

脳は計算機科学者に解明される
のを待っている
— 機械学習器としての脳 —

言語処理学会第19回年次大会
(NLP2013) 招待講演

2013-03-13

産業技術総合研究所

ヒューマンライフテクノロジー研究部門

脳機能計測研究グループ

— 杉裕志

計算論的神経科学の現状

- 脳を情報処理機械に見立ててその機能を調べるといふ脳研究の一分野 (Wikipedia)
- 従来の「脳のモデル」
 - ある特定の実験データを説明するだけの限定的なもの。
- **ベイジアンネット**を使った大脳皮質のモデル
 - 脳の認識と学習の基本原理にせまる、適用範囲の広いモデルを指向。

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年より産業技術総合研究所 脳神経情報研究部門に異動。(現ヒューマンライフテクノロジー研究部門)

私の研究の目的

- 人間のように知能の高いロボットを作って、社会の生産性を飛躍的に向上させ、人々の生活をより豊かにすること。
- 役に立つロボットに必要な能力：
 - 赤ん坊のように自律的に知識を獲得できる。
 - 自然言語を通じて知識の伝達ができる。
 - 幅広い状況に知識を応用できる。
- そのための近道として、まず脳の情報処理原理を解明したい。

脳に関する誤解

- 脳についてまだほとんど何も分かっていない
→ **すでに膨大な知見がある。**
- 脳は計算機と全く違う情報処理をしている。
→ **脳はとても普通の情報処理装置である。**
- 脳はとても複雑な組織である。
→ **心臓等に比べれば複雑だが、意外と単純。**
- 計算量が膨大すぎてシミュレーションできない。
→ **ヒトの脳全体でも計算量的にすでに可能。**
- 労働力としては人間よりも高くつく。
→ **将来は人間よりもコストが低くなる。**

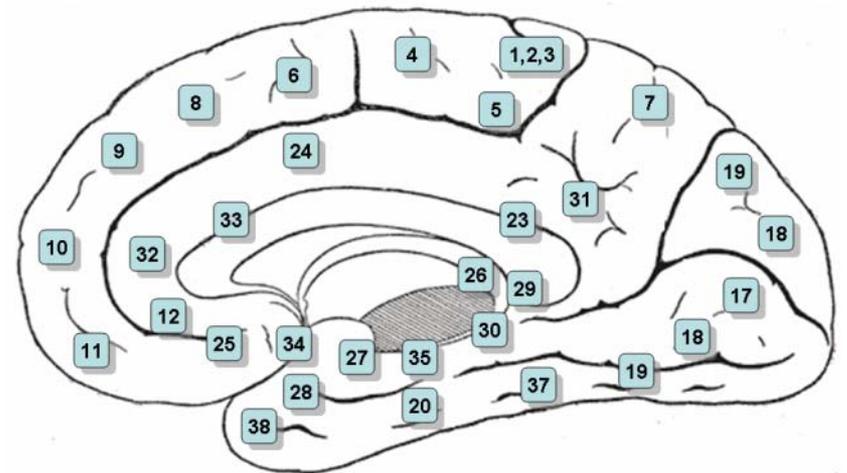
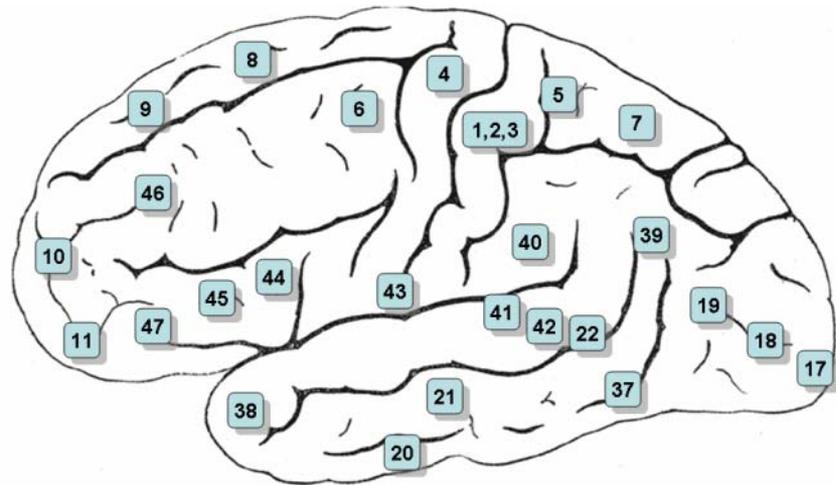
なぜ今、脳の理解が可能に？

- 脳の理解に必要な知識はこの十数年の間に揃いつつある。
 - 機械学習分野の要素技術の成熟
 - ベイジアンネットの教科書 [Pearl 1988]
 - 強化学習の教科書 [Sutton 1998]
 - 独立成分分析の教科書 [Hyvarinen 2001]
 - **大規模ニューラルネット Deep Learning [Hinton 2006]**
 - 「脳の10年」(1990～1999)以降の神経科学の急速な進歩
 - ドーパミンニューロンTD誤差の論文 [Schultz 1997]
 - V1のスパース符号化の論文 [Olshausen 1996]
 - **大脳皮質のベイジアンネットモデル [Lee and Mumford 2003] etc.**

大脳皮質に関する 神経科学的知見

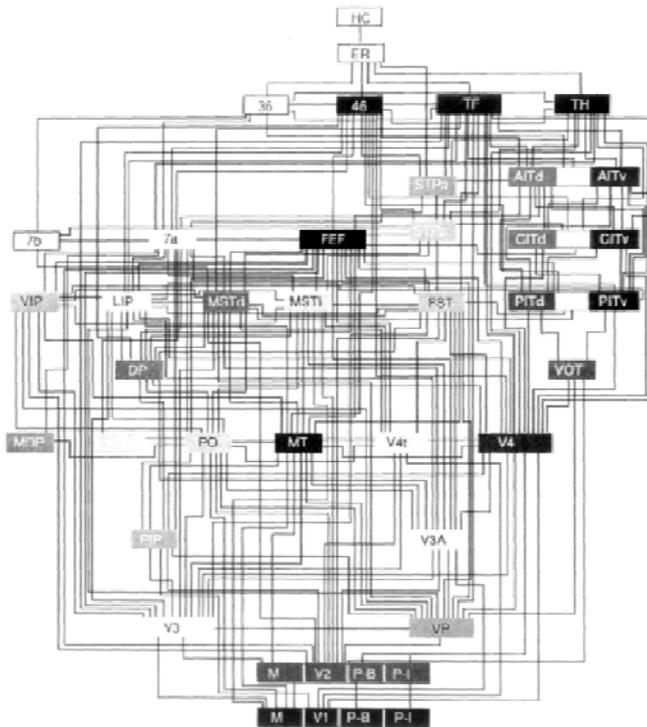
大脳皮質

- 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、...



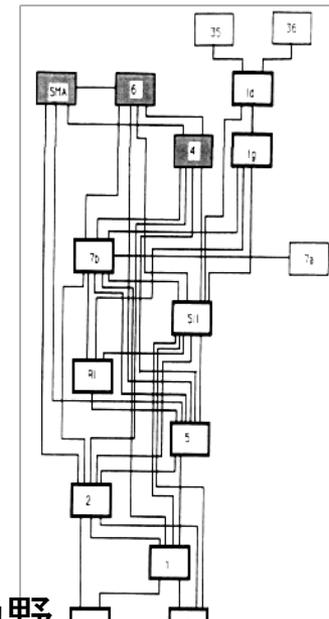
大脳皮質の領野

- 各領野の機能、接続構造はかなり明らかになりつつあるが、具体的な情報処理原理は不明。



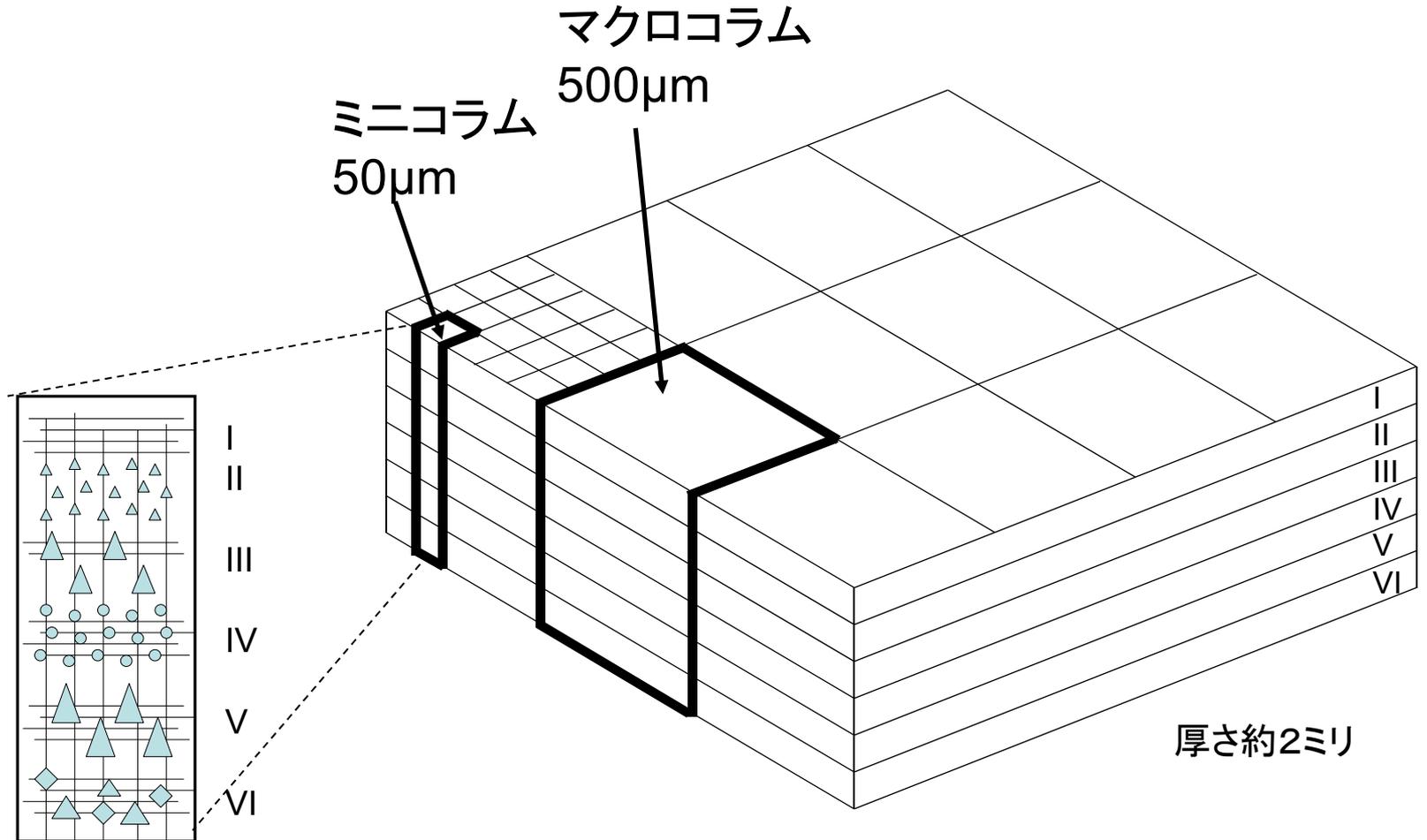
視覚系領野

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



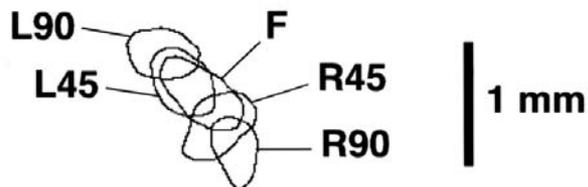
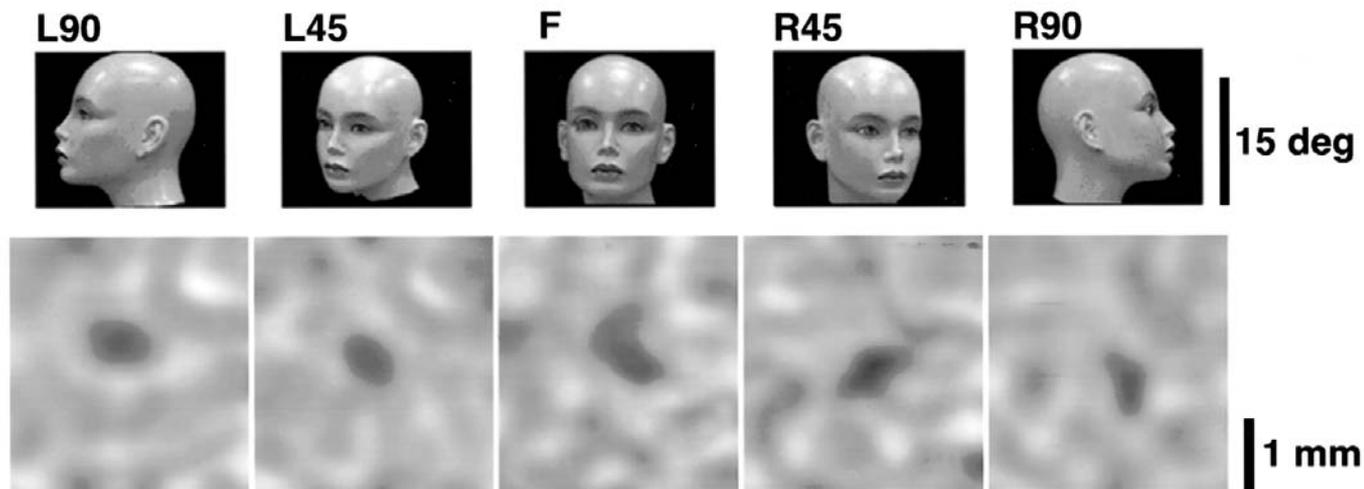
運動野

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

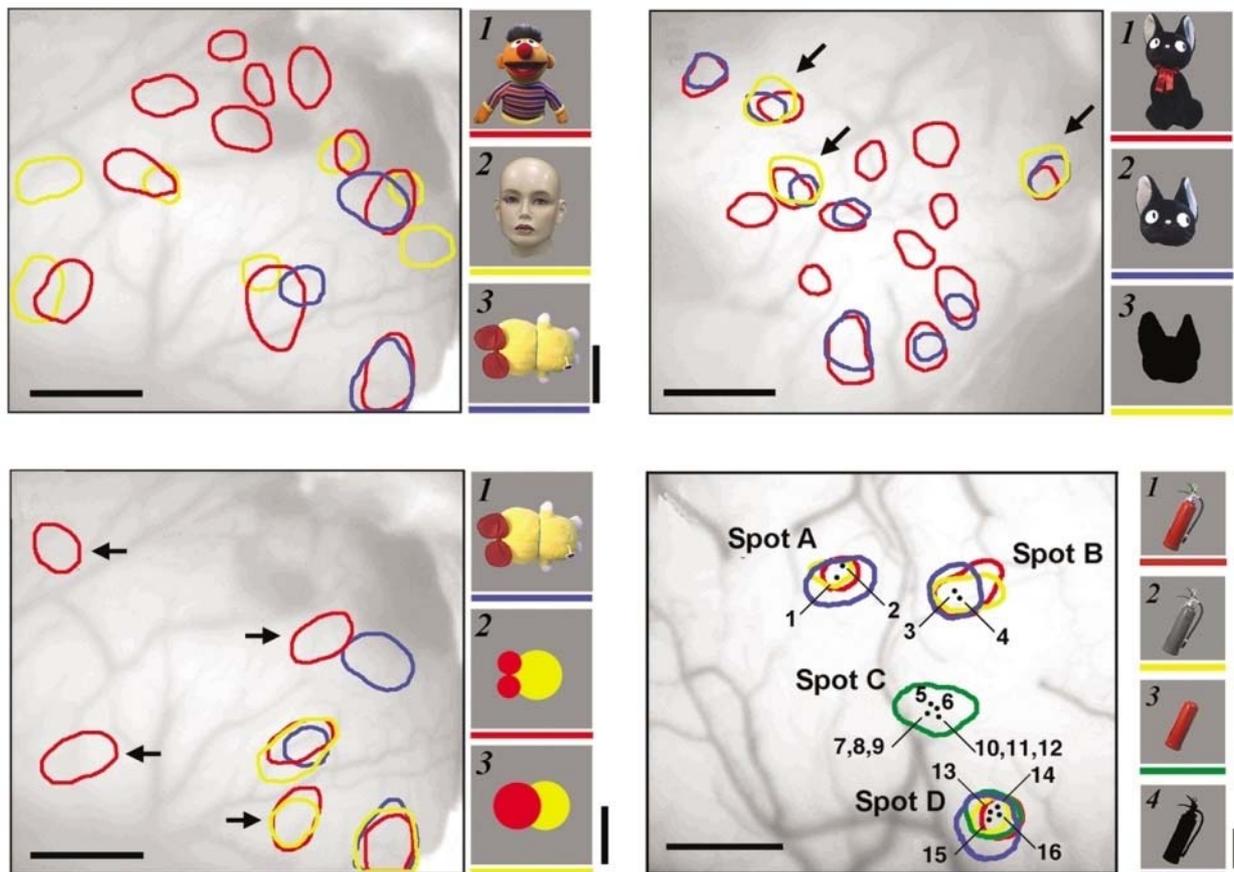
[Wang et al. 1996]



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

IT野の活動 [Tsunoda et al. 2001]

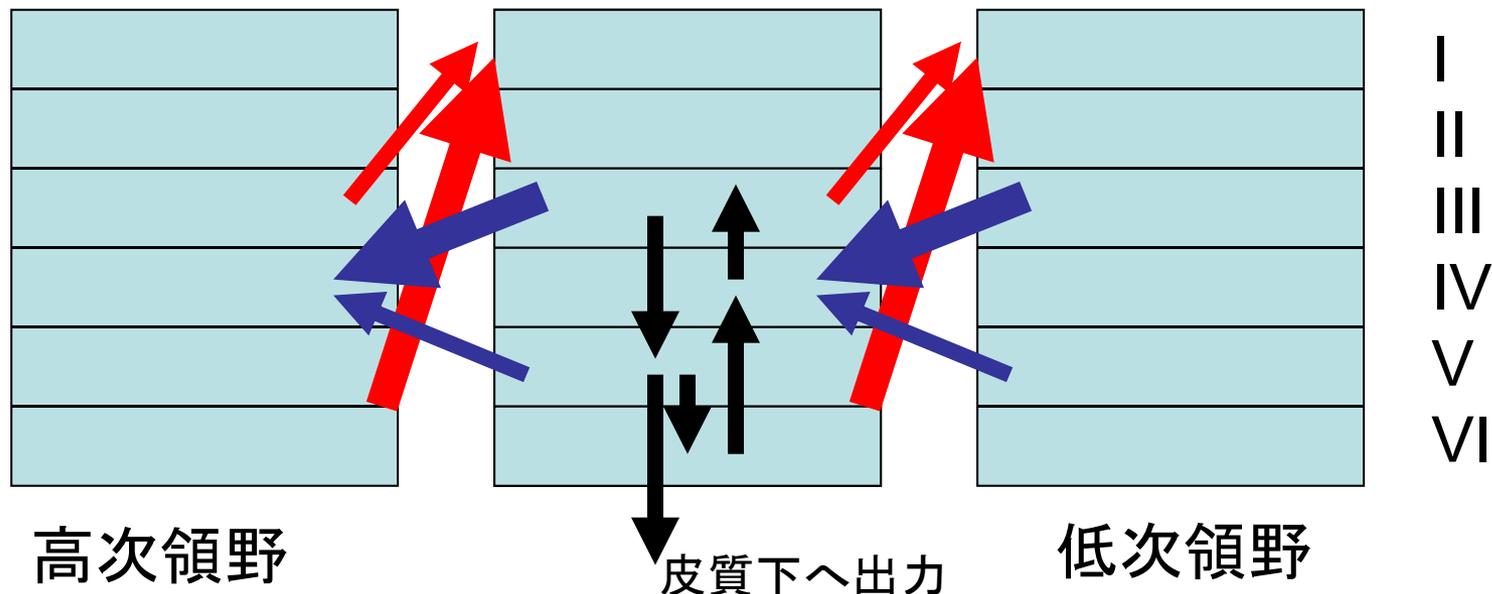


物体の形は脳内では「特徴コラム」の活動パターンで表現されているらしい。

Tsunoda K, Yamane Y, Nishizaki M, Tanifuji M (2001)
Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. Nat Neurosci 4:832-838.

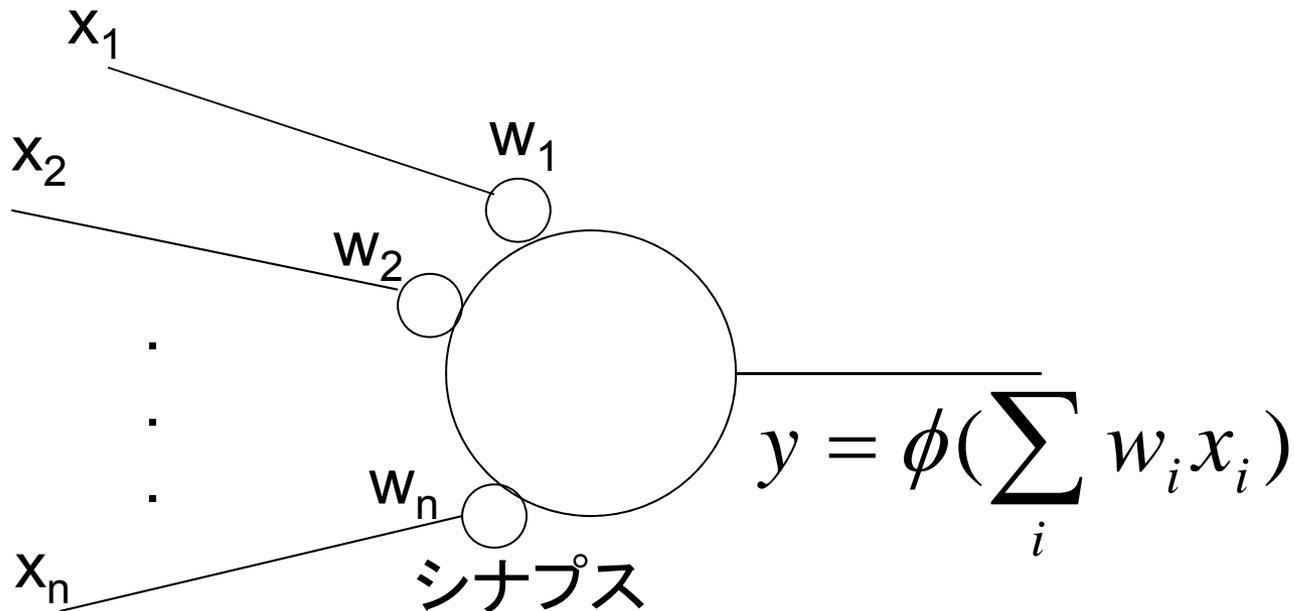
大脳皮質の解剖学的特徴 [Pandya and Yeterian 1985][Gilbert 1983]

- 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される。
 - その機能的意味は分かっていない。



ニューロン(神経細胞)

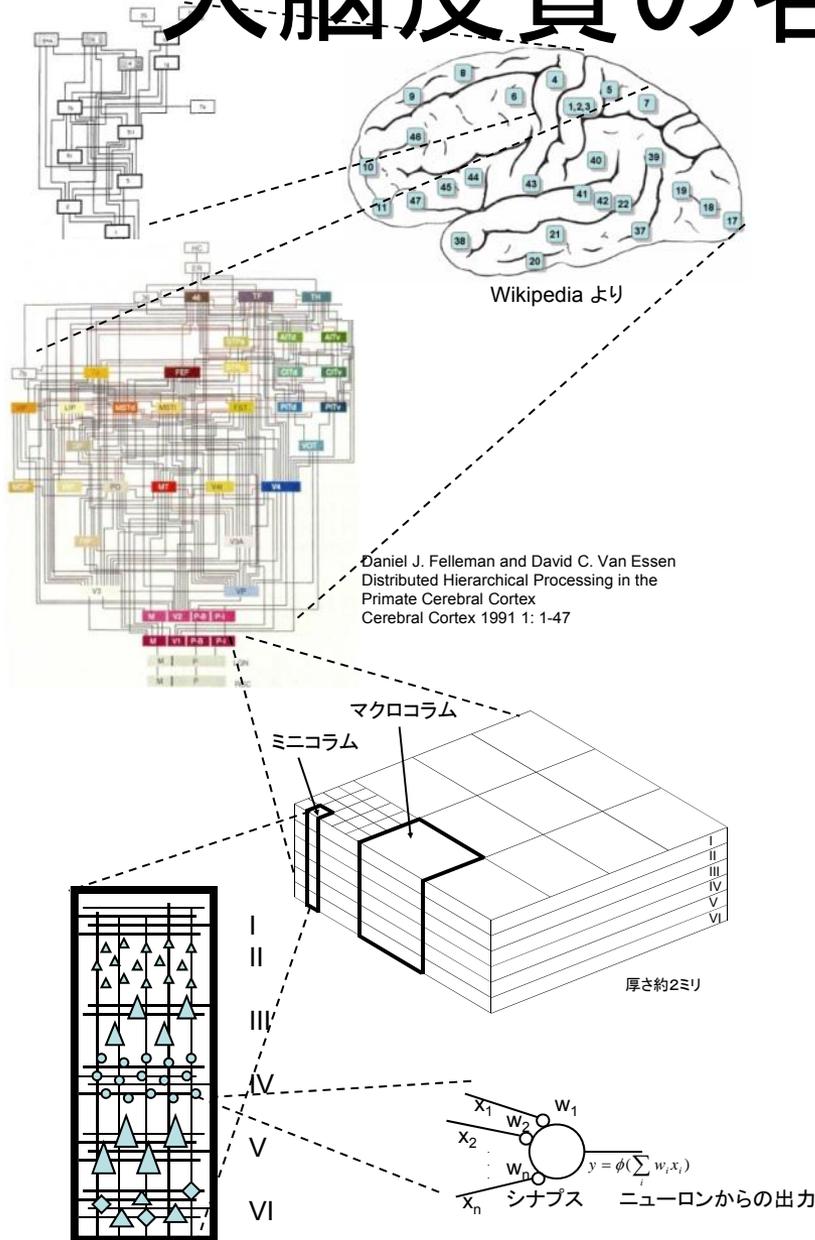
- ヒトの脳で1000億個。
- 内積計算のような簡単な演算しか行えない。
- シナプスが w という重みを学習。



ニューロンへの入力

ニューロンからの出力

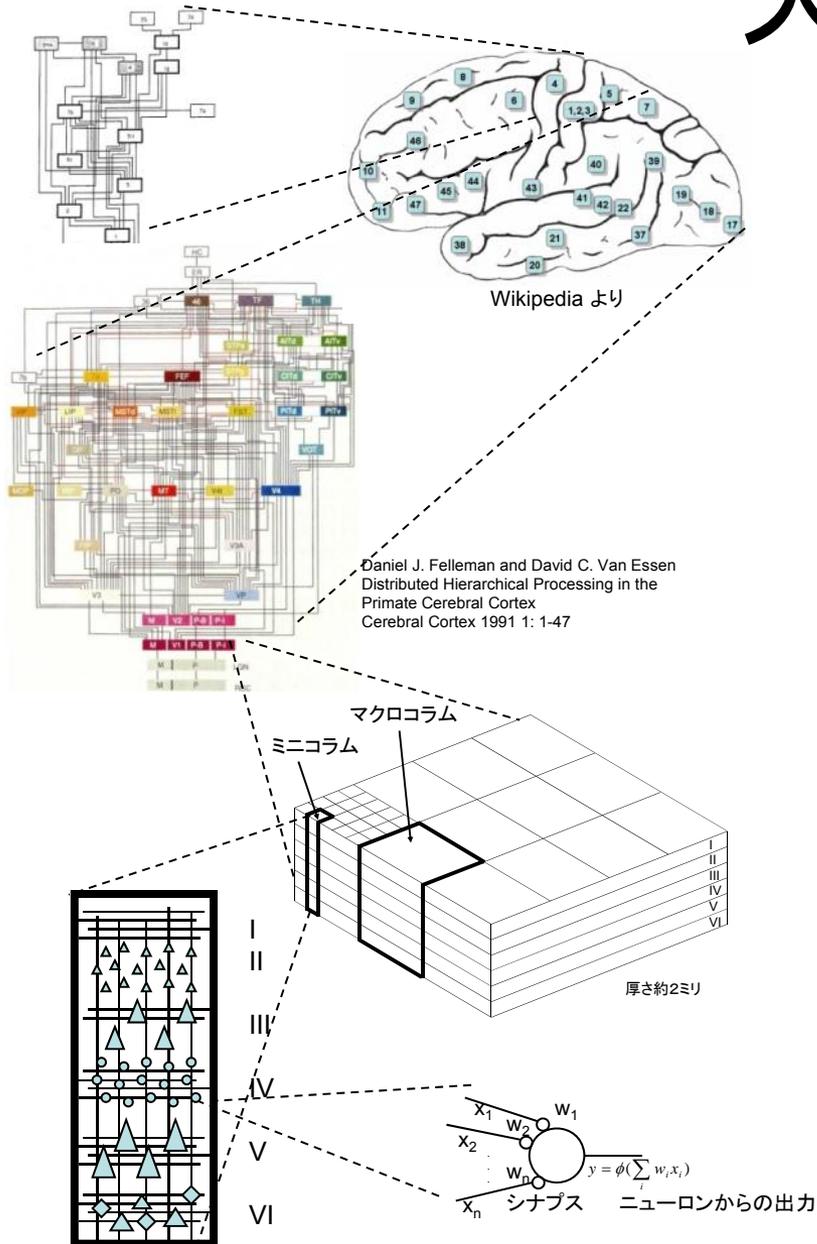
大脳皮質の各スケールでの構造



- 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

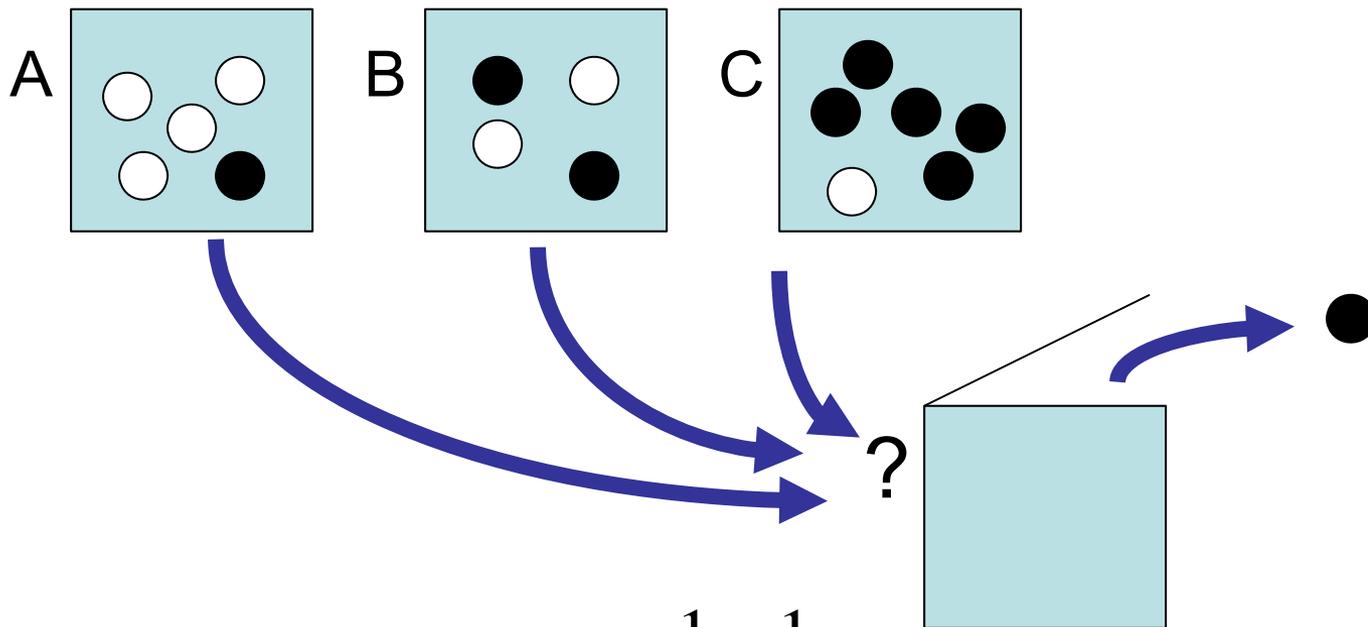
大脳皮質の不思議さ

- 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。



ベイジアンネット

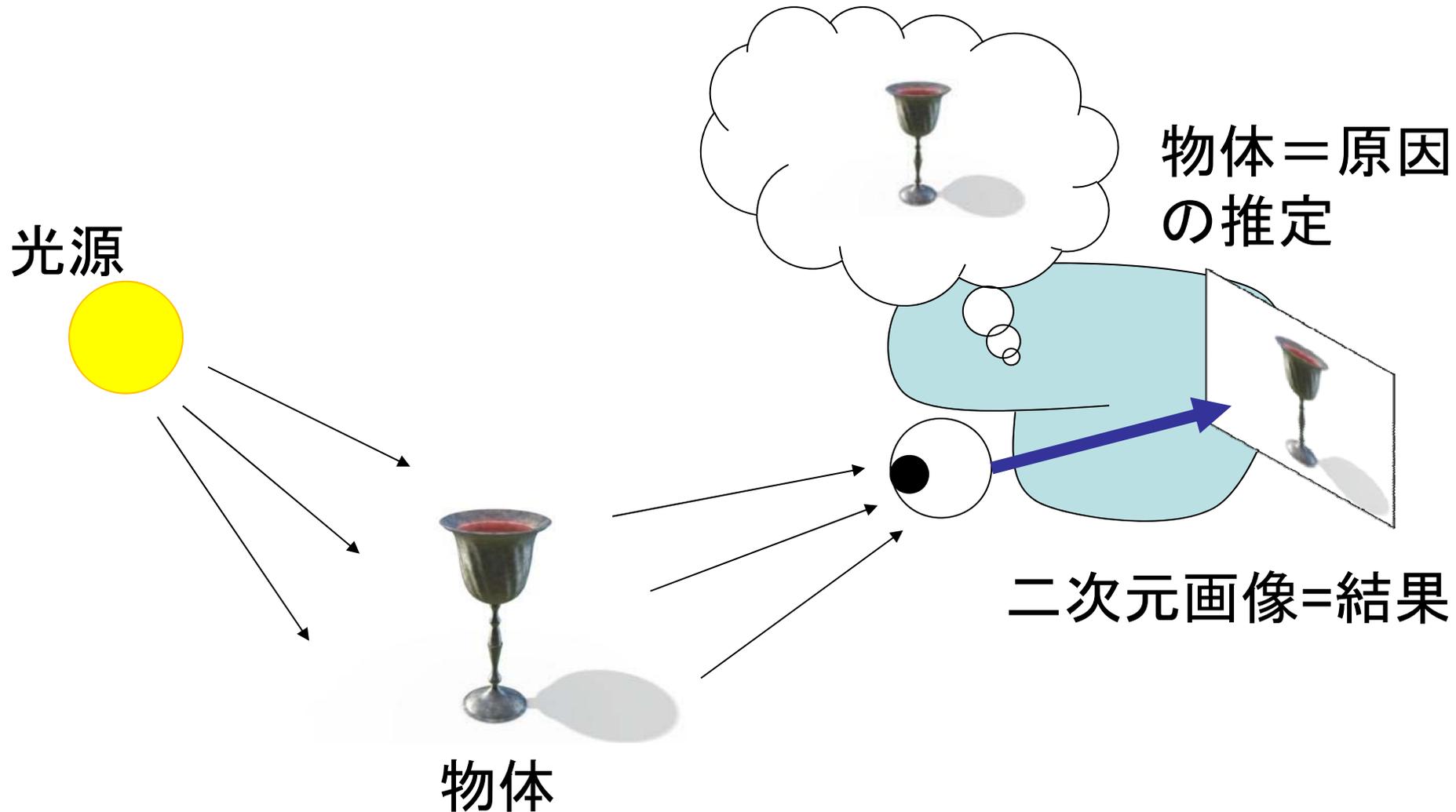
ベイズの定理=結果から原因を推定



$$P(\text{箱} = A | \text{色} = \text{黒}) = \frac{\frac{1}{5} \times \frac{1}{3}}{\frac{1}{5} \times \frac{1}{3} + \frac{2}{4} \times \frac{1}{3} + \frac{5}{6} \times \frac{1}{3}} = \frac{6}{46}$$

$$P(\text{箱} = B | \text{色} = \text{黒}) = \frac{15}{46} \quad P(\text{箱} = C | \text{色} = \text{黒}) = \frac{25}{46}$$

脳の認識も、 結果からの原因の推定

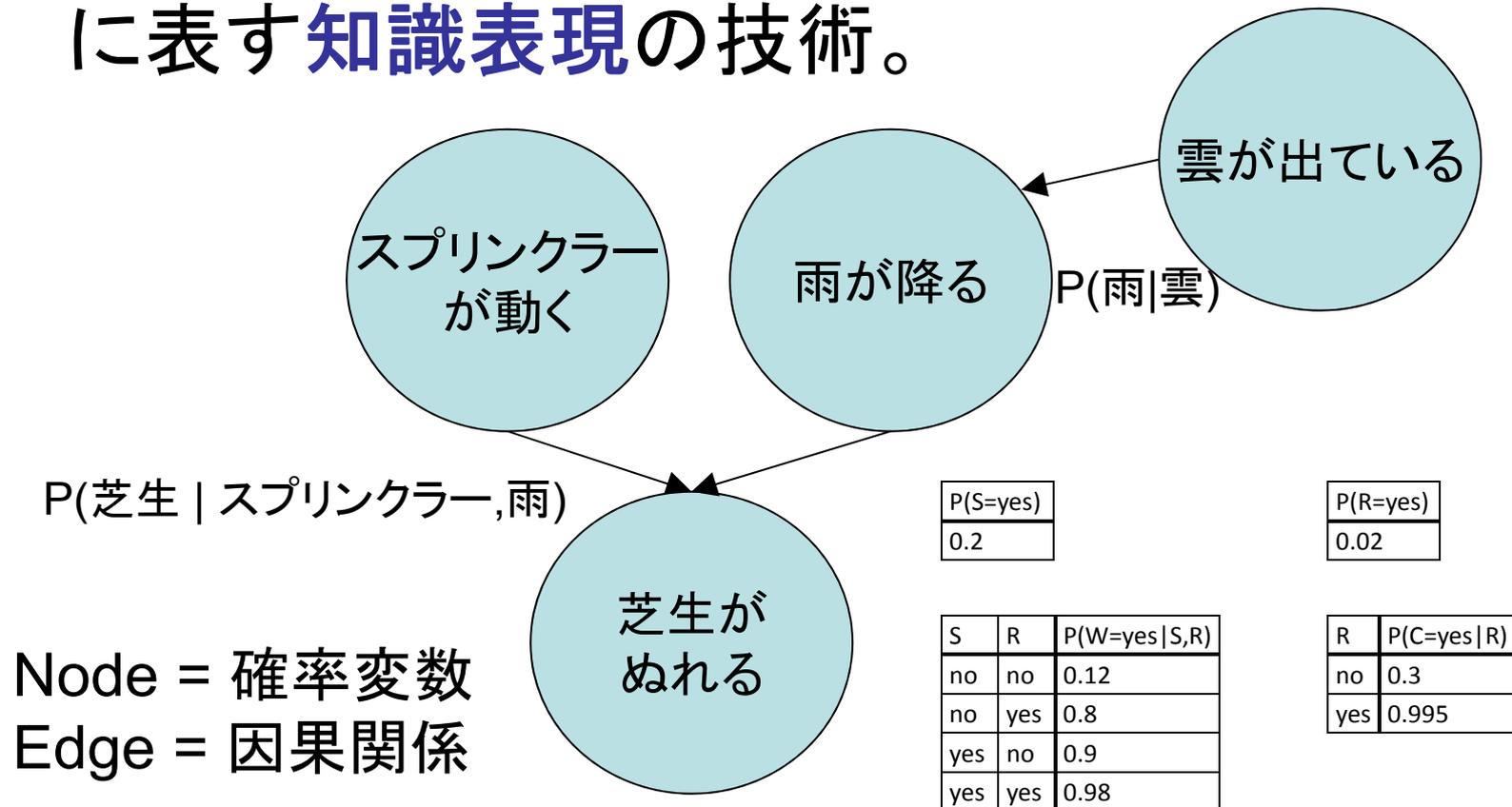


ベイズの効率化技術

- 脳はベイズを使っているらしい。
- 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**
 - マルコフ確率場

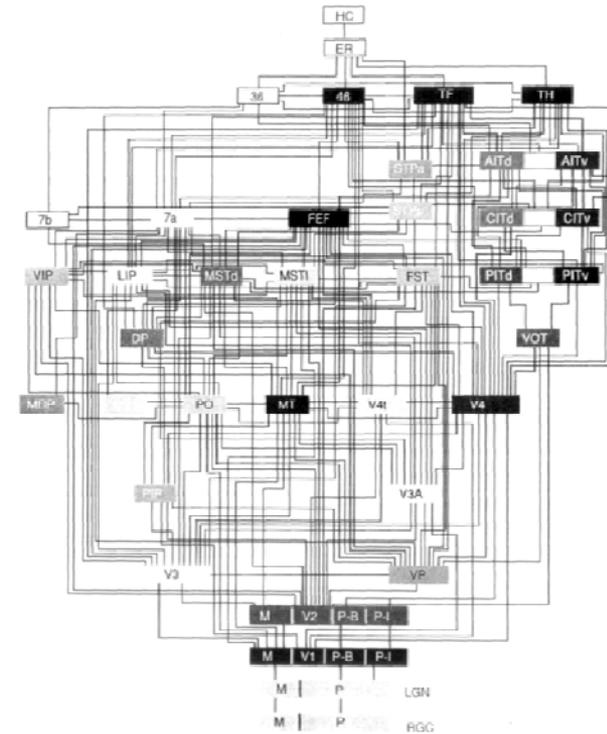
ベイジアンネットワーク [Pearl 1988] とは

- 脳の「直感・連想記憶」と似た働きをする。
- 確率変数の間の因果関係をグラフで効率的に表す知識表現の技術。



大脳皮質とベイジアンネットの類似点

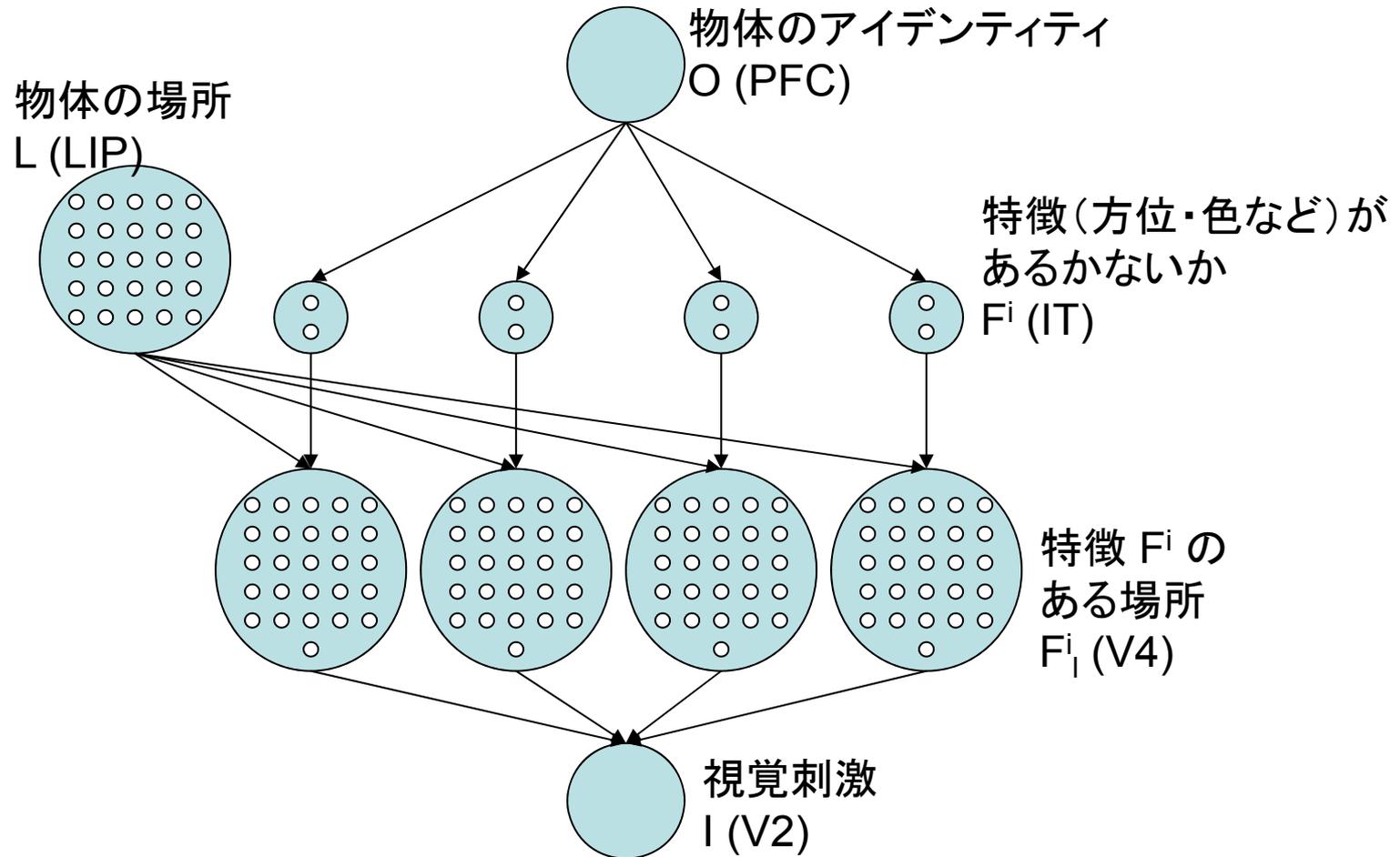
- トップダウンとボトムアップの非対称の接続
- 局所的かつ非同期な情報のやり取りだけで動作
- 値が非負
- 情報が正規化される
- ヘブ則学習
- 文脈や事前知識に依存した認識
- ベイズに基づく動作



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]

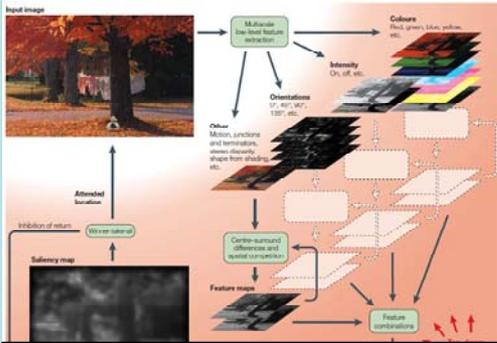
[Chikkerur, Serre, Tan and Poggio 2010] のモデル



既存の複数のモデルを統合

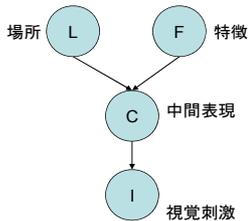
[Itti and Koch 2001]

Computational Modeling of Visual Attention



[Rao 2005] のモデル

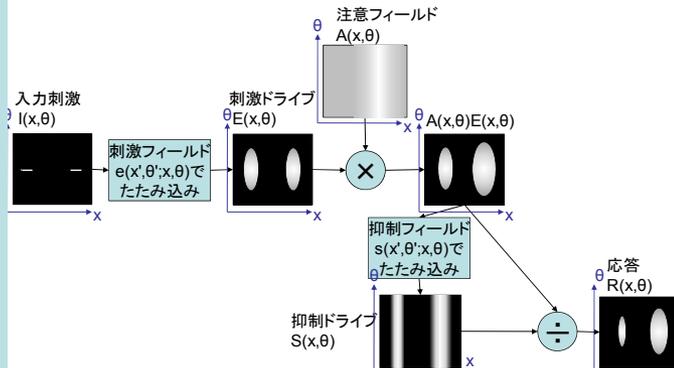
- 注意に関する電気生理実験の結果を説明するために Rao が考えたモデル



R. Rao. Bayesian inference and attention in the visual cortex. Neuroreport 16(16), 1843-1848, 2005.

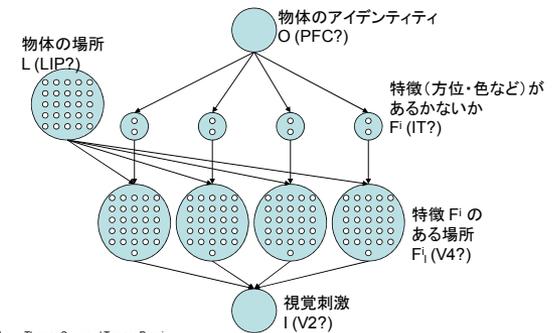
[Reynolds and Heeger 2009]

注意の正規化モデル



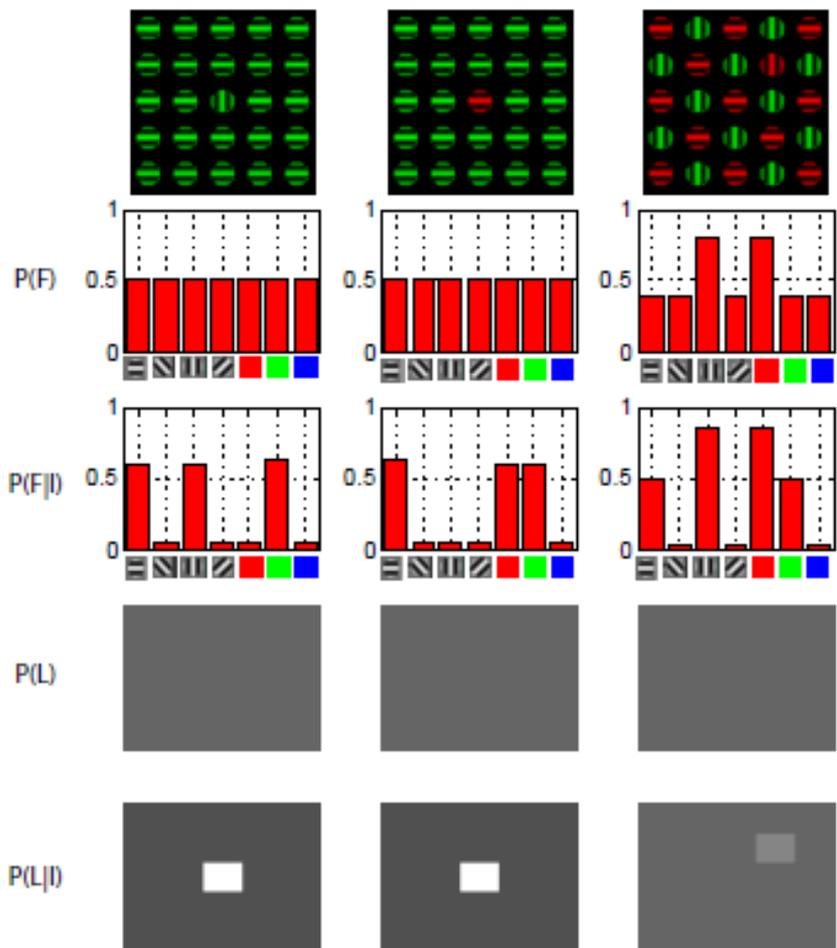
[Chikkerur, Serre, Tan and Poggio 2010]

のモデル

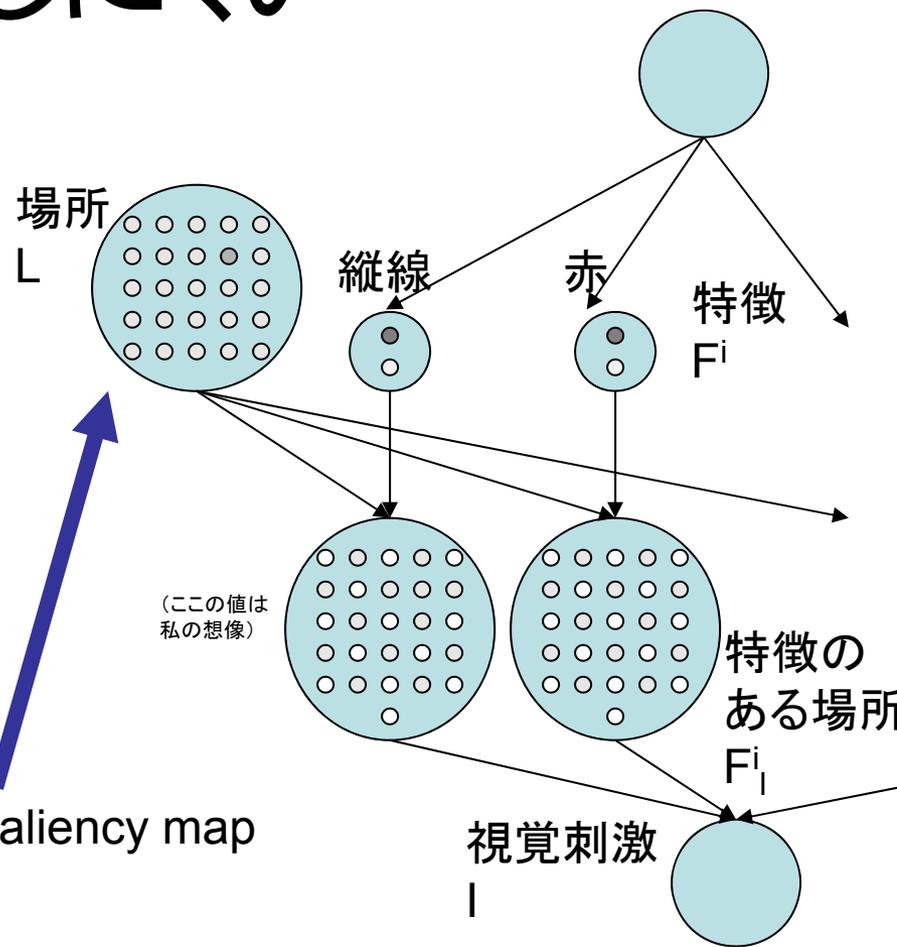


Sharat Chikkerur, Thomas Serre and Tomaso Poggio:
A Bayesian inference theory of attention: neuroscience and algorithms,
Computer Science and Artificial Intelligence Laboratory Technical Report,
MIT-CSAIL-TR-2009-047 CBCL-280, October 3, 2009

ちらかっている状況では 物体を探しにくい



(a) Parallel vs. serial search.

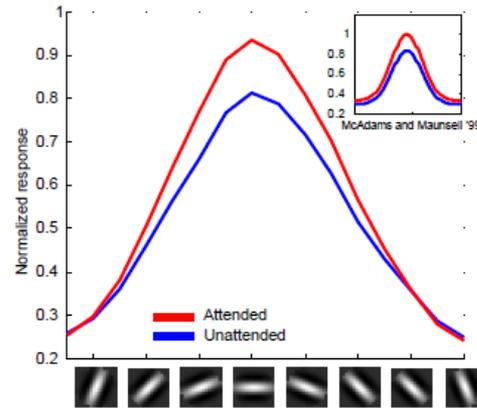


Saliency map

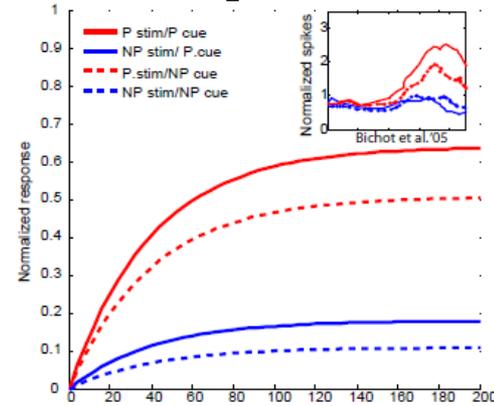
(この値は私の想像)

「縦線と赤」に注意を向けても、それがあつ場所がポップアウトしない。

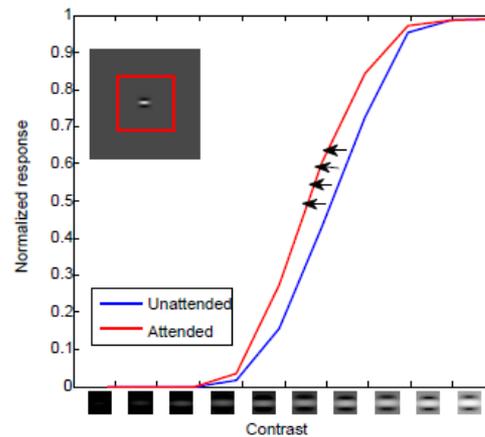
注意とニューロン応答の 関係 [Reynolds and Heeger 2009] の再現



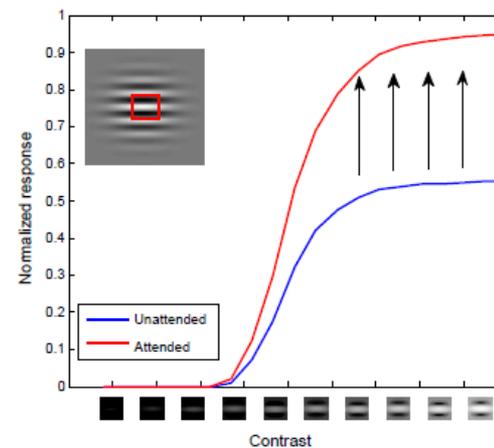
(a) Effect of spatial attention on tuning response.



(b) Effect of feature attention on neuron response.



(c) Contrast gain under attention.



(d) Response gain under attention.

事後確率

$$P(F_l^i | I)$$

がニューロン応答

確率伝播アルゴリズム [Pearl 1988]

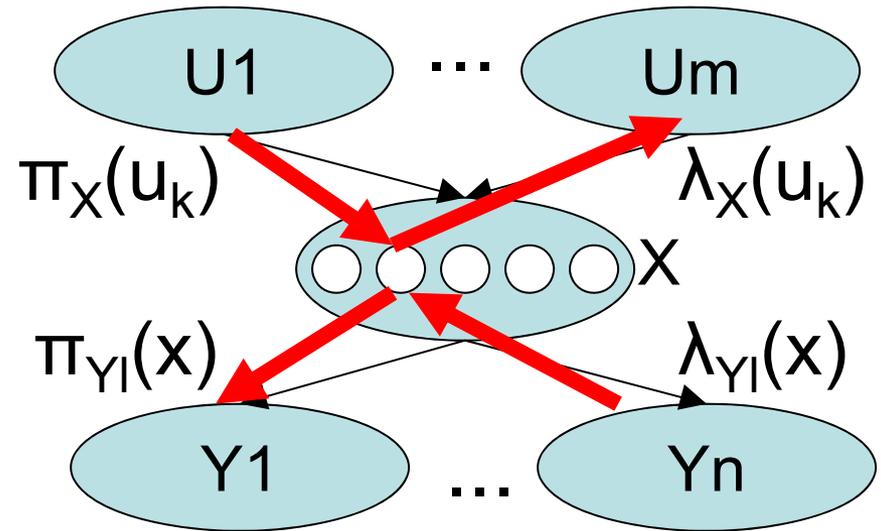
$$BEL(x) = \alpha \lambda(x) \pi(x)$$

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

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

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

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



近似確率伝播アルゴリズム [Ichisugi 2007]

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

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

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

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

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

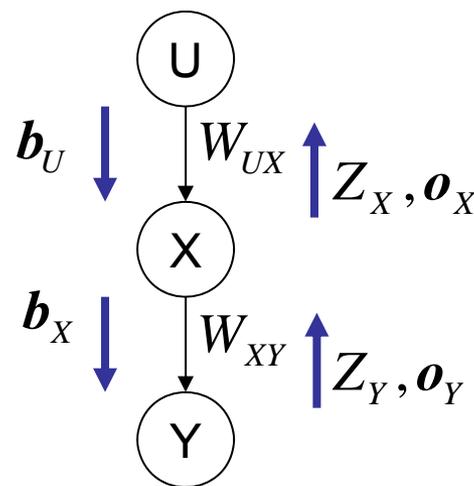
$$Z_X^{t+1} = \sum_i (r_X^{t+1})_i \quad (= \|r_X^{t+1}\|_1 = o_X^{t+1} \bullet p_X^{t+1})$$

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

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

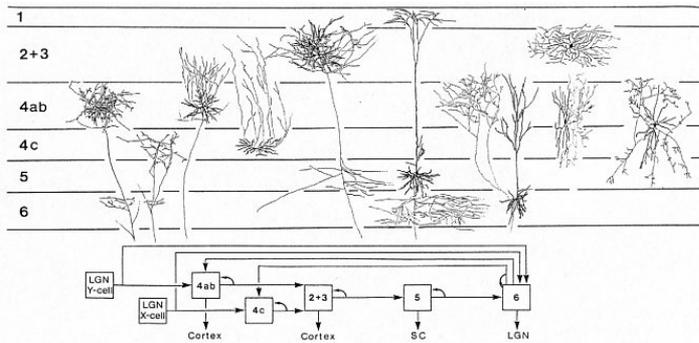
ただし、 $\mathbf{x} \otimes \mathbf{y} = (x_1 y_1, x_2 y_2, \dots, x_n y_n)^T$

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

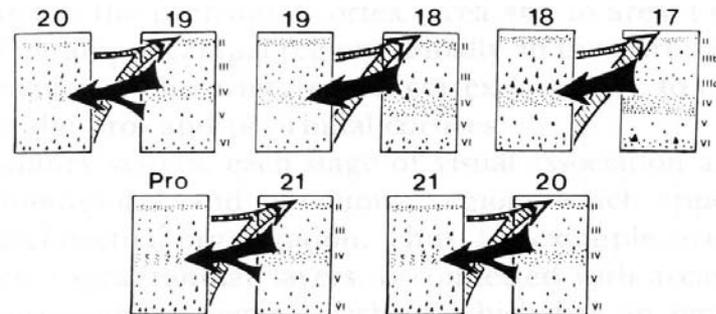


- ・神経回路で実現可能
- ・大規模化可能な計算量・記憶量

コラム構造・6層構造との一致



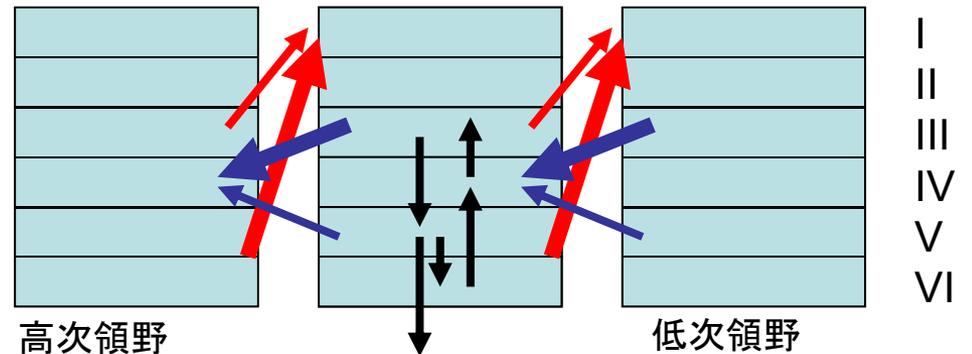
[Gilbert 1983]



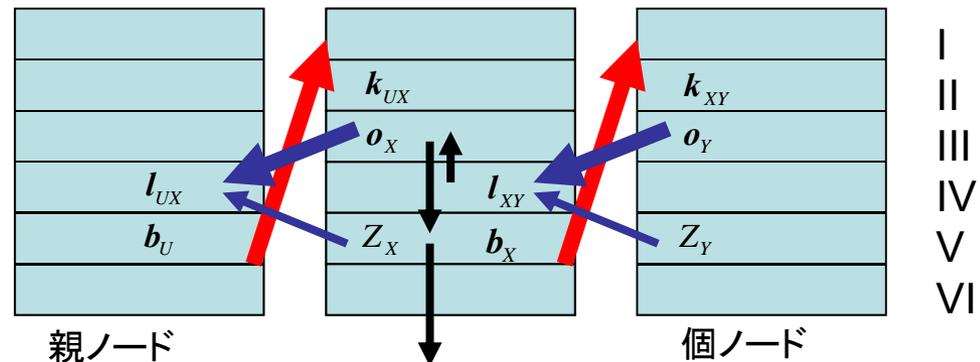
[Pandya and Yeterian 1985]

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.

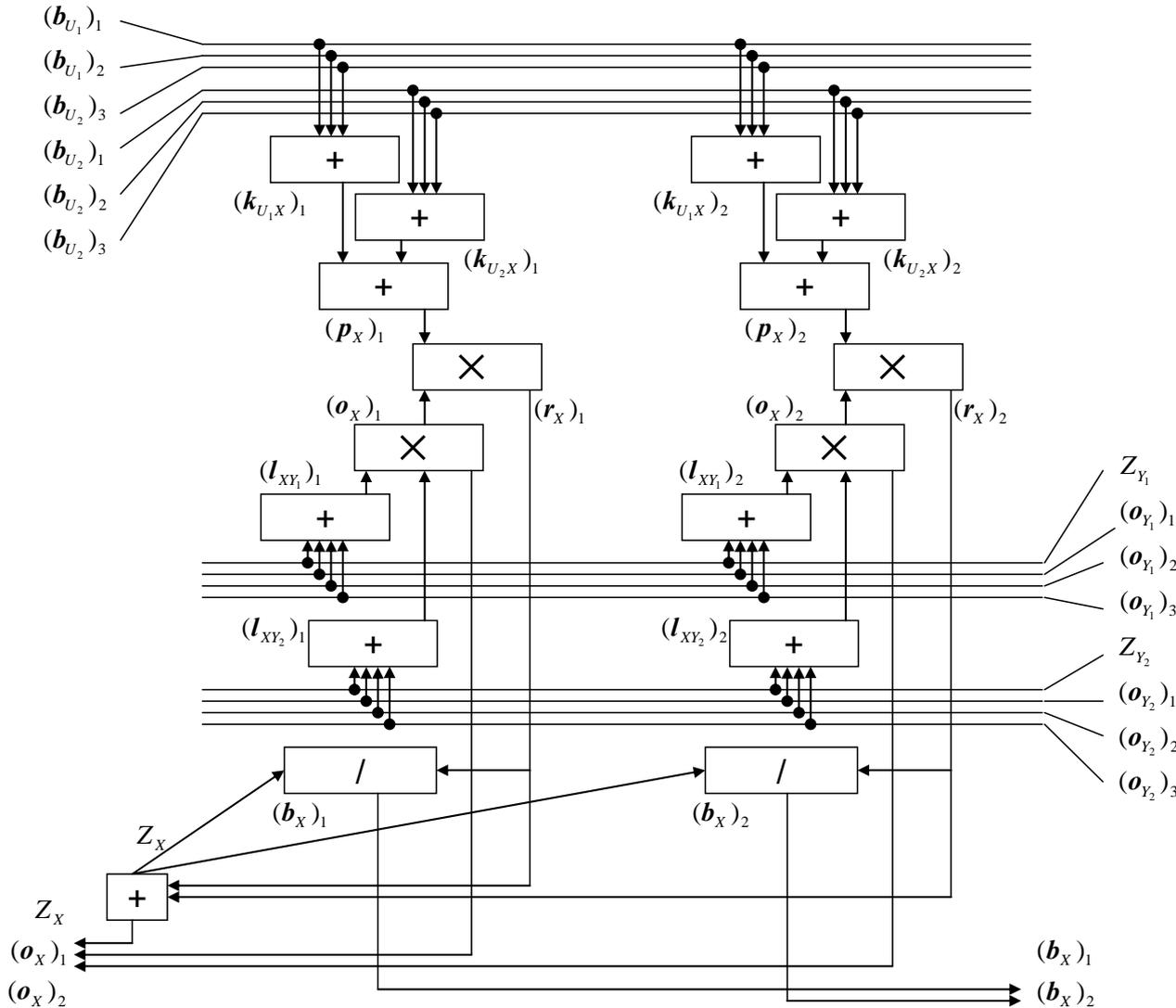


解剖学的構造



モデル

各変数の値を計算する回路



I

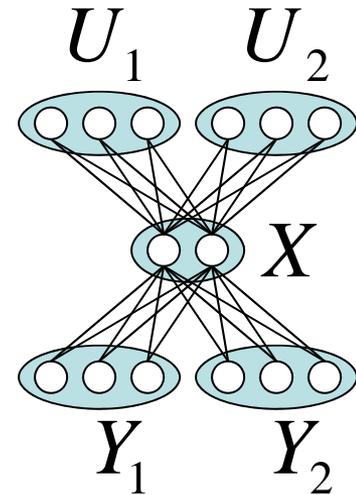
II

III

IV

V

VI

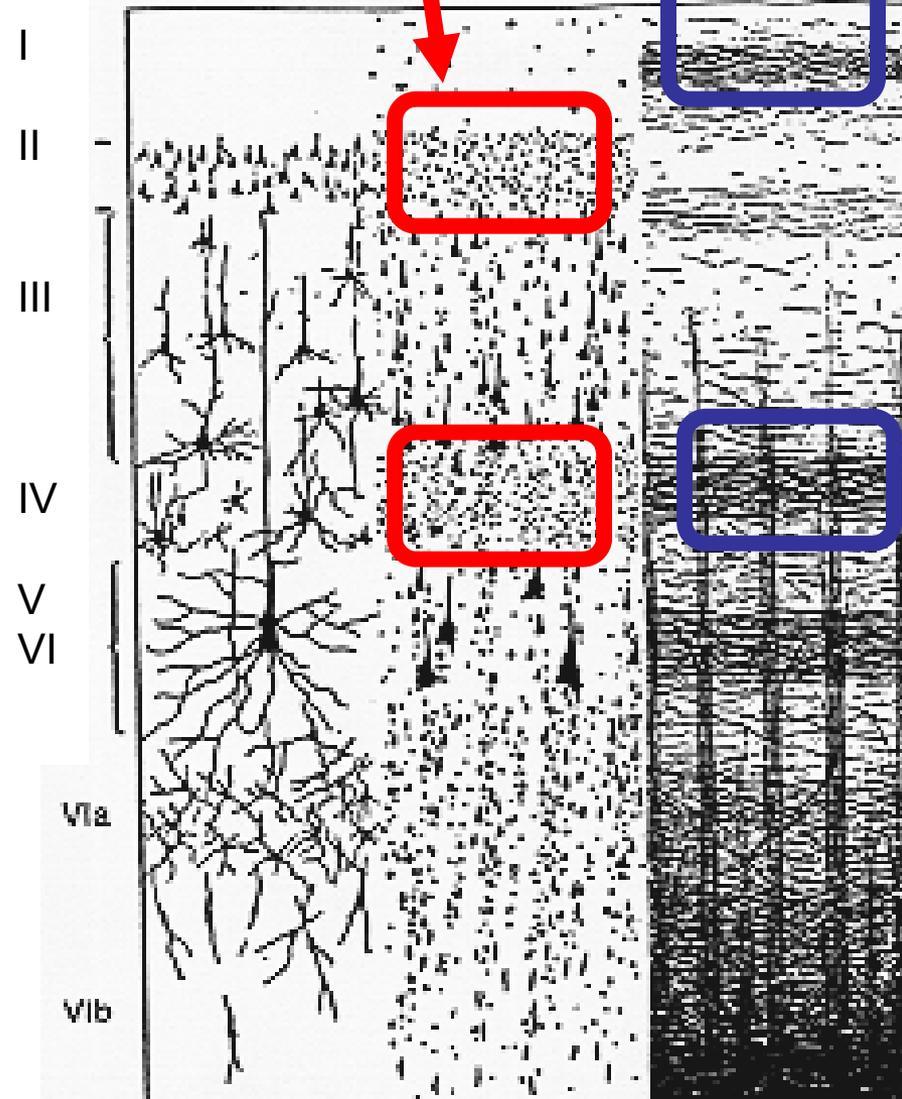
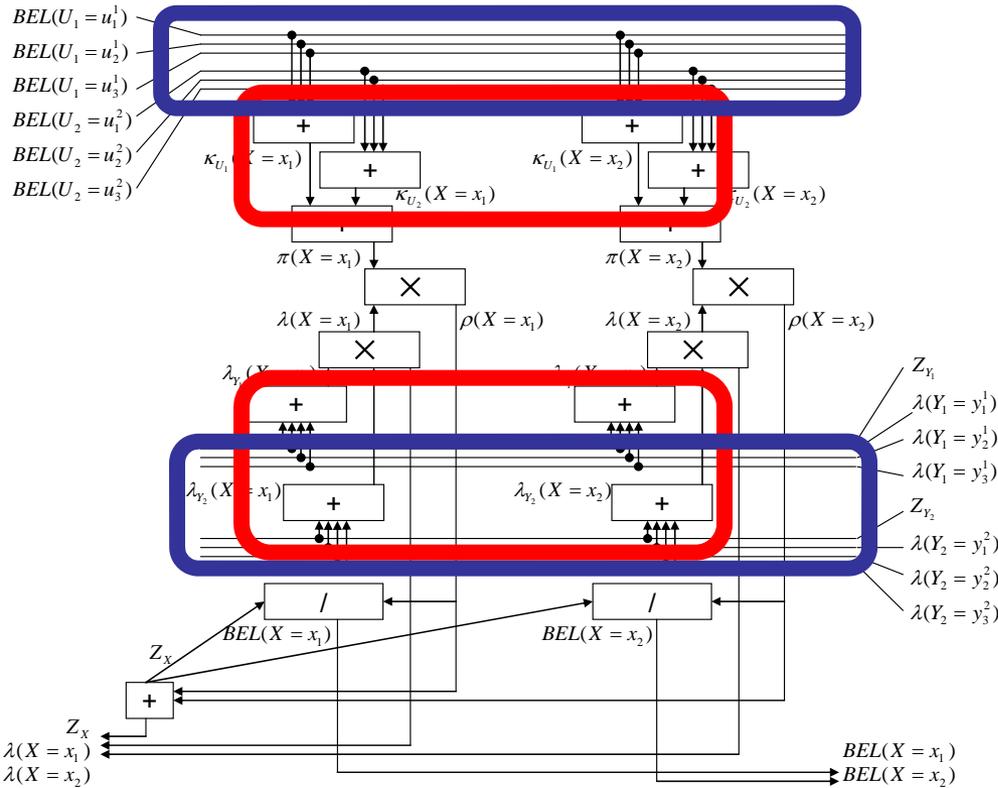


左図は、上記
BESOMネットワー
クのノードXのユ
ニットの値を計算
する回路

大脳皮質の構造との一致

1層、4層の
水平線維

2層、4層の細かい細胞

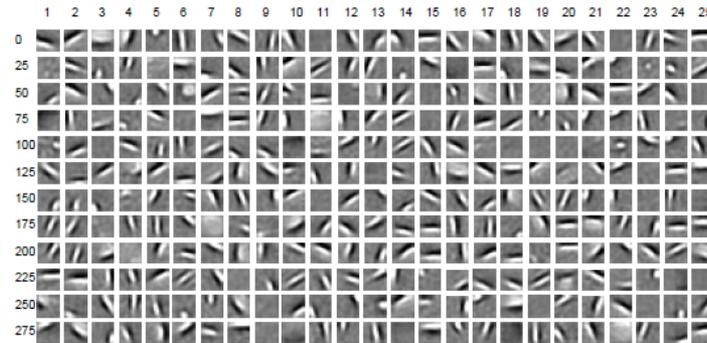
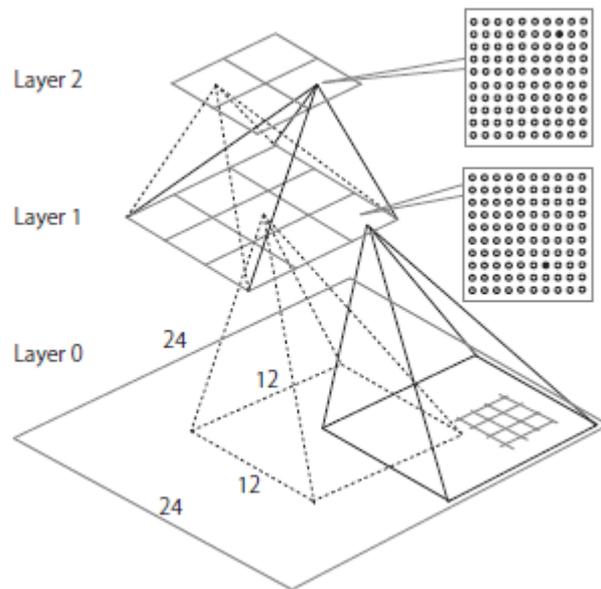


K. Brodmann, Vergleichende Lokalisation der Grosshirnrinde. in: ihren Prinzipien dargestellt auf Grund des Zellenbaues., J.A. Barth, Leipzig, 1909.

This figure is taken from the following Web page.
<http://web.sc.itc.keio.ac.jp/anatomy/brodal/chapter12.html>

ベイジアンネットの学習 [Hosoya 2012]

- 一次視覚野、2次視覚野の応答特性の再現
- 複数の電気生理学的知見の再現
 - Cross-Orientation Suppression, Filling-In.



Hosoya, Haruo ,
Multinomial Bayesian Learning for Modeling Classical and Nonclassical Receptive Field
Properties
NEURAL COMPUTATION, 24 (8): 2119-2150 AUG 2012

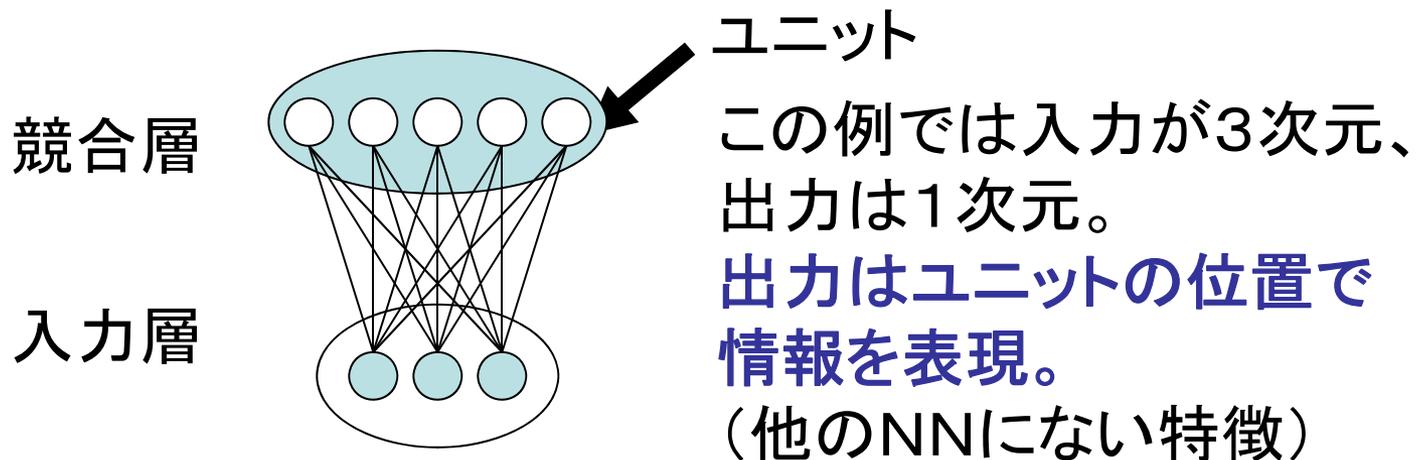
ここまでのまとめ

- 大脳皮質がある種のベイジアンネットであることはまず間違いない。
- 脳の情報処理原理を理解するために**非常に重要な手掛かり**。
- 過去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る。

SOM(自己組織化マップ)と ベイジアンネット

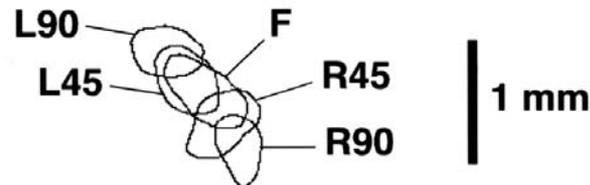
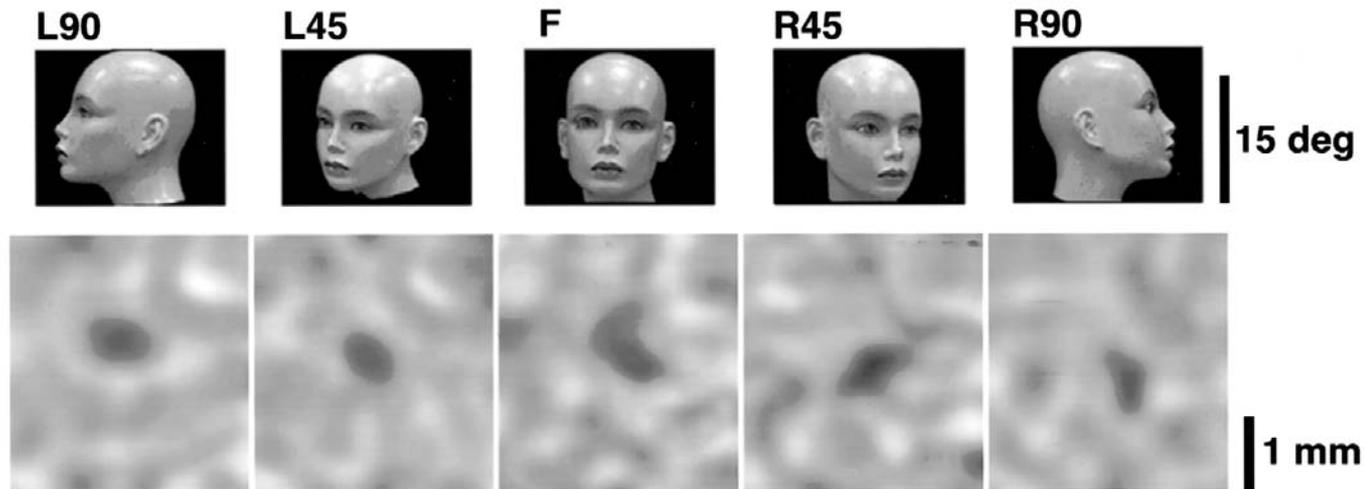
SOM (自己組織化マップ) [Kohonen 1995]

- 脳の一次視覚野のモデルを工学的に扱いやすくしたものの。
- 脳の「概念獲得」と似た働きをする、教師なし学習アルゴリズムの1つ。
- 高次元の入力ベクトルを低次元に圧縮する。



IT野の顔の向きに応答するコラム

[Wang et al. 1996]

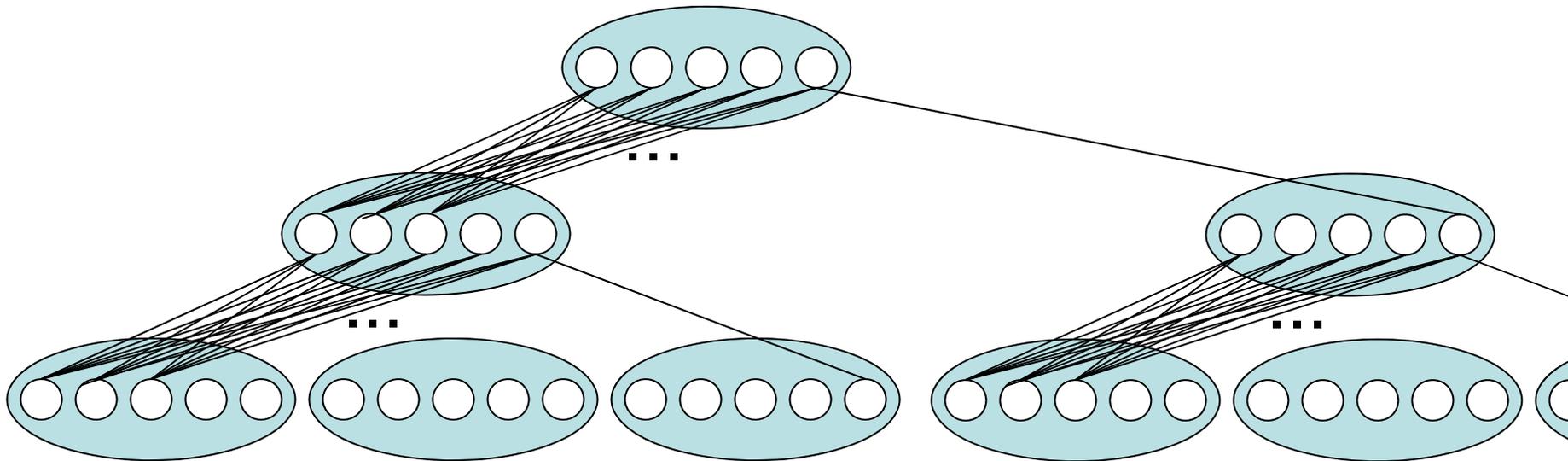


Wang G, Tanaka K and Tanifuji M,
Optical imaging of functional organization in the
monkey inferotemporal cortex
SCIENCE 272 (5268): 1665-1668 JUN 14 1996.

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

前提とする2つの知見

1. 大脳皮質の1つのマクロコラムは、1つのSOMのようなものらしい。
2. 上位領野のマクロコラムは、下位領野のマクロコラムの出力を受け取る。



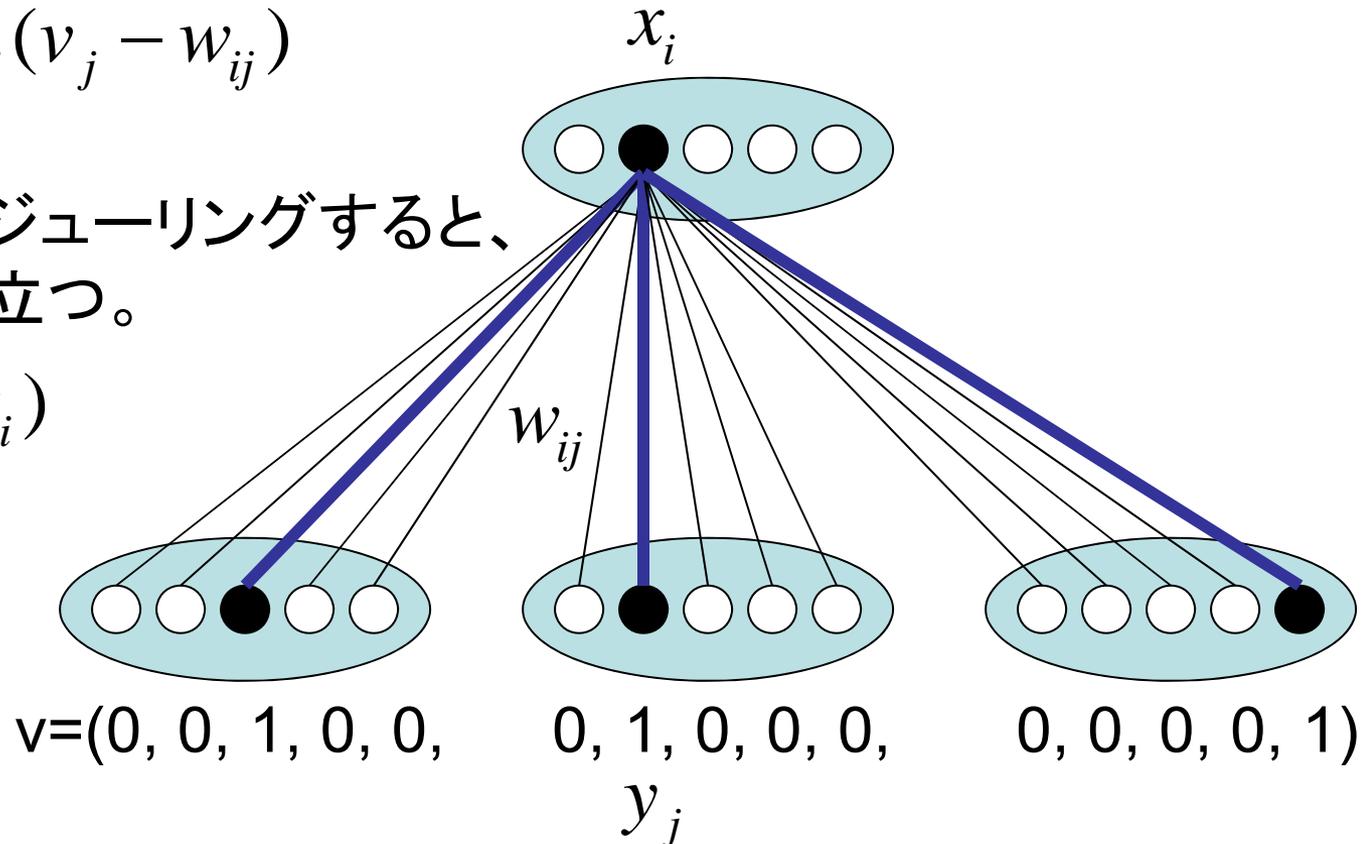
SOMの階層はほぼ必然的にベイジアンネットになる [Ichisugi 2007]

近傍学習を無視すると学習則は

$$w_{ij} \leftarrow w_{ij} + \alpha_i (v_j - w_{ij})$$

α_i を適切にスケジューリングすると、
下記の式が成り立つ。

$$w_{ij} = P(y_j | x_i)$$

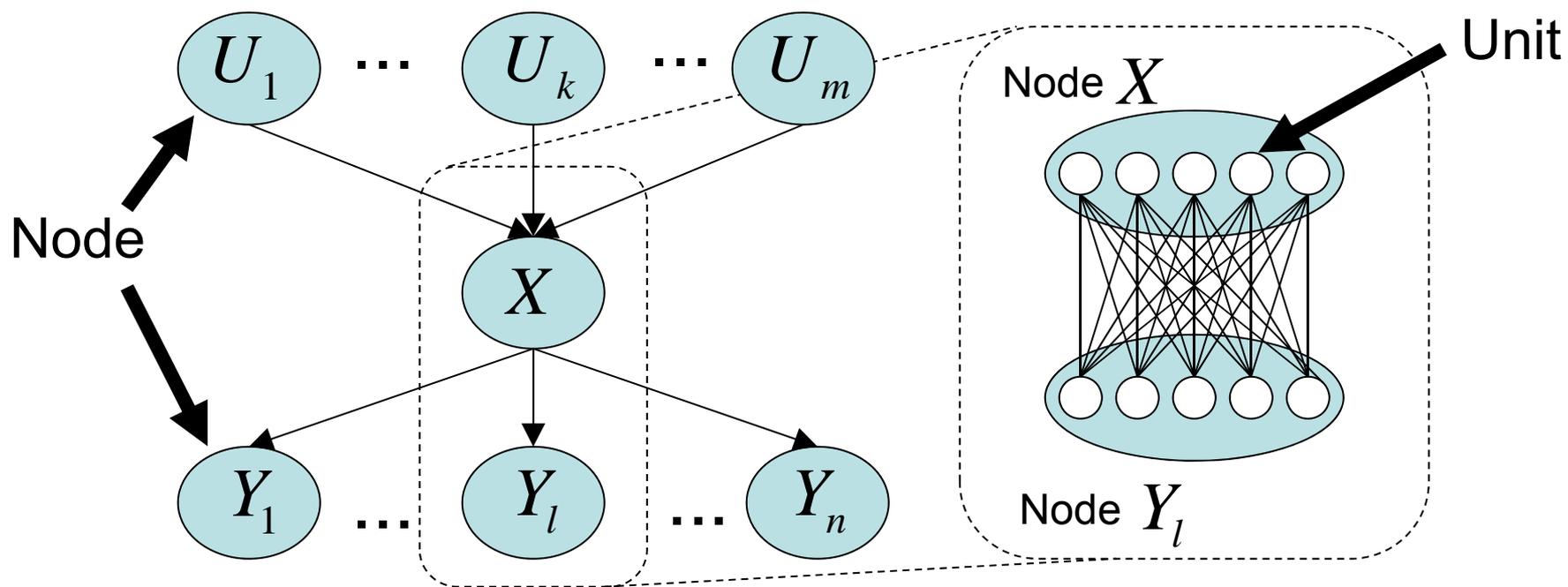


BESOMモデル

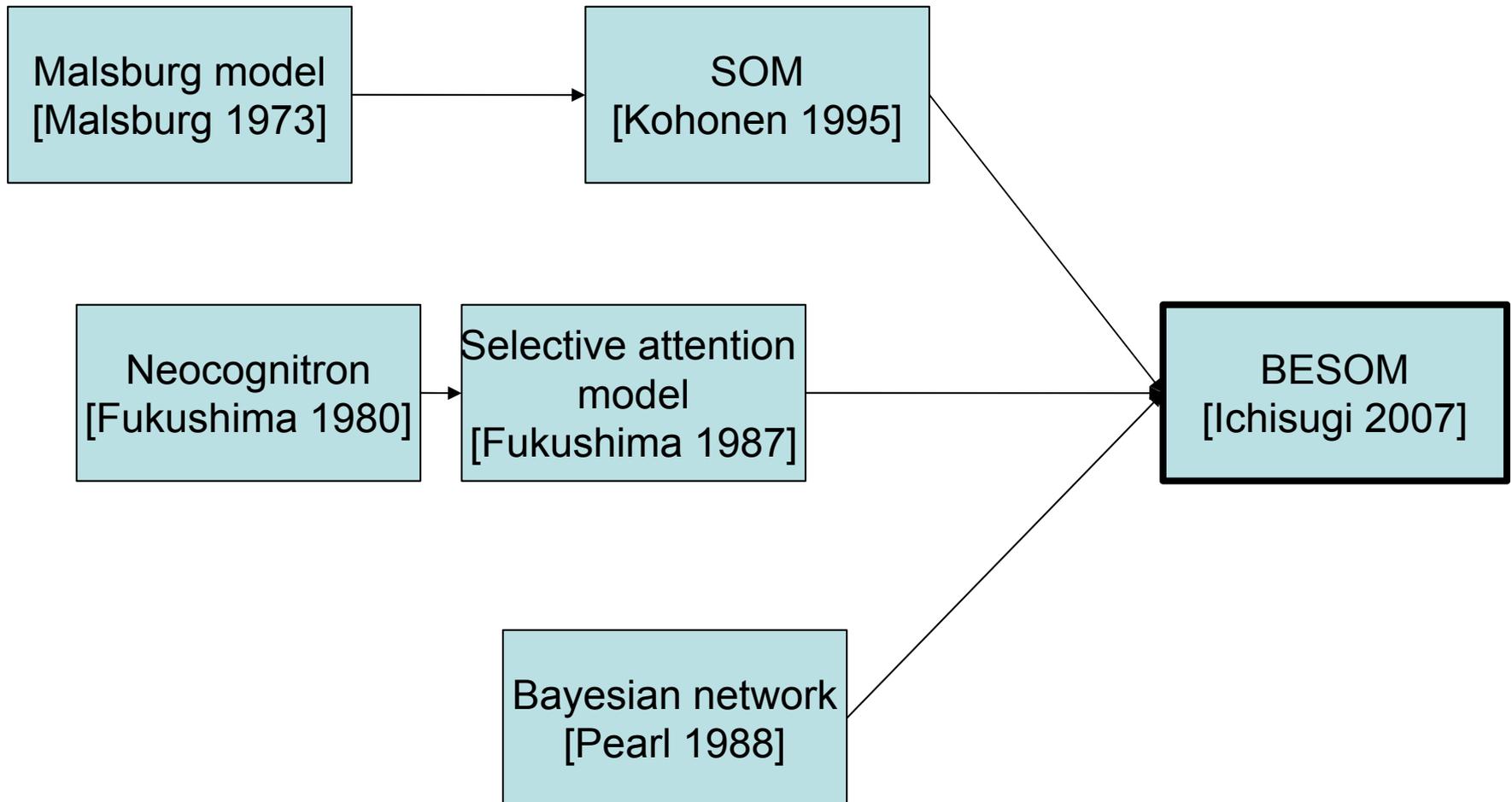
注: BESOMモデルは未完成です。

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

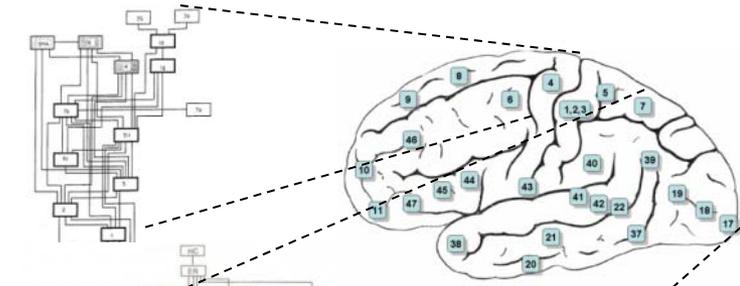
BESOM	自己組織化マップ	ベイジアンネットワーク	大脳皮質
ノード	競合層	ノード (確率変数)	マクロコラム
ユニット	入力ベクトルの要素, 競合層のユニット	確率変数を取りうる値	ミニコラム
親ノード	入力層から見た競合層	親ノード (原因)	上位領野
子ノード	競合層から見た入力層	子ノード (結果)	下位領野
ユニットの出力	入力との類似度	事後確率	コラムの発火
結合の重み	参照ベクトルの要素	条件付確率	シナプスの重み



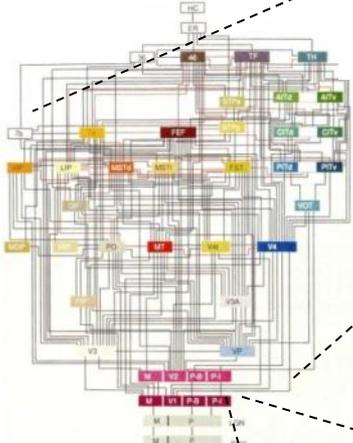
BESOM が統合する 大脳皮質のモデル



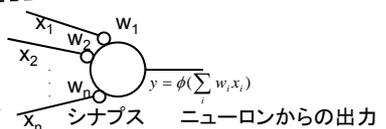
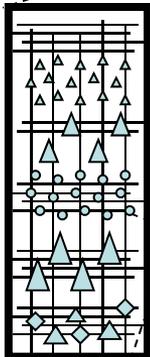
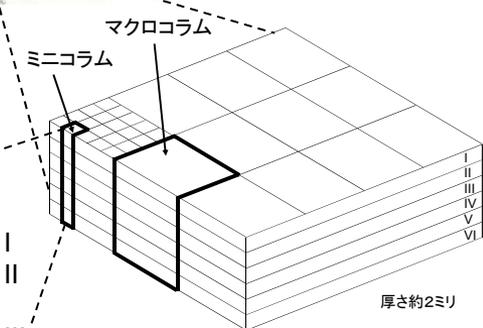
各スケールでの構造と機能の一致



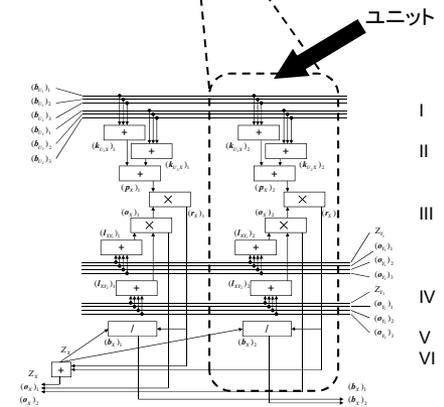
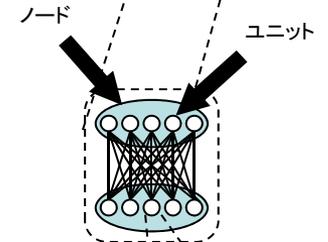
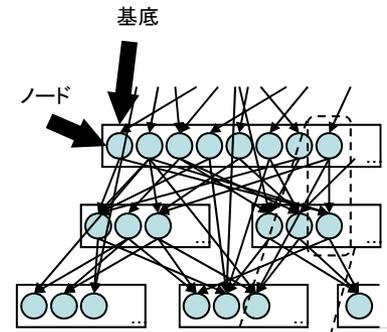
Wikipedia より



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

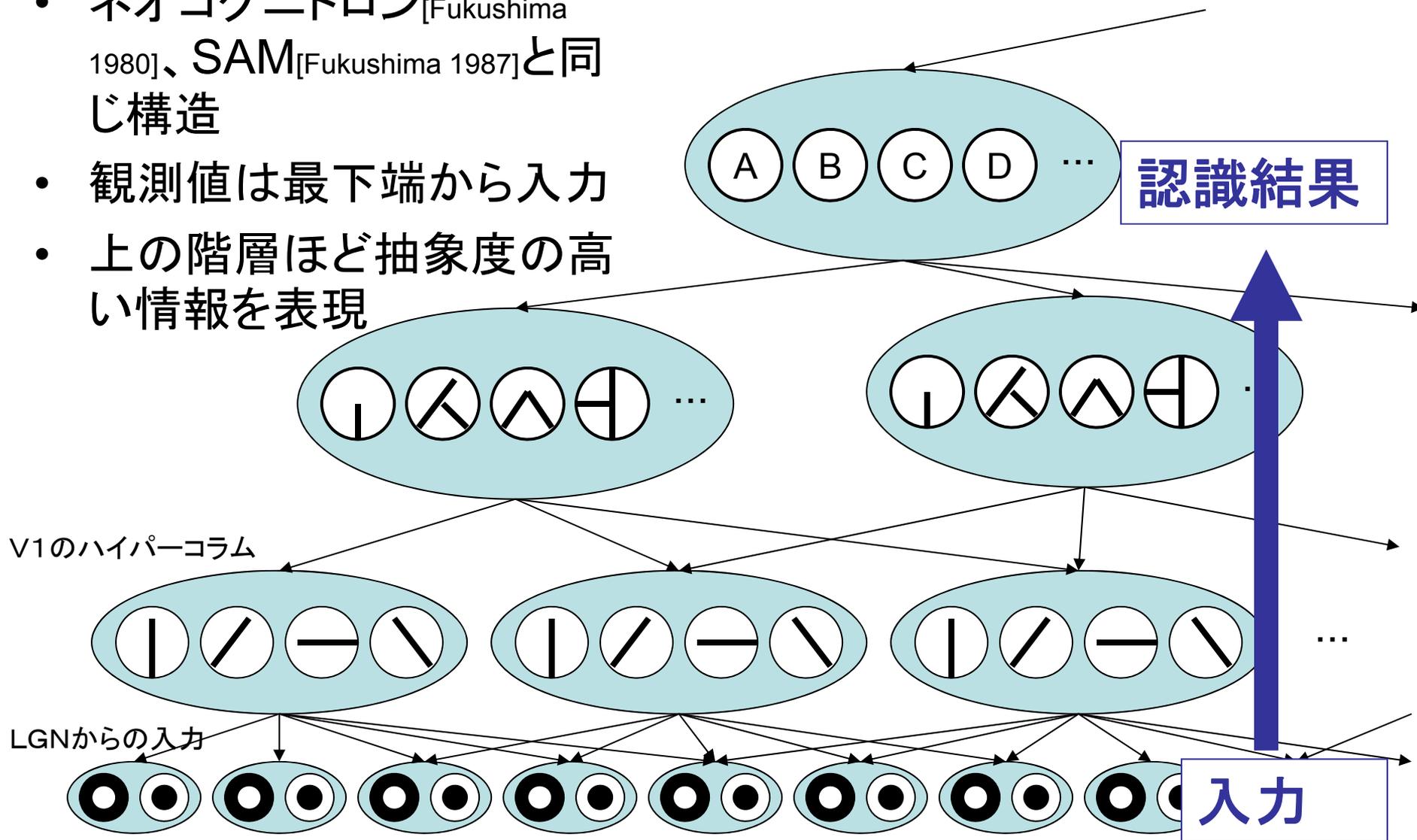


脳の構成要素	BESOM モデルの構成要素
大脳皮質	BESOM ネット
領野階層	基底の階層
領野(の一部)	基底
マクロコラム	ノード
ミニコラム	ユニット
ニューロン	変数
シナプス	ユニット間の結合の重み



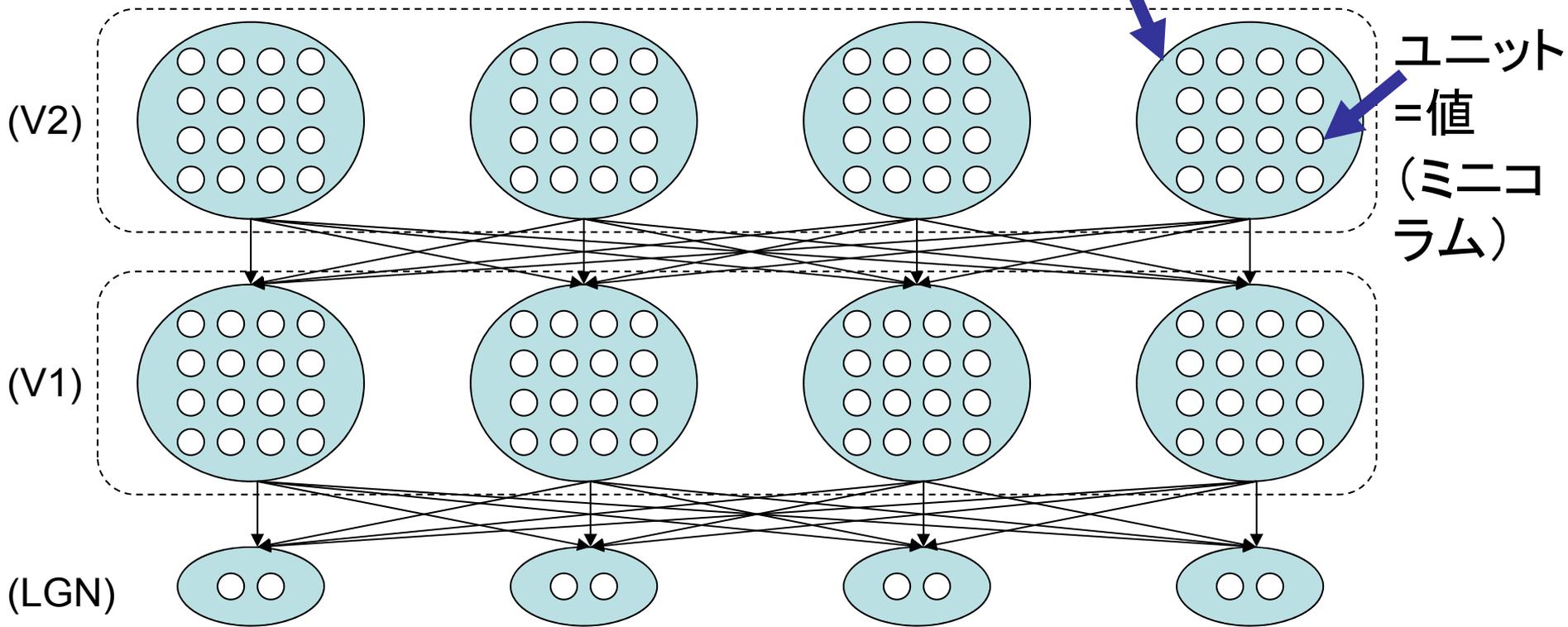
視覚野の模式的説明

- ネオコグニトロン[Fukushima 1980]、SAM[Fukushima 1987]と同じ構造
- 観測値は最下端から入力
- 上の階層ほど抽象度の高い情報を表現



BESOM の構造

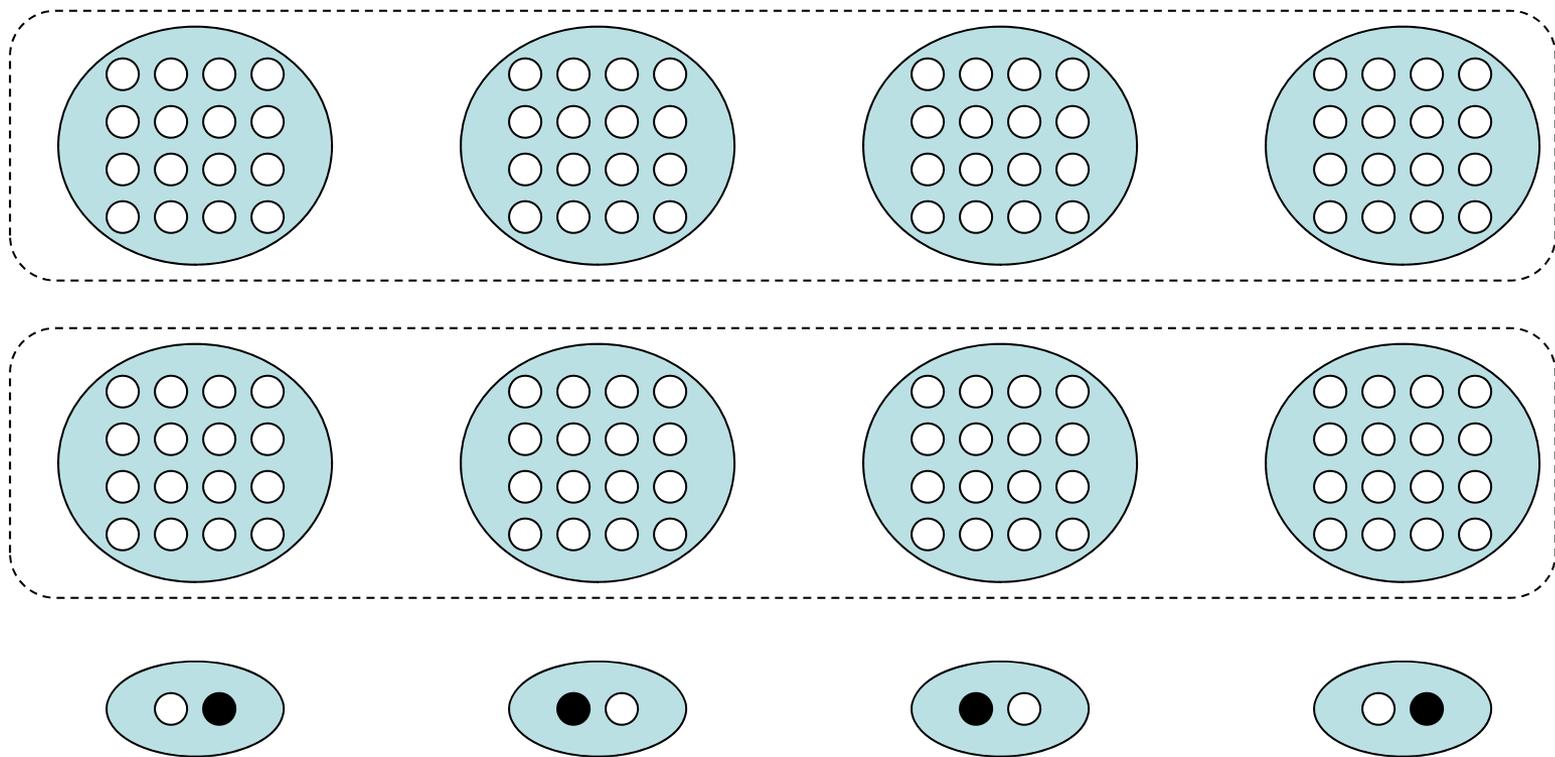
ノード=確率変数=SOM(マクロコラム)



同一階層のユニットどうしは結合がない。
それ以外はすべて結合。

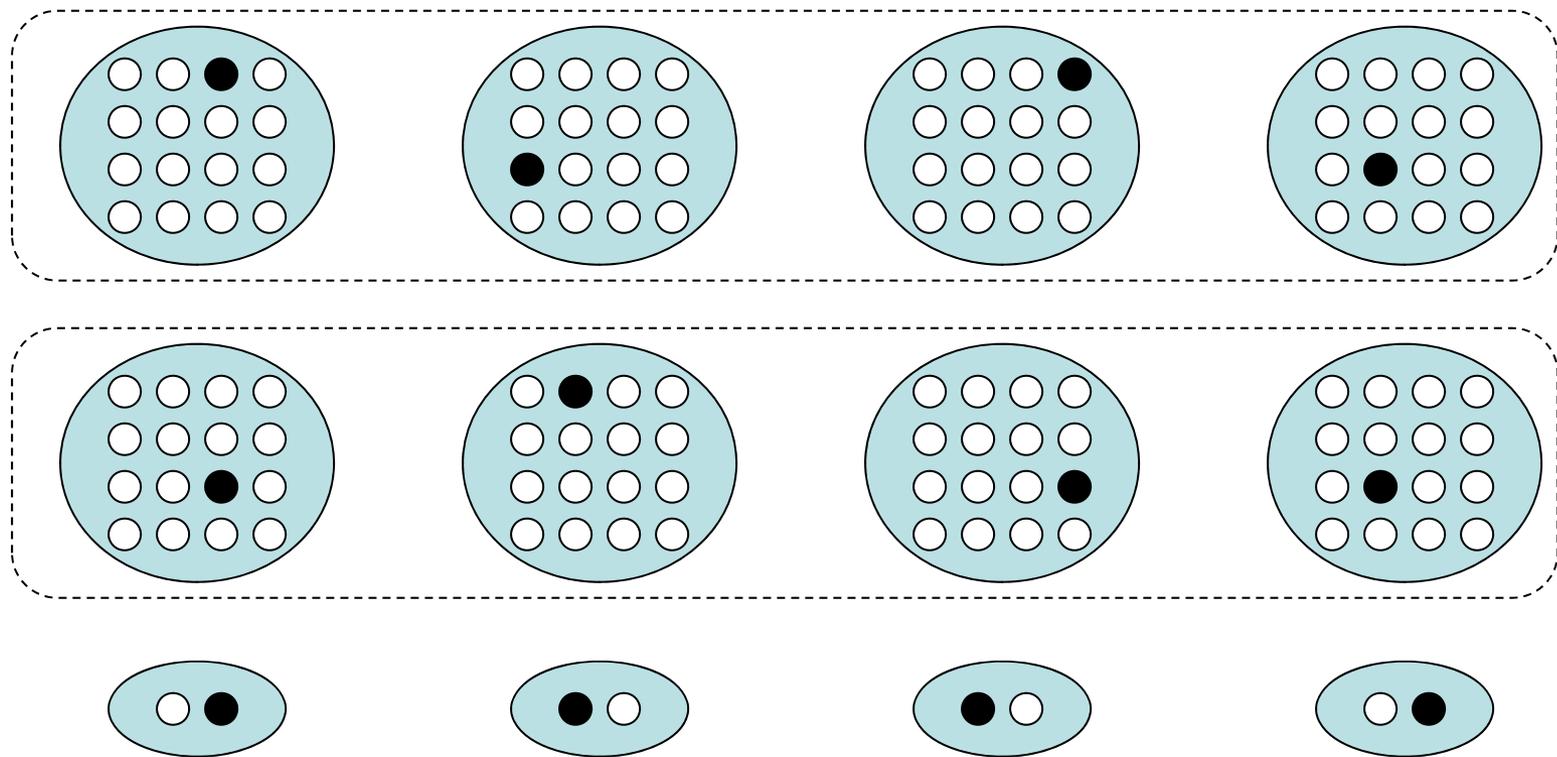
結合の重み
=条件付き確率
(シナプスの重み)

BESOM の動作



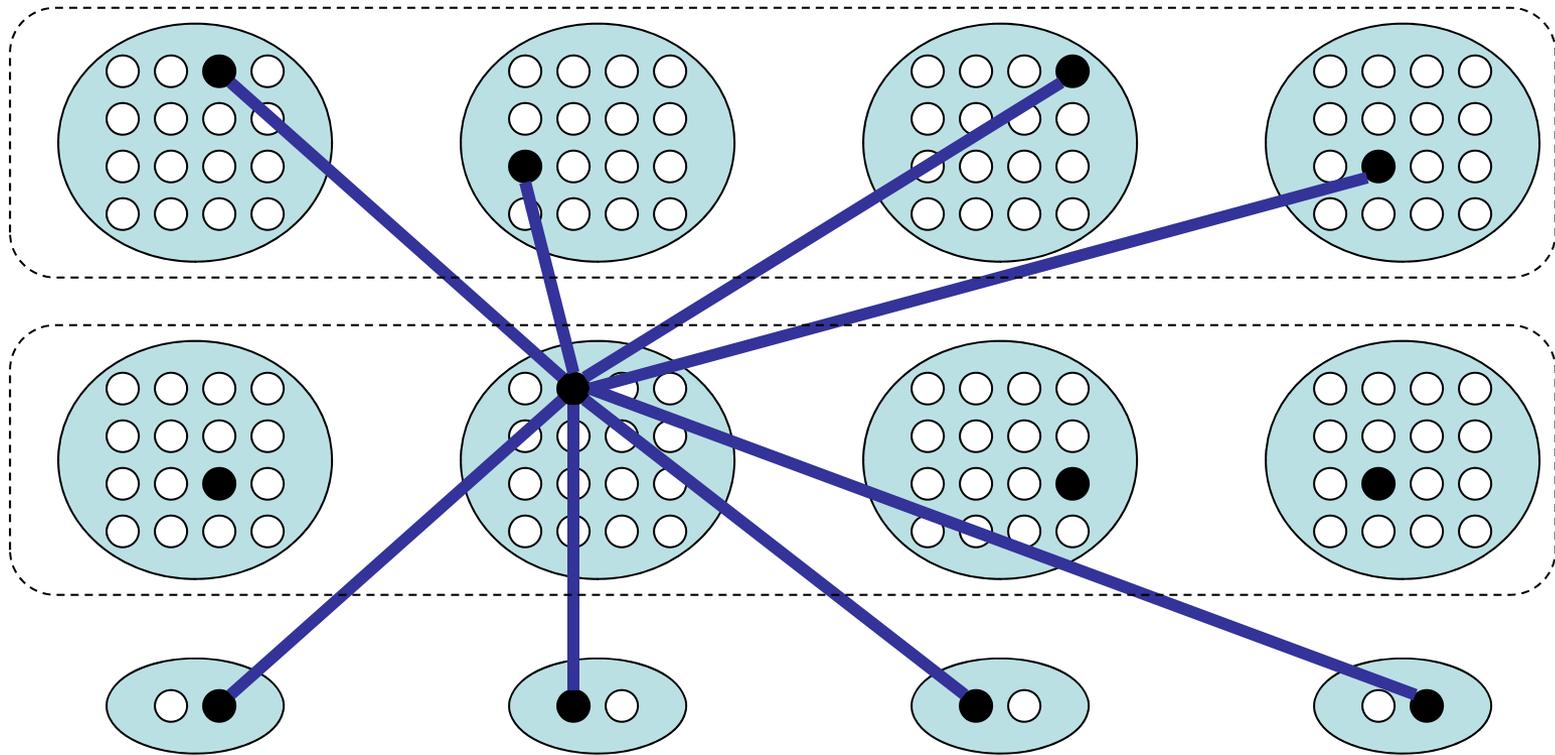
再下端のノードに観測値が与えられる。

BESOM の動作



認識：観測データとの同時確率が最大となる値の組み合わせ(MPE: most probable explanation)を求める。

BESOM の動作



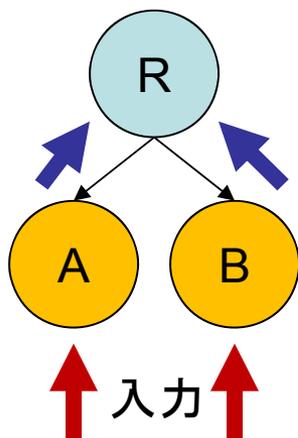
学習：発火したコラムどうしの結合を強め、他の結合は弱める。近傍学習もする。

BESOMの基本動作

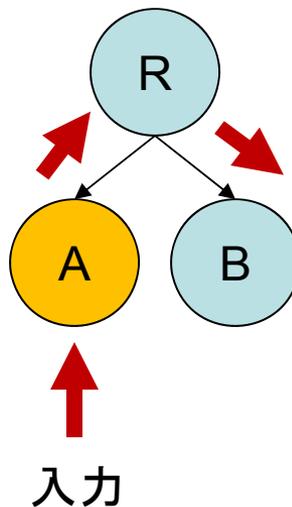
1ノードの場合

BESOMのノード(またはノードの集合)は「データベース」

学習



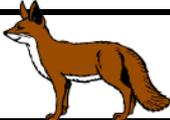
想起

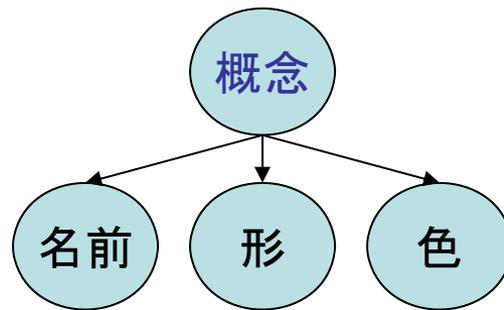


ノードRには、同時入力されるAとBの値の組がデータベースとして記憶される。

Aの値が与えられると、データベースを検索し、対応するBの値が想起される。

データベースの学習の例

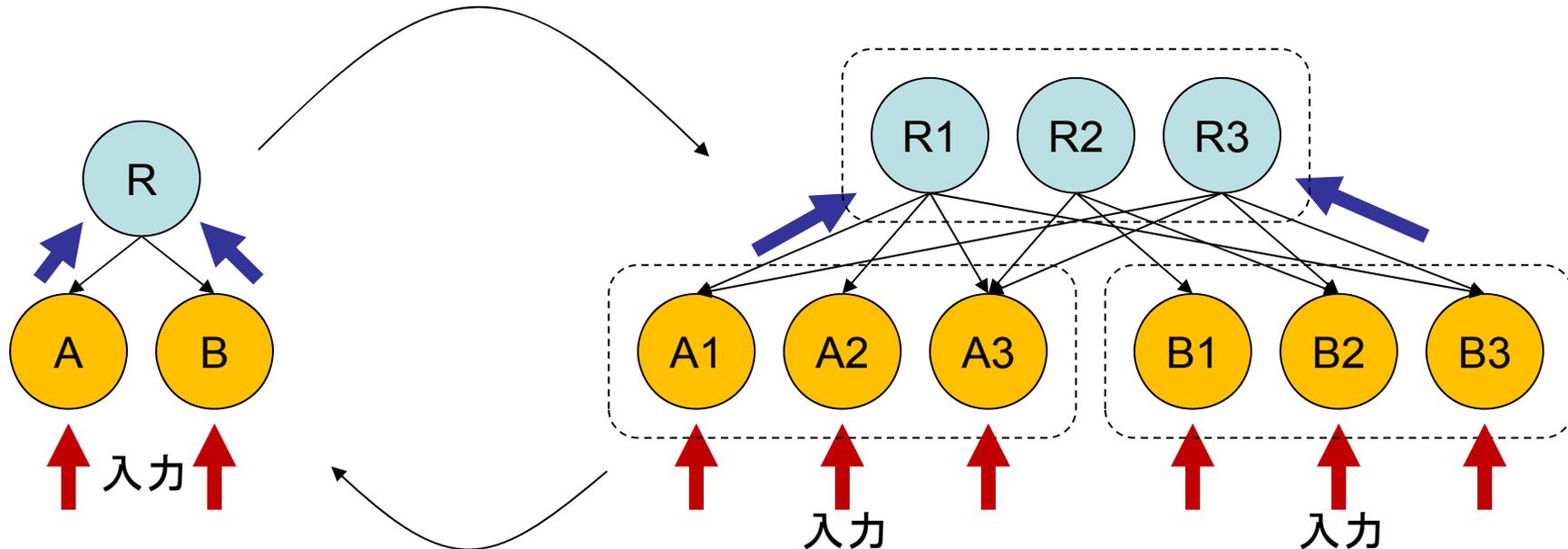
ユニットID	名前	形	色
U01	きつね		茶色
U02	うさぎ		白
U03	かめ		緑
U04	からす		黒
U05	はと		灰色
U06	やぎ		白
...			



注: 競合学習を使って圧縮・抽象化して記憶する。

2層BESOMもデータベースとして 動作

各変数を複数の独立成分に分解



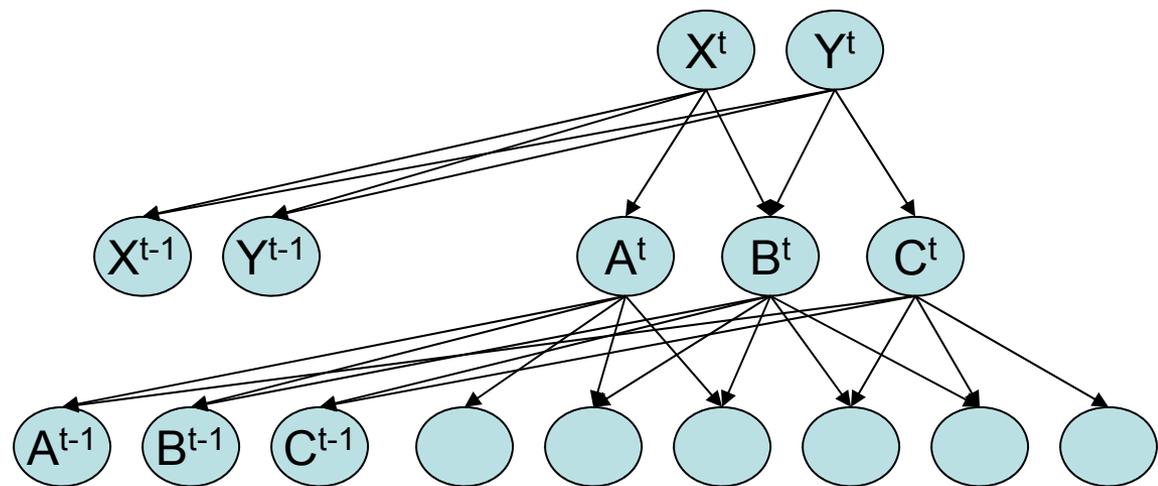
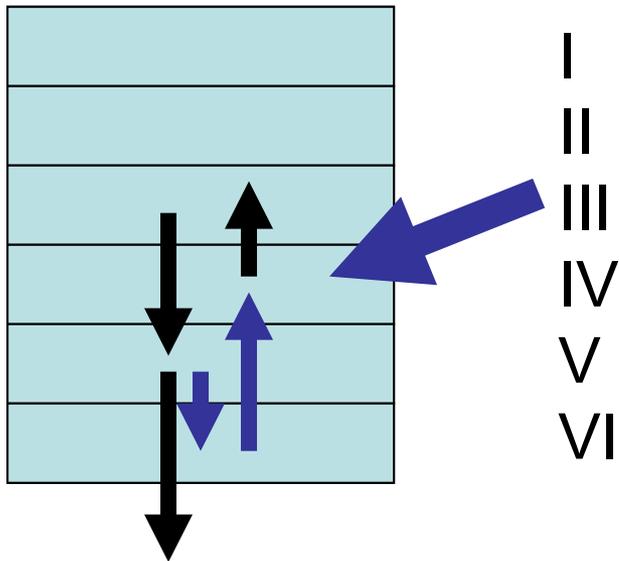
複数の変数を1つに統合

2つのベイジアンネットの機能は等価

大脳皮質のその他の機能 についてのモデルの構想

時系列学習は？

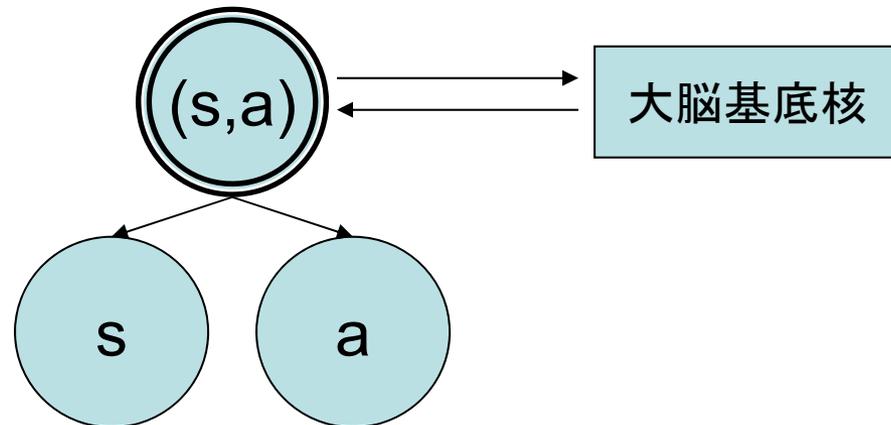
- 6層から4層への入力は、時間遅れの再帰的入力？ → そうならば時系列学習が可能



運動野は？

- ノードが状態行動対を学習。行動価値は大脳基底核(線条体)に投射するシナプスが学習 [Ichisugi 2012]

Yuuji Ichisugi, A Computational Model of Motor Areas Based on Bayesian Networks and Most Probable Explanations, In Proc. of The International Conference on Artificial Neural Networks (ICANN 2012), Part I, LNCS 7552, pp.726--733, 2012.



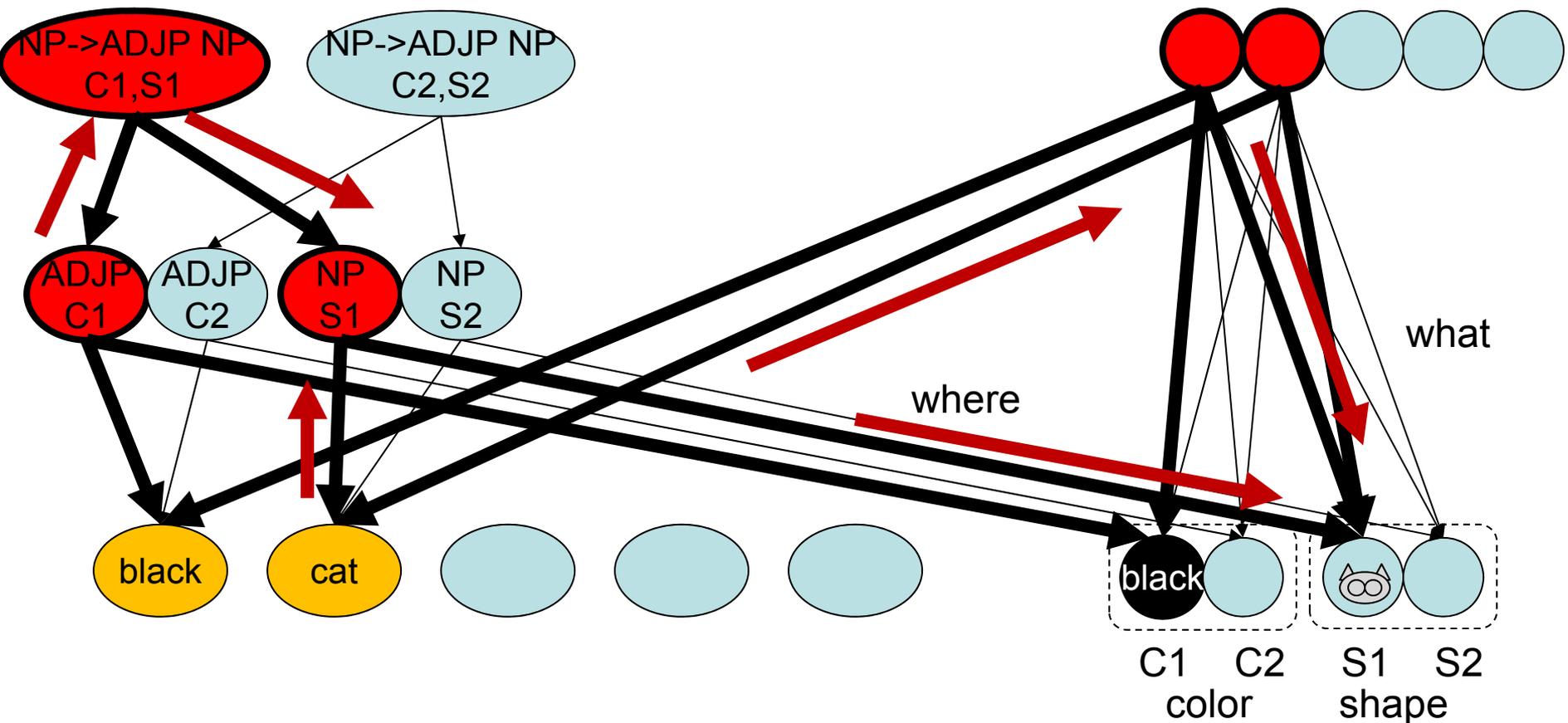
- 基底核－皮質ループの解剖学的構造と一致。
 - [Doya 2000] のモデルの拡張になっている:
K. Doya, Complementary roles of basal ganglia and cerebellum in learning and motor control, Current Opinion in Neurobiology 10 (6): 732-739 Dec 2000.

言語野 = チャートパーサ？

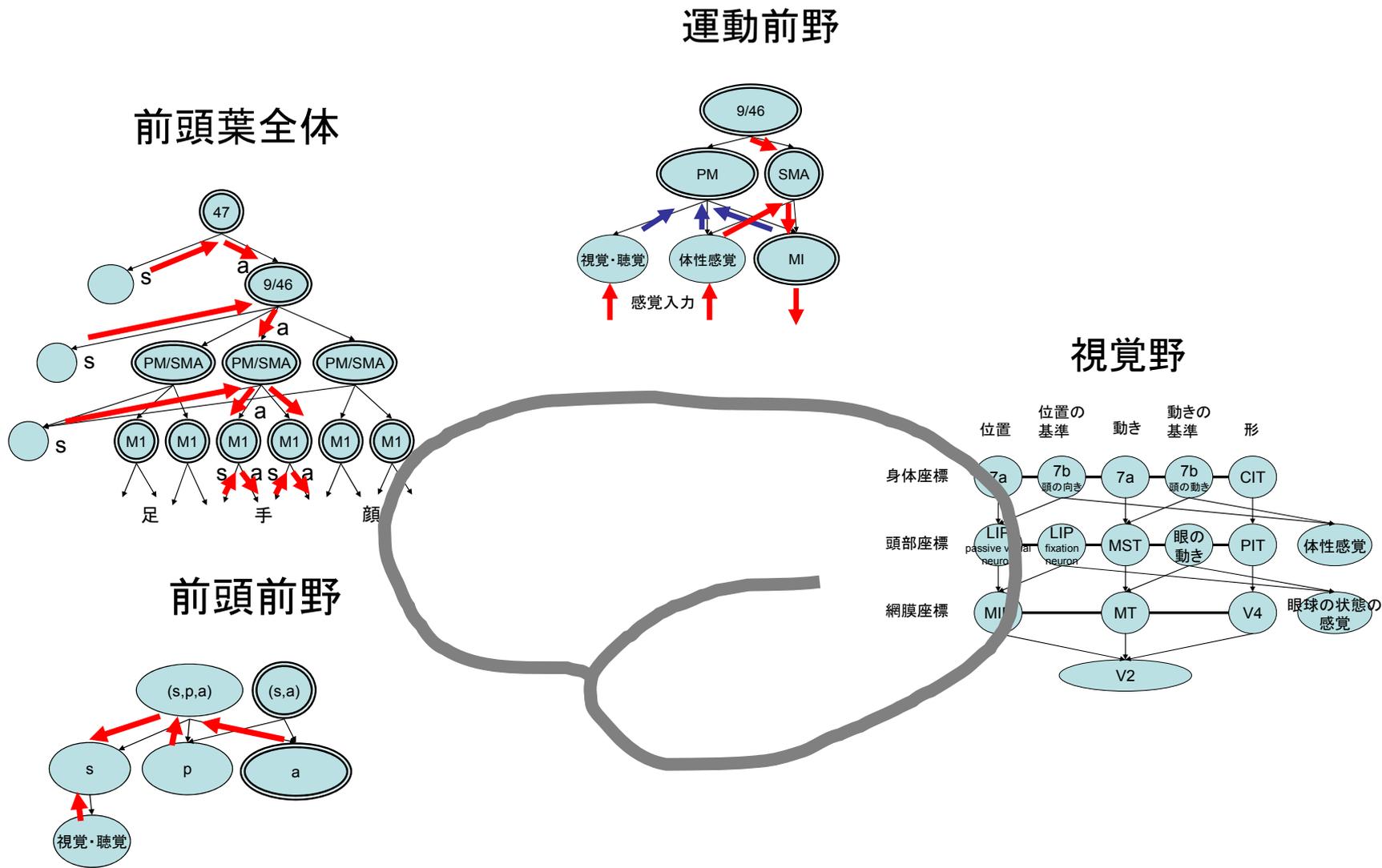
研究構想説明資料:「BESOMを使った言語野モデルの構想」

<http://staff.aist.go.jp/y-ichisugi/besom/201212language.pdf>

(単語,意味) x n

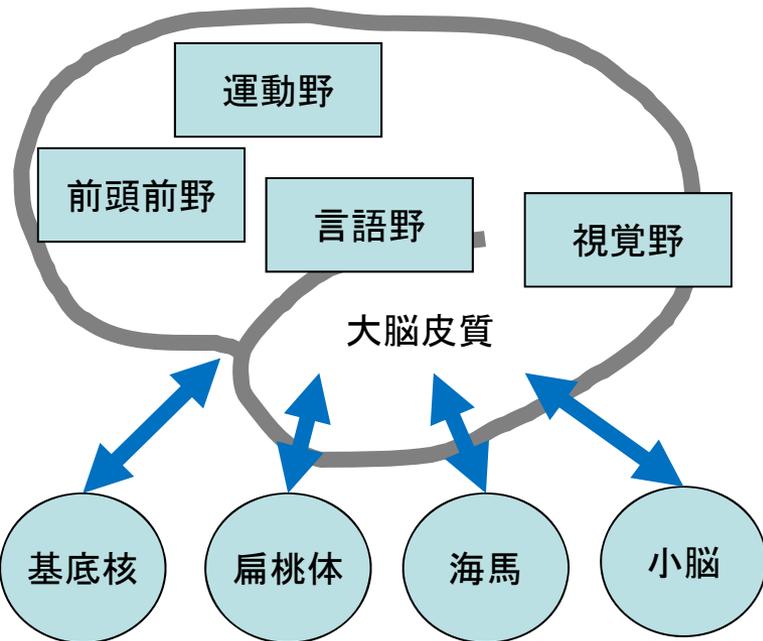


脳の様々な部分が表現可能



大脳皮質以外の組織モデル

脳を構成する主要要素



脳の各組織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: 相互結合型ネットワーク

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: 状態行動対?

前頭前野: 状態遷移機械?

言語野: チャートパーサ?

実用化の可能性

大脳皮質モデルの実用化の動き

- Numenta 社
 - 2004年より継続的に研究開発
 - 旧バージョンでベイジアンネットを使用。
 - ジェフホーキンス「考える脳 考えるコンピューター」2005

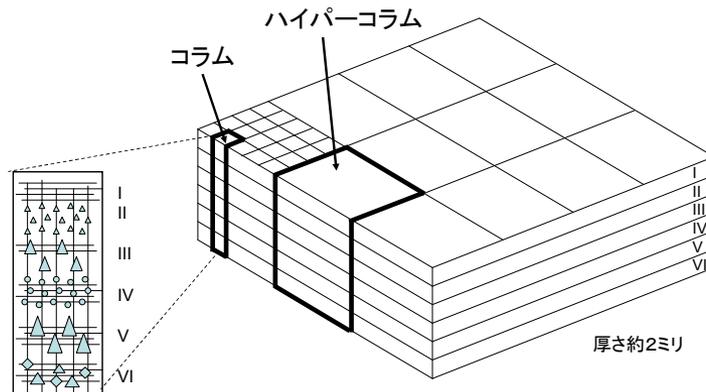


目指すべき目標

- 短期的目標
 - 脳にヒントを得て作られた有用なアルゴリズム
 - 工学・産業への貢献
 - 実際の脳が行っている情報処理アルゴリズム
 - 計算論的神経科学という自然科学の一分野への貢献
- 長期的目標
 - 人間と同じ機能を実現。
 - **人工知能が人工知能研究。**
 - すべての労働を自動化する。

ヒトの大脳皮質の計算速度

- ニューロンの動作を発火頻度モデルで素直に計算する場合：
- ニューロン100億、シナプス1000個、演算数毎秒100回とすると、**1PFLOPS**。
- 2008時点でのトップクラスのスパコン程度。
- メモリ量(シナプス数)は10TB程度。



人工脳の特徴

- 知能：調整可能、ゼロ～賢い人間程度
- 思考速度：調整可能、ゼロ～無限大、コストとトレードオフ
- 記憶力：調整可能、ゼロ～人間程度
- 感情、欲求：ユーザに役立つようにメーカーがデザイン
- 自己保存欲求：調整可能
- 常識：あり
- 自由意志：あり
- 自己認識：あり
- 創造性・ひらめき：人間程度
- 1個体の教育コスト：人間の教育コストと同程度
- 教育ずみの知識の複製コスト：ゼロ
- ランニングコスト：将来は人間の労働コストより安い
- 知能の寿命：なし
- 自己改変能力：きびしく規制
- 自己複製能力：きびしく規制

労働力としての自然脳と人工脳

	ソフトウェア	自然脳	人工脳
開発コスト・ 教育コスト	高い	安い	安い
複製	可	不可	可
ランニング コスト	安い	高い	安い

参考:「脳とベイジアンネットFAQ」

- ベイジアンネットが「意識」を持ち得るのでしょうか？
- ベイジアンネットが短期記憶を持ち得るのでしょうか？
- ベイジアンネットが意思を持ち得るのでしょうか？
- ベイジアンネットが「ひらめき」を持ち得るのでしょうか？
- ベイジアンネットが感情を持ち得るのでしょうか？

脳 ベイジアンネット

検索

脳とベイジアンネットFAQ

脳の情報処理原理の解明の鍵となる技術がベイジアンネットです。しかし、大半の研究者は大脳皮質とベイジアンネットの鮮やかな対応についてまだ知りません。

脳の情報処理原理に基づいた知能の高いロボットの表現に向け、一人でも多くの神経科学者・計算機科学者に、ベイジアンネットと大脳皮質の関係を知ってもらいたいと思います。

ベイジアンネットそのものについての質問

ベイジアンネットとは何ですか？

ベイジアンネット (ベイジアンネットワーク、Bayesian network) とは、確率論に基づいた推論を効率的に行うための技術です。脳の機能の一つである直観と似た働きをします。

ベイジアンネットは、複数の事象の間の因果関係をネットワーク構造で表現し、同時に因果関係の強さを表す数値も記録したものです。

このように表現された「知識」を用いれば、得られた観測データに基づいて様々な事象の確率を「ベイズ」に基づいて合理的かつ効率的に計算することができます。

音声認識などのパターン認識はベイジアンネットの応用の一つです。与えられた音声信号 (観測データ) が、どの単語に対応する確率が一番高いかを計算することができます。

ベイジアンネットはニューラルネットワークとは違うのですか？

パターン認識に使えるという点では同じですが、全く別のものです。代表的なニューラルネットワークである「パーセプトロン」と比べると以下の点が違います。

1. 入力されるものも出力されるものも「値」ではなく「値の確率分布」です。

言語の専門家の方々に教えていた ただきたいこと

- 言語野に作りこまれている事前分布に関するヒント
 - 「言語現象」にはどういうものがあるのか。
 - 言語の学習・理解にどういうバイアスがあるか。

まとめ

- 大脳皮質は**ベイジアンネット**である。
 - 既存のモデルの多くがベイジアンネットを核にして統合されつつある。
- 脳の機能の再現に必要な計算パワーはすでにある。
- 脳のアルゴリズムの詳細を解明するヒントとなる膨大な神経科学的知見があるが、
- **それを解釈・統合できる人材が圧倒的に不足！**