳科学辞典」

<https://bsd.neuroinf.jp/wiki/%E5%89%8D%E5%B8%AF%E7%8A%B6%E7%9A%AE%E8%B3%AA>

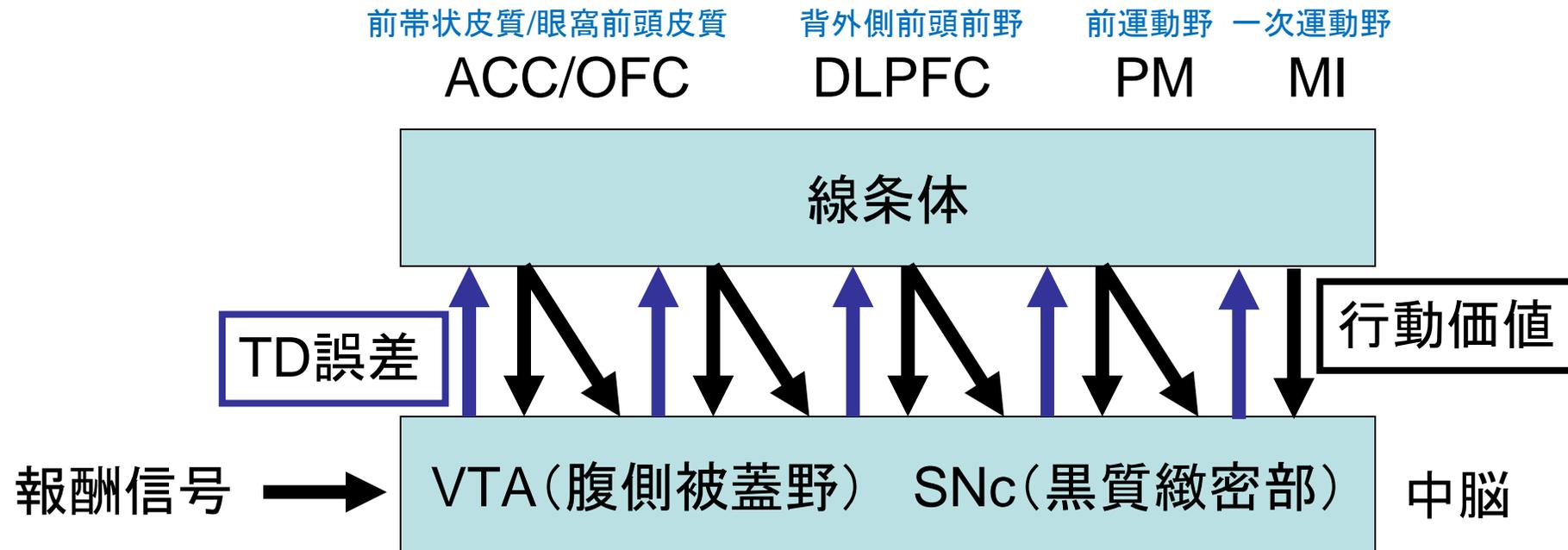
強化学習を行う並行した解剖学的結合のループ



Principles of Neural Science 5th ed.
Eric R. Kandel et al. , McGraw-Hill, p. 987

並行ループの間の階層

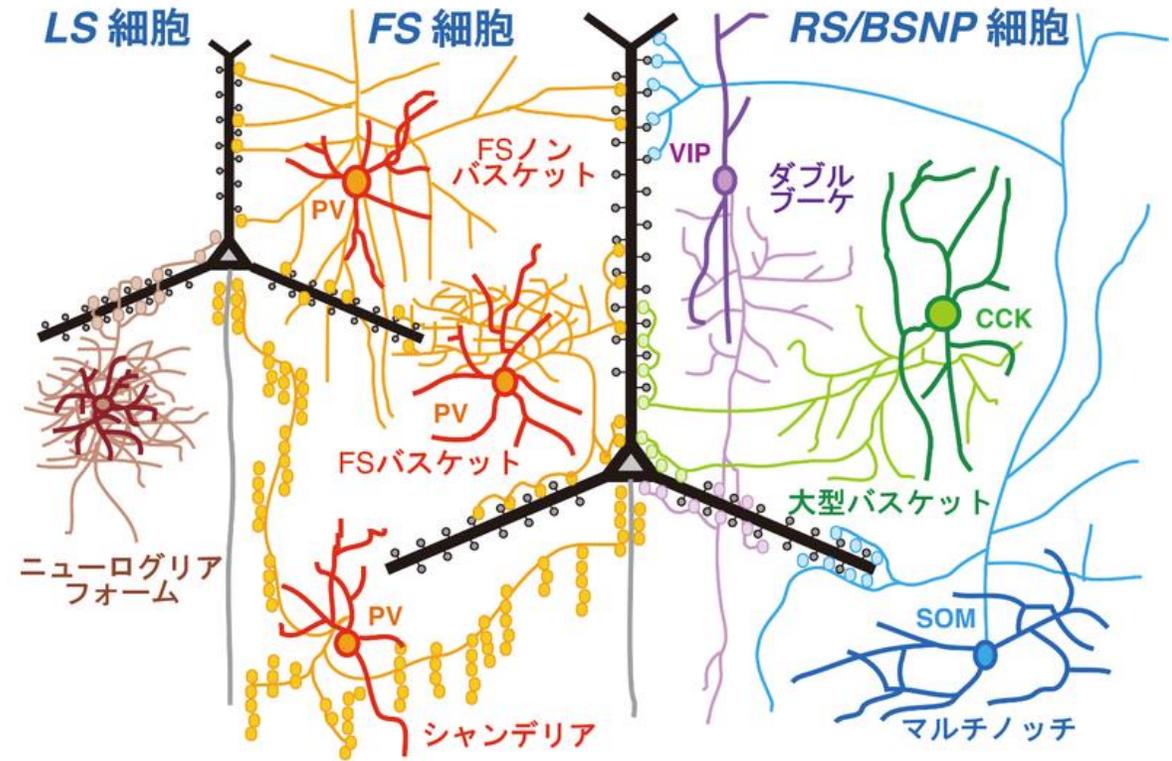
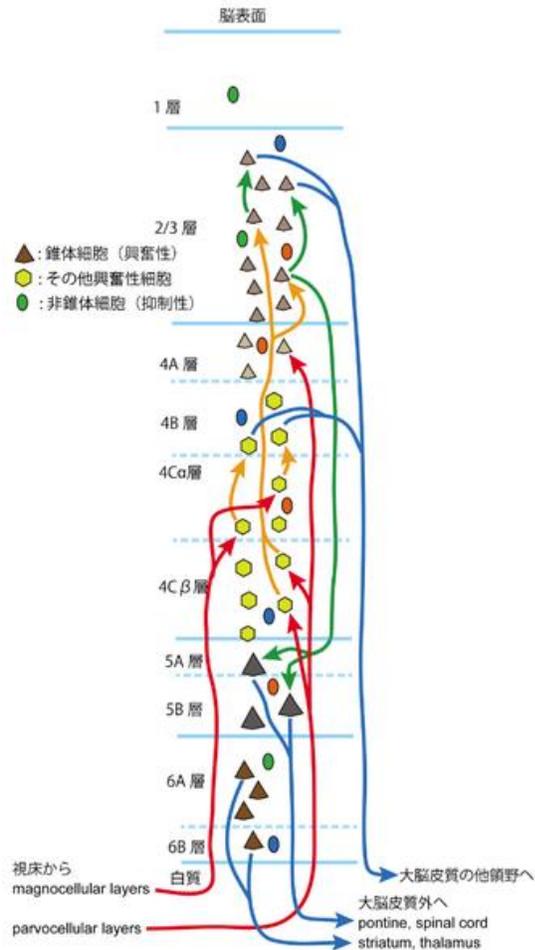
前帯状皮質(ACC)が最上位の階層



Haber SN. , The primate basal ganglia: parallel and integrative networks. J Chem Neuroanat. 2003 Dec;26(4):317-30.

大脳皮質の局所回路と ニューロン

大脳皮質内の局所回路



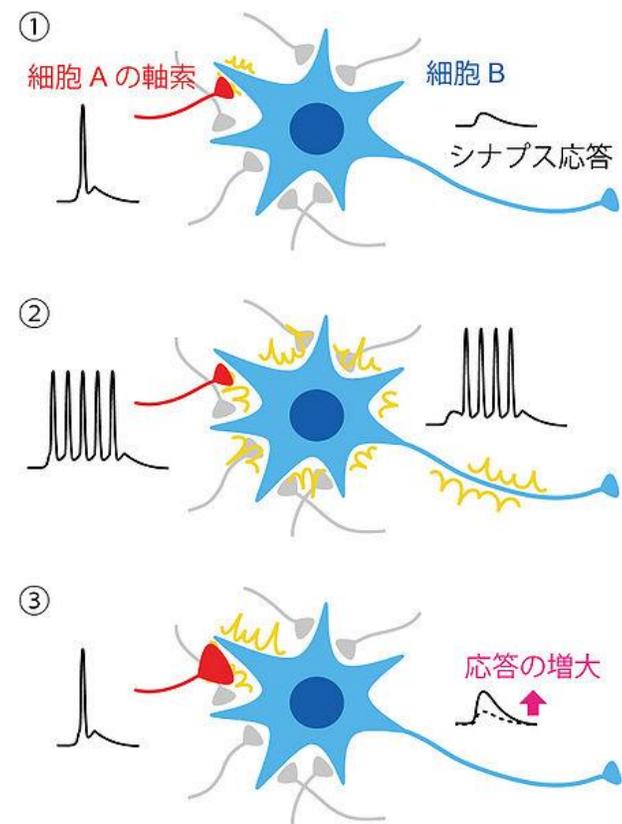
「大脳皮質の局所神経回路 - 脳科学辞典」窪田芳之、川口泰雄
<https://bsd.neuroinf.jp/wiki/大脳皮質の局所神経回路>

シナプスの学習

シナプスはニューロン間の結合部位、
結合の強さ(信号の伝えやすさ)で情報を記憶

脳内でバックプロパゲーションが行われているとは
考えられていない

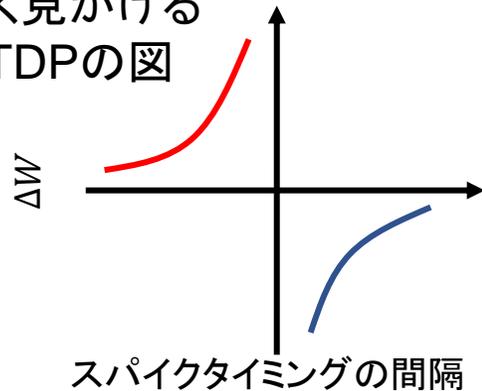
よく想定される学習則: ヘブ則
「ニューロンAの発火がニューロンBを発火させると
2つのニューロンの結合が強まる」



「ヘブ則 - 脳科学辞典」高橋直矢、池谷裕二、松木則夫
<https://bsd.neuroinf.jp/wiki/ヘブ則>

スパイクタイミング依存可塑性 (spike timing-dependent plasticity; STDP)

よく見かける
STDPの図



STDP の振る舞いは発火頻度にも依存

P. J. Sjöström, G. G. Turrigiano, and S. B. Nelson. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. *Neuron*, 32:1149–1164, 2001.
<https://www.ncbi.nlm.nih.gov/pubmed/11754844>

WTA 回路と組み合わせて
EM アルゴリズムを実現可能

STDP enable spiking neurons to detect hidden causes of their inputs
NIPS 2009
B Nessler, M Pfeiffer, W Maass
<http://papers.nips.cc/paper/3744-stdp-enables-spiking-neurons-to-detect-hidden-causes-of-their-inputs.pdf>

樹状突起で論理演算 [Koch, Poggio, Torre 1982]

Philos Trans R Soc Lond B Biol Sci. 1982 Jul 27;298(1090):227-63.
Retinal ganglion cells: a functional interpretation of dendritic morphology.
Koch C, Poggio T, Torre V.

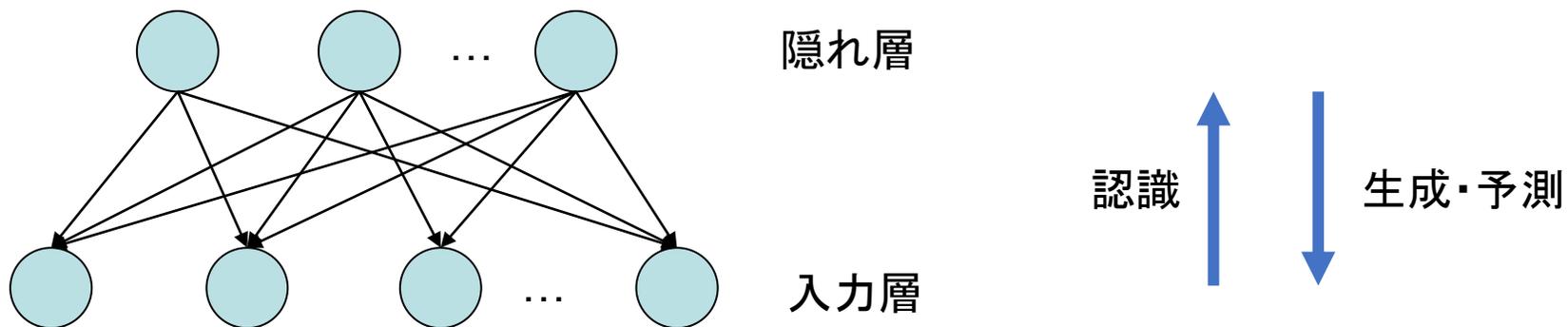
<http://cbcl.mit.edu/publications/ps/koch-poggio-torres-PRSL-1982.pdf>

- ミクロなレベルの現象は複雑で詳しくわかっていないことが多い
→ ミクロな知見の積み上げだけで脳の機能を人工的に再現
(whole brain emulation) できる日は遠い
- マクロな知見と計算論的モデルで補完することが重要

大脳皮質の 計算論的モデル

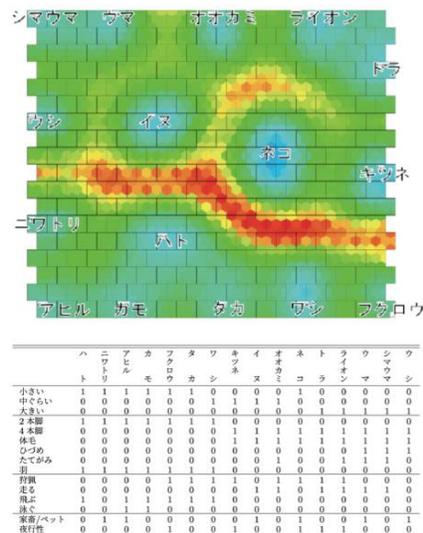
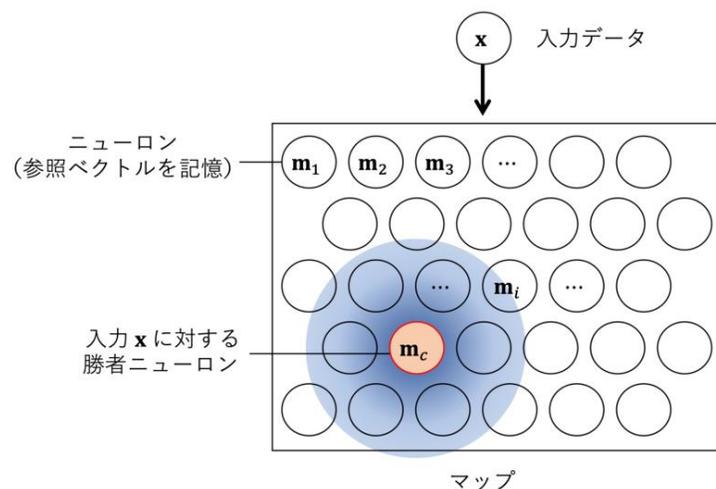
大脳皮質の多くのモデルはオートエンコーダー

- センサーからの入力を使って脳内の情報表現を教師なし学習
- 例：
 - 自己組織化マップ (Self-organizing map) [Kohonen 1995]
 - スパース符号化 (Sparse coding) [Olshausen and Field, Nature, 1996]
 - 予測符号化 (Predictive coding) [Rao and Ballard 1999]
 - BESOM [Ichisugi 2007]



自己組織化マップ (self-organizing map) [Kohonen 1995]

- 脳の一次視覚野のモデル ([Linsker 1988] など) を工学的に扱いやすくしたもの
- 高次元の入力ベクトルを低次元に圧縮
- 勝者総どり (winner-take-all) 型の学習 (競合学習)



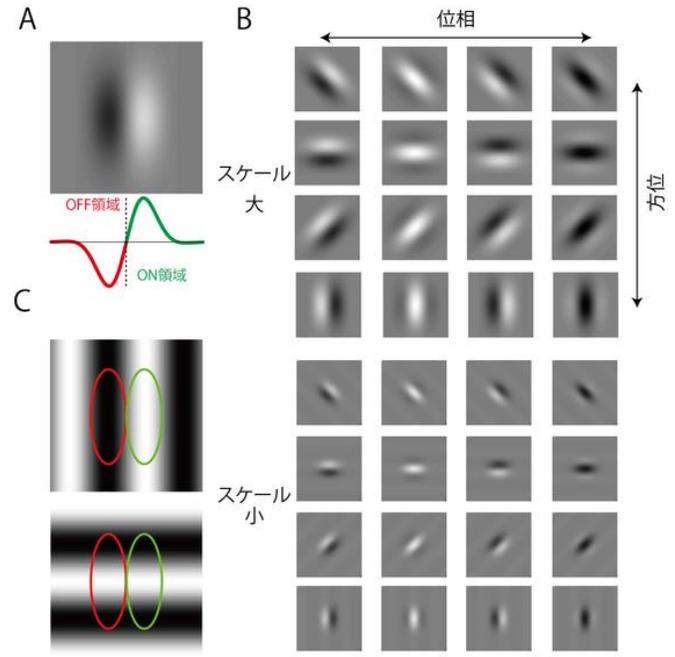
「自己組織化マップ - 脳科学辞典」古川 徹生
<https://bsd.neuroinf.jp/wiki/自己組織化マップ>

“Emergence of simple-cell receptive field properties by learning a sparse code for natural images”.
Bruno A. Olshausen and David J. Field
Nature, 381:607-609 (1996)

Sparse coding [Olshausen and Field, Nature, 1996]

- ・ 大脳皮質一次視覚野 単純型細胞のモデル
- ・ できるだけスパースに重み付けされた基底ベクトルの線形和で入力ベクトルを近似
- ・ 不思議な形をした受容野を単純なアルゴリズムで再現

脳の一次視覚野単純型細胞の受容野構造



自然画像に対してスパース符号化によって得られた基底局所性、方位選択性、空間周波数選択性がある

「受容野 - 脳科学辞典」田中 宏喜
<https://bsd.neuroinf.jp/wiki/受容野>

いろいろな「予測符号化」モデル

**画像圧縮技術としての
予測符号化**
= 予測値との差分を圧縮

MPEGなどで
使われている

網膜が予測符号化をしているとい
う論文もある
「Dynamic predictive coding by the retina |
Nature」
<https://www.nature.com/articles/nature03689>

知覚に関する仮説としての予測符号化

= 知覚は観測と予測を統合して行われる

知覚 観測 予測

≡ ベイズ脳 $P(X|Y) \propto P(Y|X)P(X)$

学習の目的としての予測符号化

= 予測誤差を最小化するように学習

≡ オートエンコーダー

Self-Organizing Map

Sparse coding

[Rao and Ballard 1999]

[Rao 2005]

HTM [George and Hawkins 2005]

BESOM [Ichisugi 2007]

双方向結合の役割の仮説

としての予測符号化 =
feedback 信号は予測誤差

実験事実との整合性には疑問

脳とベイズ

ベイズの効率化技術

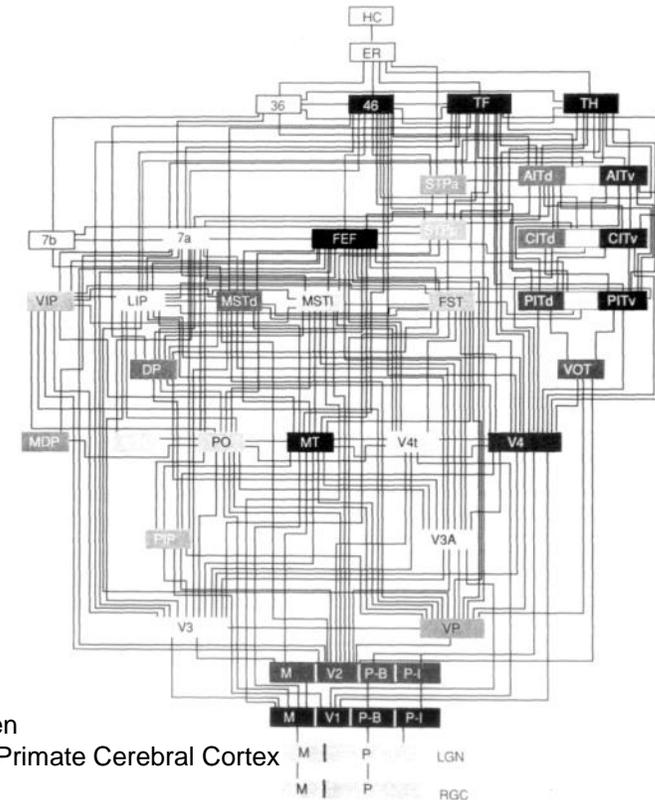
- 脳はベイズ推論を行っているらしい
- しかし、ベイズを使った計算は、普通は計算量がとても多い（指数関数的）
- 効率的推論を可能にする多くの技術
 - マルコフ連鎖モンテカルロ (MCMC)
 - ギブスサンプリング
 - 変分ベイズ
 - グラフィカルモデル
 - **ベイジアンネットワーク** 神経科学的妥当性が非常に高いと私が思うモデル
 - マルコフ確率場

大脳皮質と ベイジアンネット

大脳皮質とベイジアンネットの類似点

- トップダウンとボトムアップの非対称の接続
- 局所的かつ非同期的な情報のやり取りだけで動作
- 値が非負
- 情報が正規化される
- ヘブ則学習
- 文脈や事前知識に依存した認識
- ベイズに基づく動作

一杉裕志, 解説: 大脳皮質とベイジアンネット、
日本ロボット学会誌 Vol.29 No.5, pp.412--415, 2011.

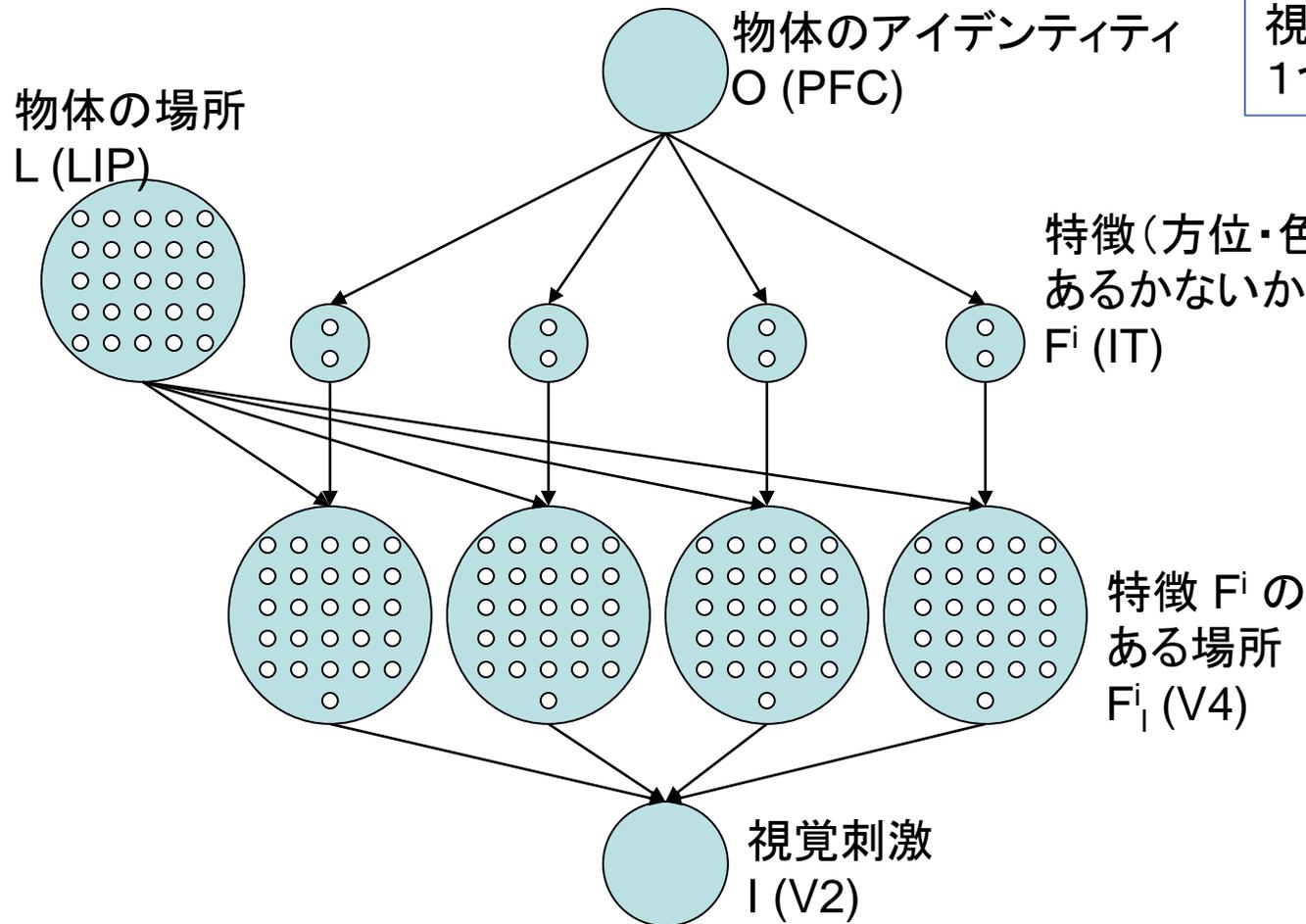


ベイジアンネットを使った 大脳皮質モデル

- 視覚野の機能、運動野の機能、解剖学的構造、電気生理学的現象などを説明
 - [Lee and Mumford 2003]
 - [George and Hawkins 2005]
 - [Rao 2005]
 - [Ichisugi 2007] [Ichisugi 2010] [Ichisugi 2011] [Ichisugi 2012]
 - [Rohrbein, Eggert and Korner 2008]
 - [Hosoya 2009] [Hosoya 2010] [Hosoya 2012]
 - [Litvak and Ullman 2009]
 - [Chikkerur, Serre, Tan and Poggio 2010]
 - [Hasegawa and Hagiwara 2010]
 - [Dura-Bernal, Wennekers, Denham 2012]

[Chikkerur, Serre, Tan and Poggio 2010]

のモデル



視覚野の腹側経路・背側経路というマクロな構造を模した1つベイジアンネットで、たくさんのモデルを同時に再現

他のモデルとの比較

	Proposed	[Bruce and Tsotsos, 2006]	[Zhang et al., 2008]	[Deco and Rolls, 2004]	[Ehinger et al., 2009]	[Fukushima, 1986]	[Hou and Zhang, 2007]	[Harel et al., 2007]	[Itti and Koch, 2001]	[Rao, 2005]	[Torralba, 2003]	[Walther and Koch, 2007]	[Wolfe, 2007]	[Yu and Dayan, 2005]
Biologically plausible	✓	✓	✓	✓	×	✓	×	×	✓	✓	✓	✓	✓	✓
Real world stimuli	✓	✓	✓	×	✓	×	✓	✓	✓	×	✓	✓	×	×
Pop-out	✓	✓	✓	×	✓	×	✓	✓	✓	×	✓	✓	×	×
Feature-based attention	✓	×	×	✓	✓	×	×	×	×	×	✓	✓	✓	✓
Spatial attention	✓	×	×	×	✓	×	×	×	×	✓	✓	✓	×	×
Parallel vs. serial search	✓	×	×	×	×	×	×	×	×	×	×	×	✓	×
Explicitly models ventral/parietal	✓	×	×	✓	×	×	×	×	×	✓	×	×	×	×

Table 3: The matrix compares the features of prior computational models

確率伝搬アルゴリズムと 解剖学的構造との対応

確率伝搬アルゴリズム [Pearl 1988]

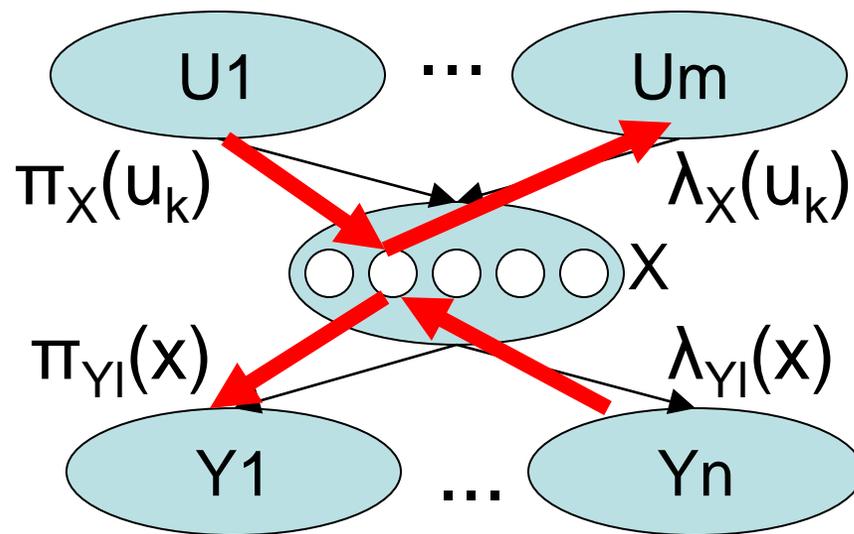
$$BEL(x) = \alpha \lambda(x) \pi(x)$$

$$\pi(x) = \sum_{u_1, \dots, u_m} P(x | u_1, \dots, u_m) \prod_k \pi_X(u_k)$$

$$\lambda(x) = \prod_l \lambda_{Y_l}(x)$$

$$\pi_{Y_l}(x) = \pi(x) \prod_{j \neq l} \lambda_{Y_j}(x)$$

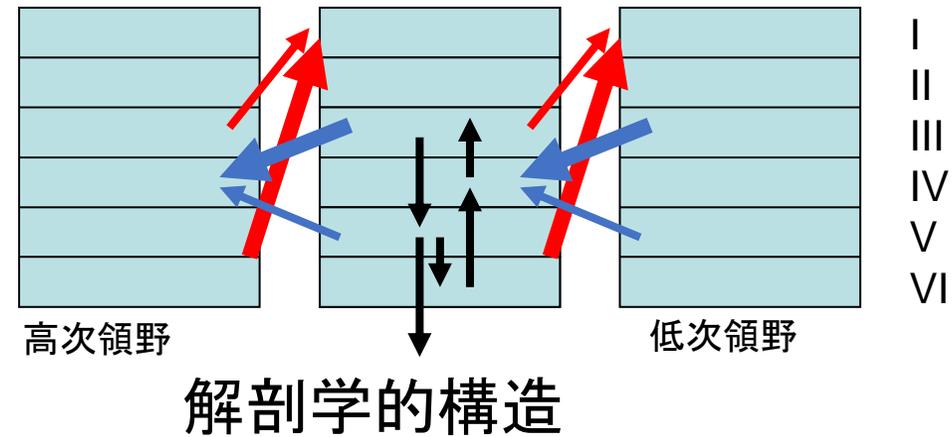
$$\lambda_X(u_k) = \sum_x \lambda(x) \sum_{u_1, \dots, u_m / u_k} P(x | u_1, \dots, u_m) \prod_{i \neq k} \pi_X(u_i)$$



近似確率伝搬アルゴリズムと コラム構造・6層構造との一致

[Ichisugi IJCNN 2007]

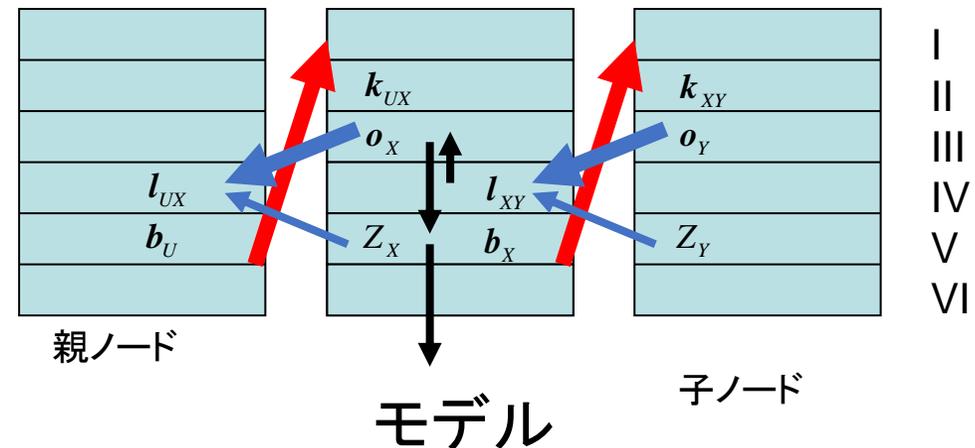
[Gilbert 1983]



[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.



確率伝搬と大脳皮質の対応 Ver. 2 [未発表]

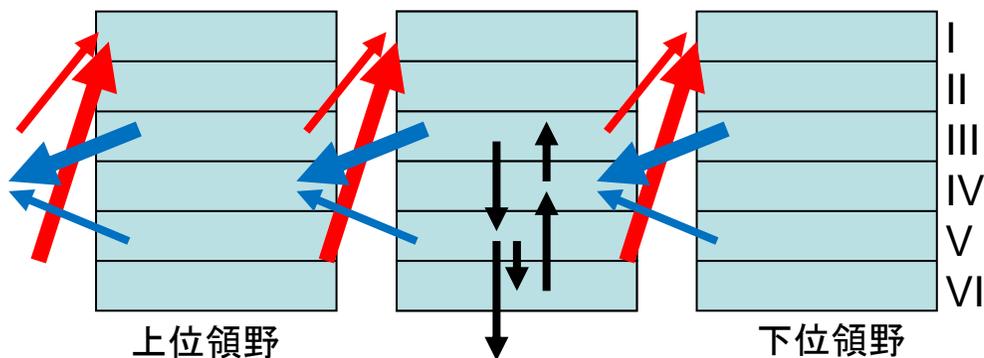
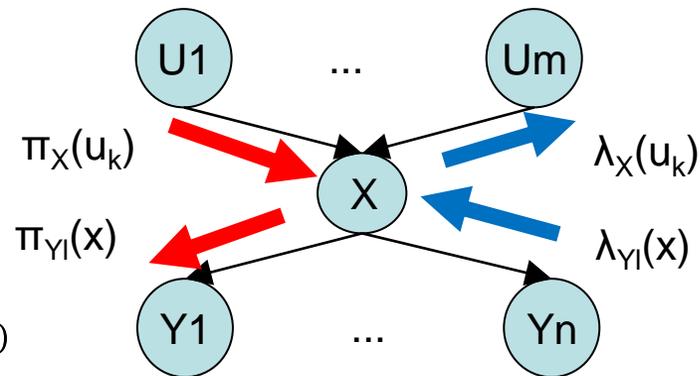
$$BEL(x) = \alpha \lambda(x) \pi(x)$$

$$\pi(x) = \sum_{u_1, \dots, u_m} P(x|u_1, \dots, u_m) \prod_k \pi_{X(u_k)}$$

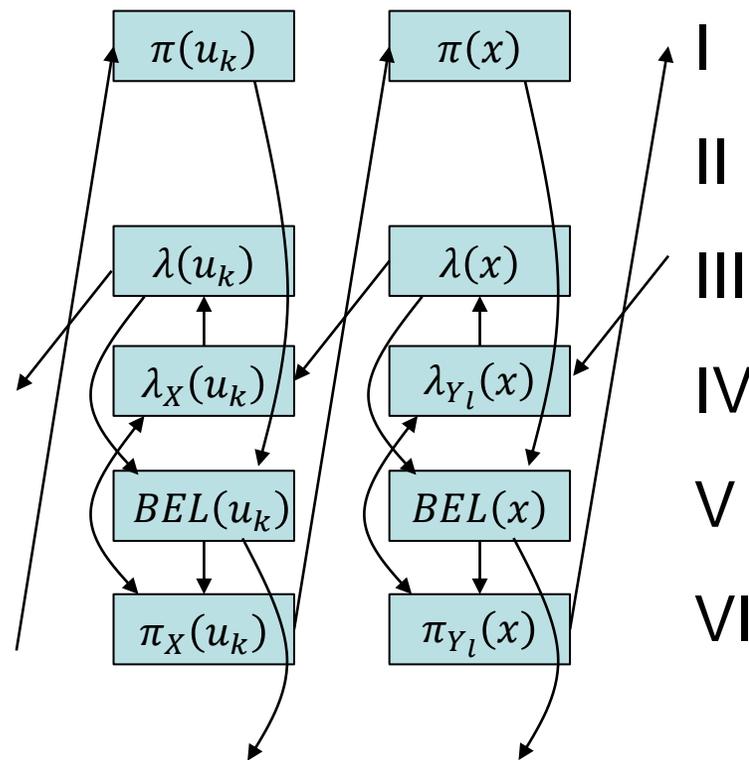
$$\lambda(x) = \prod_l \lambda_{Y_l}(x)$$

$$\pi_{Y_l}(x) = BEL(x) / \lambda_{Y_l}(x)$$

$$\lambda_X(u_k) = \sum_x \lambda(x) \sum_{u_1, \dots, u_m \setminus u_k} P(x|u_1, \dots, u_m) \prod_{i \neq k} \pi_{X(u_i)}$$



解剖学的構造



大脳皮質のベイジアンネットモデルのまとめ

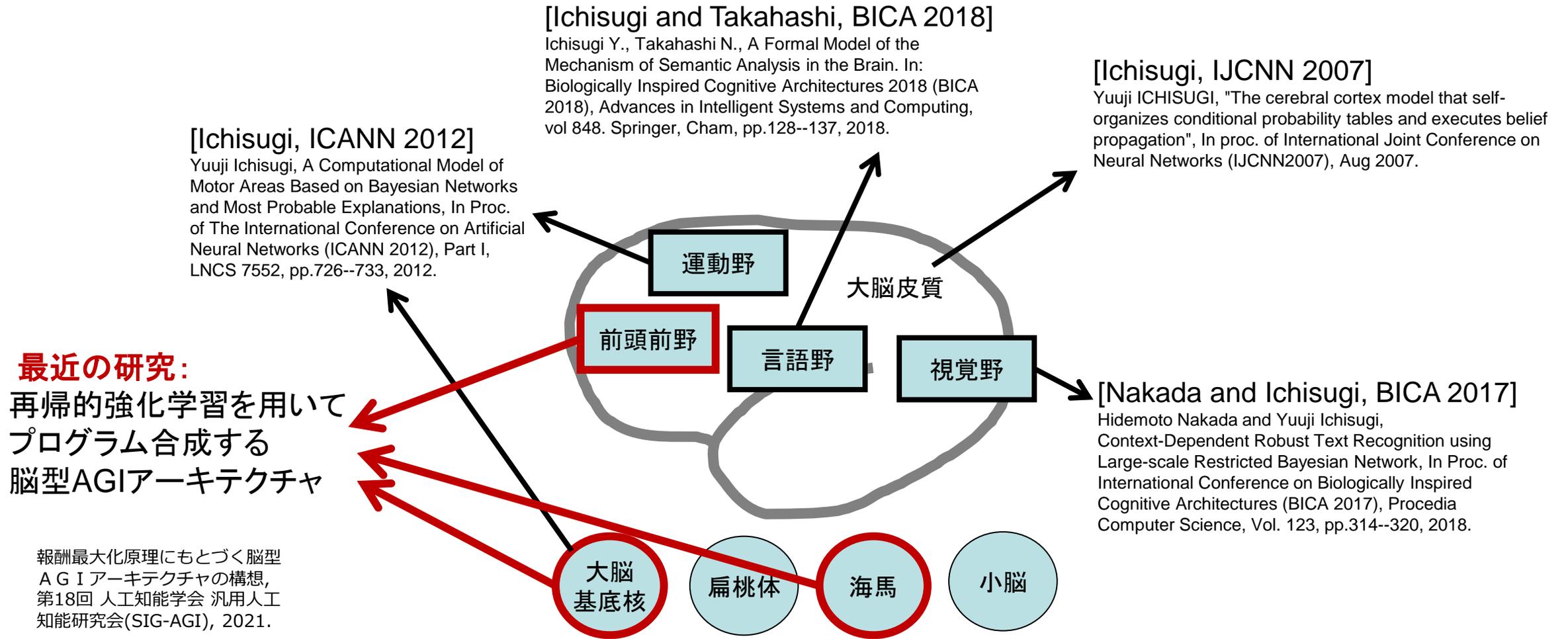
- 大脳皮質はおそらく、ある種のベイジアンネット
- 過去のたくさんのモデルがベイジアンネットを核として1つの万能モデルに統合されつつあった
- 大規模シミュレーションが難しく研究が途絶えている状況
(対策を検討中)

より詳しい情報:「脳とベイジアンネット」

<https://staff.aist.go.jp/y-ichisugi/besom/j-index.html>

私の最近の研究と 大脳皮質との関係

私の研究の長期的目標： 脳全体のアーキテクチャに学んだAGIの実現

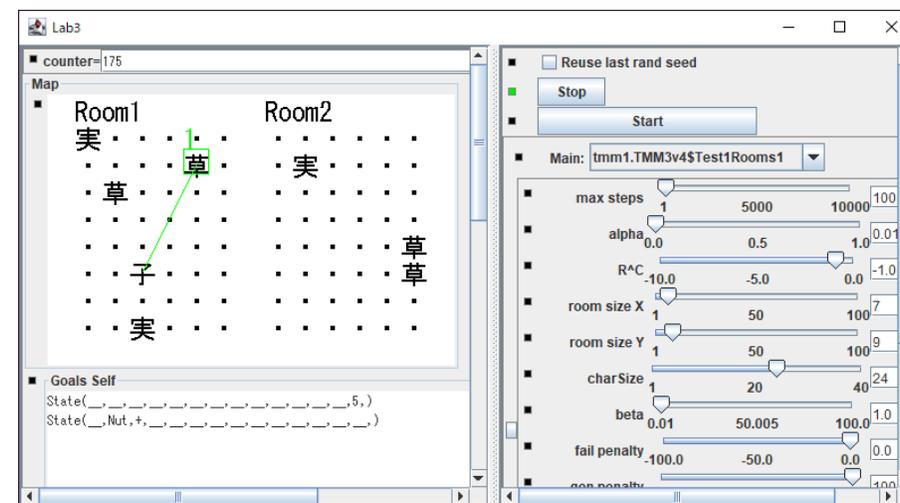


私の研究の中期的目標

- ローグライクゲームのような**実世界を極力単純化した環境**で脳型AGIのデモを動かす
- エージェントどうしが対話して生活し独自の文化を発展させるようなデモを目指す
- 中核技術
 - 大脳皮質モデル BESOM (未完成) [Ichisugi 2007]
 - 再帰的強化学習 RGoal [Ichisugi+ 2019]
 - プログラム合成対象言語 Pro5Lang [一杉+ 第20回 汎用人工知能研究会, 2022]



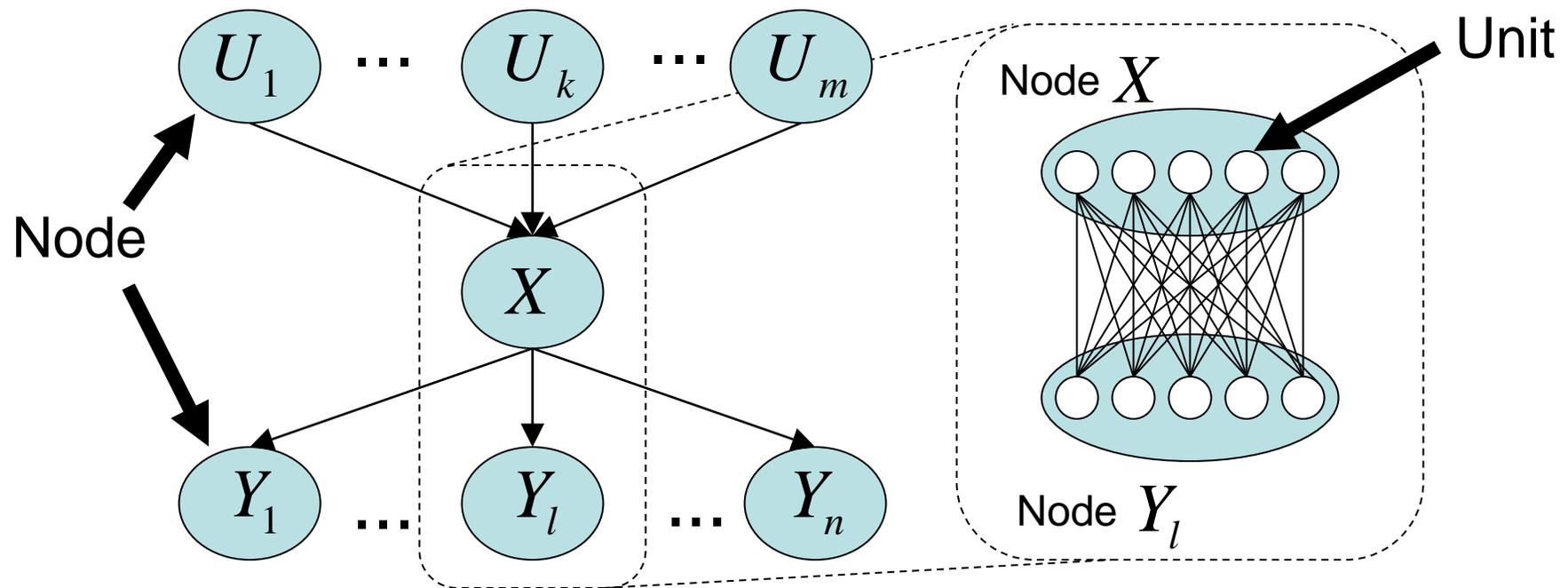
「ローグライクゲーム - Wikipedia」
<https://ja.wikipedia.org/wiki/ローグライクゲーム>



開発中のAGIエージェント実行環境

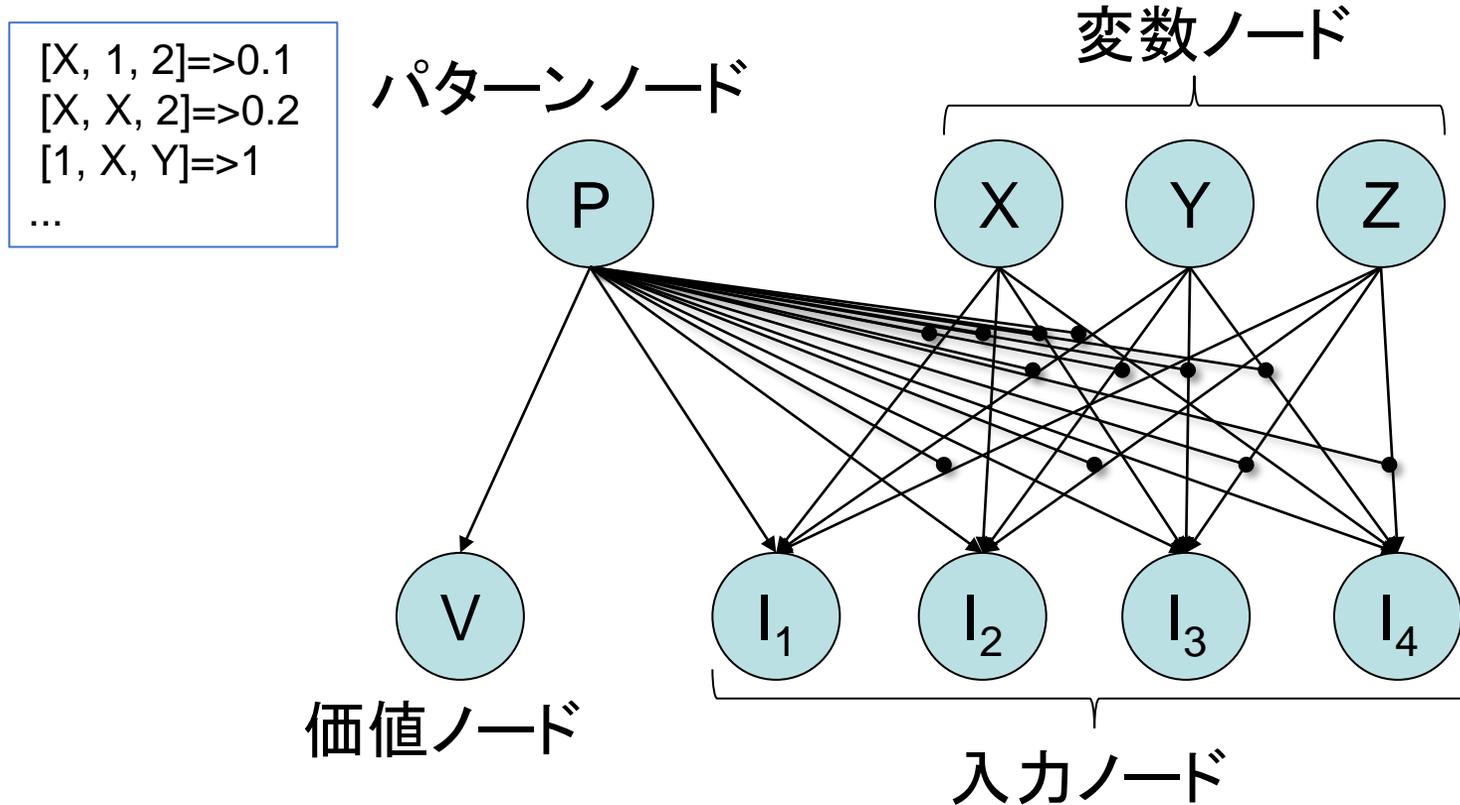
BESOM, 自己組織化マップ, ベイジアンネットワーク, 大脳皮質の構成要素の間の対応

BESOM	自己組織化マップ	ベイジアンネットワーク	大脳皮質
ノード	競合層	ノード (確率変数)	マクロコラム
ユニット	入力ベクトルの要素, 競合層のユニット	確率変数を取りうる値	ミニコラム
親ノード	入力層から見た競合層	親ノード (原因)	上位領野
子ノード	競合層から見た入力層	子ノード (結果)	下位領野
ユニットの出力	入力との類似度	事後確率	コラムの発火
結合の重み	参照ベクトルの要素	条件付確率	シナプスの重み



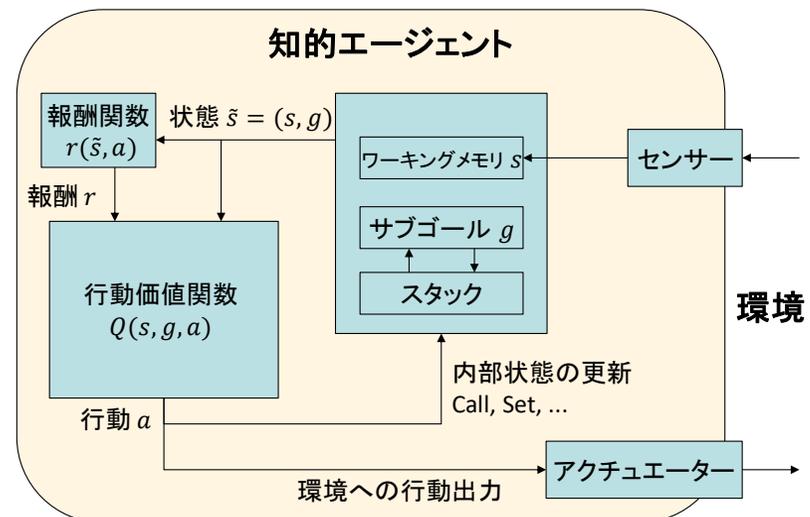
BESOM による統計的機械学習と記号処理との橋渡し

パターンマッチを行う BESOM の回路[一杉+ 第15回 汎用人工知能研究会, 2020]



再帰的強化学習 RGoal によるプログラム合成

行動価値関数の圧縮表現
= 報酬を最大化する
プログラム



再帰的強化学習 RGoal のアーキテクチャ

十分な性能を持つプログラム合成システムは AGI

Cf. AIXI, MagicHaskeller, DNC, DreamCoder, AlphaCode, ...

性能を上げるためには**合成対象言語**の設計が重要: Pro5Lang [一杉+ 第20回 汎用人工知能研究会, 2022]

一杉裕志

産業技術総合研究所 人工知能研究センター

研究テーマ：脳型汎用人工知能アーキテクチャの研究開発

人間のような知能を持つ機械を実現する最も確実な方法は、脳の動作原理を解明し、それを模倣することです。私は自ら再帰的に目標を設定する強化学習 [RGoal](#) を用いた [脳型汎用人工知能\(AGI\)アーキテクチャ](#) の構築を目指しています。また、[計算論的神経科学](#) の分野で知られている大脳皮質に関する知見をヒントにした [BESOM](#) と呼ぶ機械学習アルゴリズムの開発も行っています。BESOM は複数の機械学習技術（[自己組織化マップ](#)、[ベイジアンネット](#)、[独立成分分析](#)など）を組み合わせたもので、計算機上での効率的実行に適した有望な技術になると考えています。

解説等

- [全脳アーキテクチャ解明に向けて](#)

機械学習装置としての脳全体の機能解明に取り組む研究者を増やすための情報発信をしています。

- [脳とベイジアンネット](#)

「脳とベイジアンネットFAQ」やベイジアンネットの解説などを載せています。

- [脳を理解するための情報源メモ](#)

脳の情報処理原理の理解に必要な知識の良質な情報源を、独断で選んで紹介します。

- [BESOM \(ビーソム\) ブログ](#)

BESOMモデルに関連する研究者向けの情報発信をしています。

- [脳のアルゴリズムに関する取り急ぎメモ](#)

脳のアルゴリズムに関することを取り急ぎメモにまとめて公開しています。

研究構想の説明資料、発表資料、解説など

- 一杉裕志, 脳全体のアーキテクチャと対応付けたAGI研究 –前頭前野と前帯状皮質の機能が今後重要–, 汎用人工知能研究会(SIG-AGI) ワークショップ: AGI研究第3の波, パネル討論セッショントーク資料, 2022. [[slides](#)]
- 一杉裕志, 中田秀基, 高橋直人, 竹内泉, 佐野崇, プログラム合成対象言語 Pro5Lang のための行動価値関数圧縮アルゴリズム, 第22回 人工知能学会 汎用人工知能研究会(SIG-AGI), 2022. [[paper](#)] [[slides](#)]
- 一杉裕志, 中田秀基, 高橋直人, 竹内泉, 佐野崇,

最後に

脳と機械学習理論の両方を深く広く理解すれば
ヒトのようなAGIは作れるはず

私の最近の研究については下記ページをご覧ください

<https://staff.aist.go.jp/y-ichisugi/j-index.html>

オンラインでの議論も歓迎

以上