

大脳皮質とベイジアンネット

第1回 産総研 人工知能セミナー

「脳型人工知能

ー脳を模倣した役に立つ人工知能技術ー」

2015-10-20

産業技術総合研究所

人工知能研究センター

脳型人工知能研究グループ

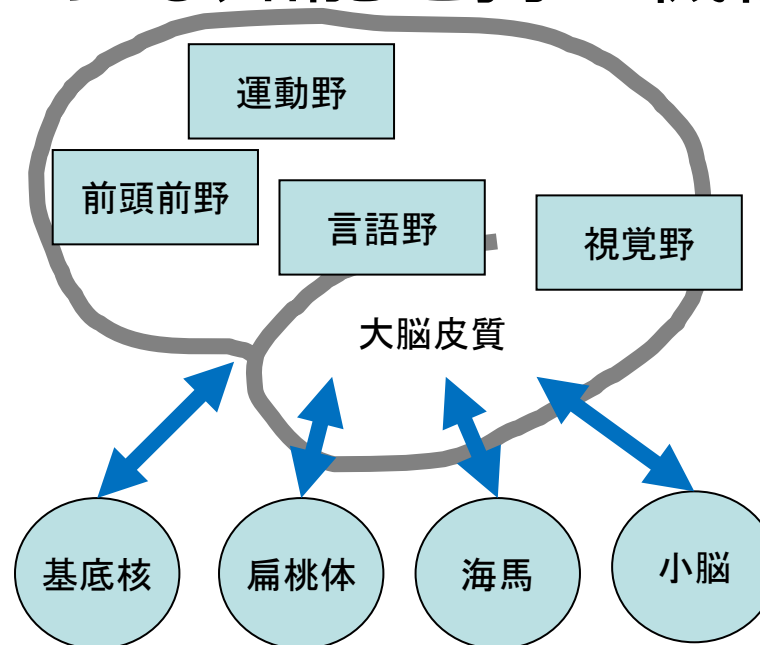
一杉裕志

を5年で目指す

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。

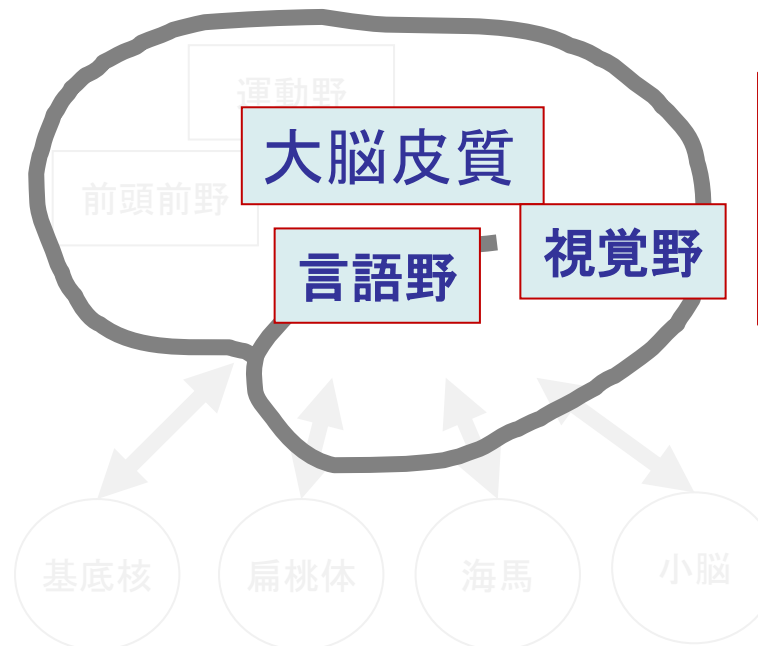
20～30年後？



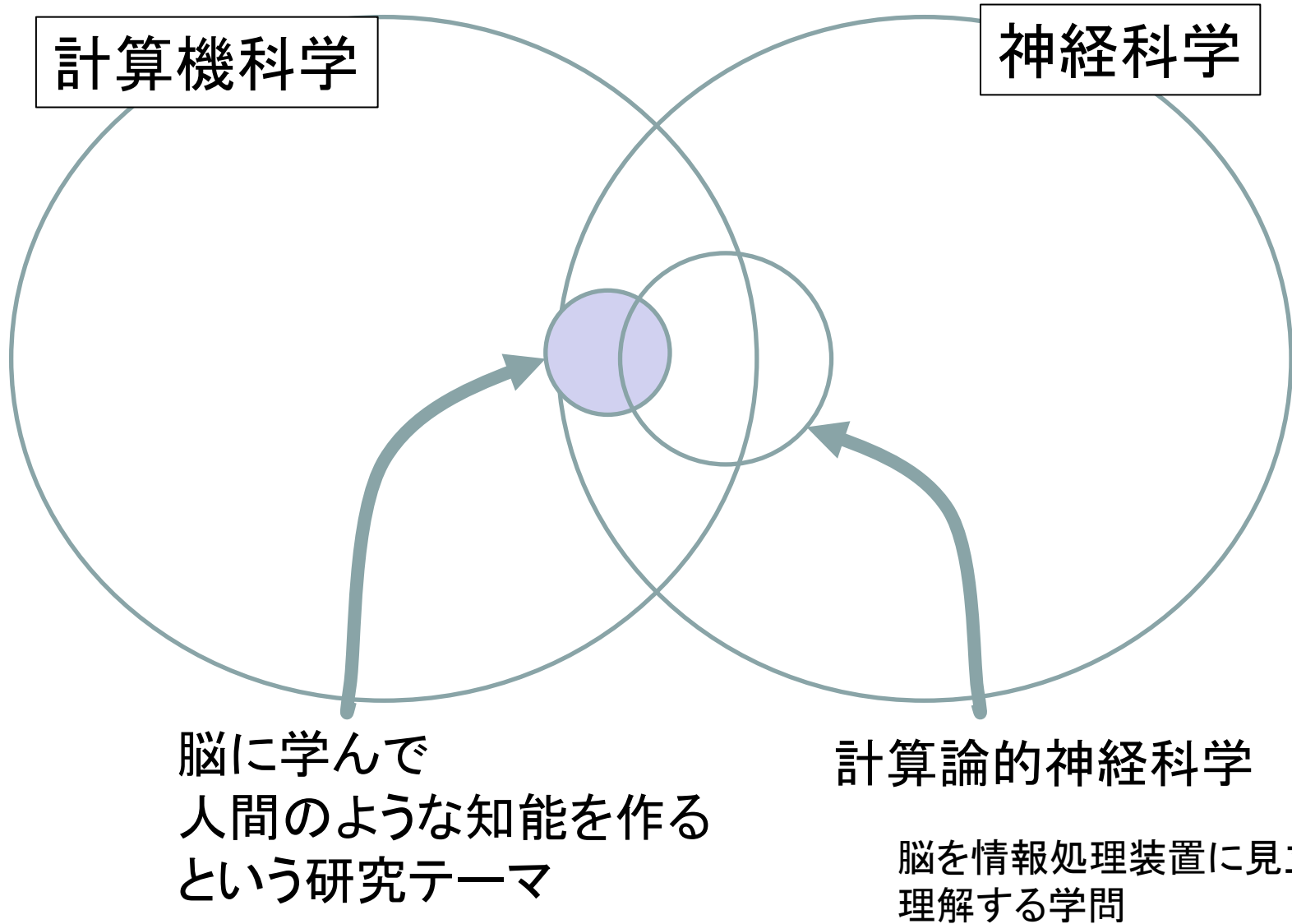
脳のリバーエンジニアリング

私の研究の短期的目標

- 5年で大脳皮質モデルの実用化の目途をつける。



今日説明する内容：
大脳皮質の
計算論的モデル



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。

脳の中のさまざまな研究対象

抽象度
↑

抽象化レベル	構成要素	要素数
様々な高次機能		
脊椎動物の脳のアーキテクチャ	大脳皮質、海馬、大脳基底核、...	5個程度
神経回路		1000億のニューロン
ニューロン	興奮性、抑制性、...	数種類(数百種類?)
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)

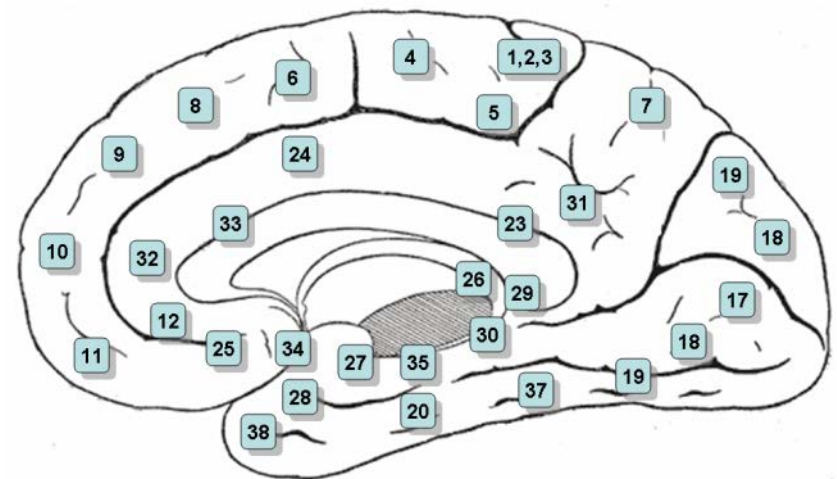
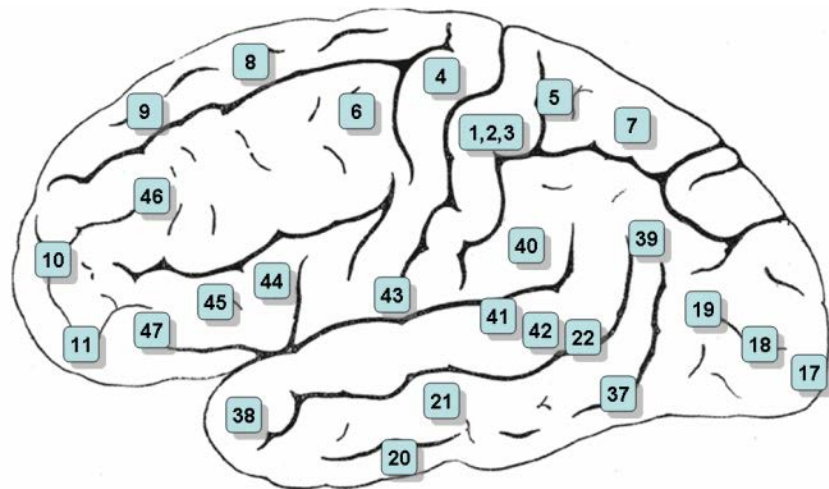
知能の再現
が目的

病気の治療や
薬の開発が
重要な目的

大脳皮質に関する 神経科学的知見

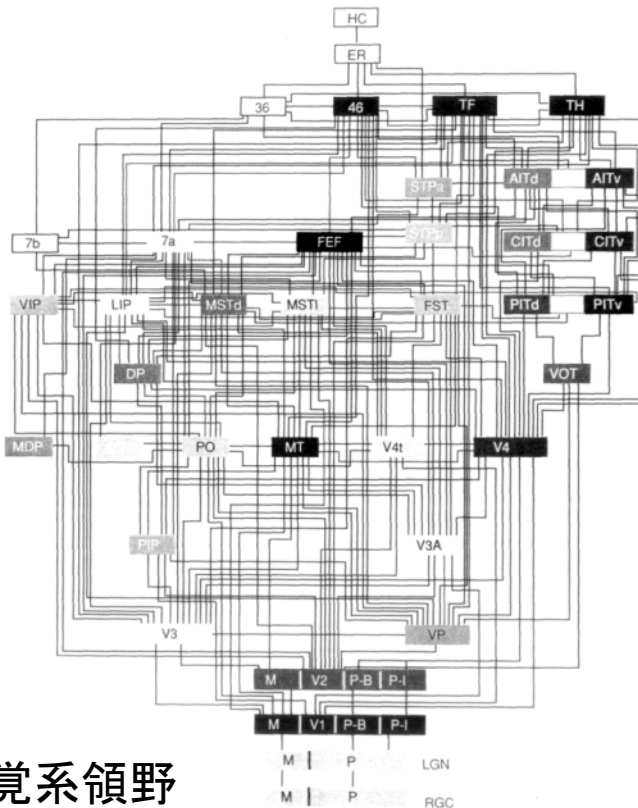
大脳皮質

- 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、・・・



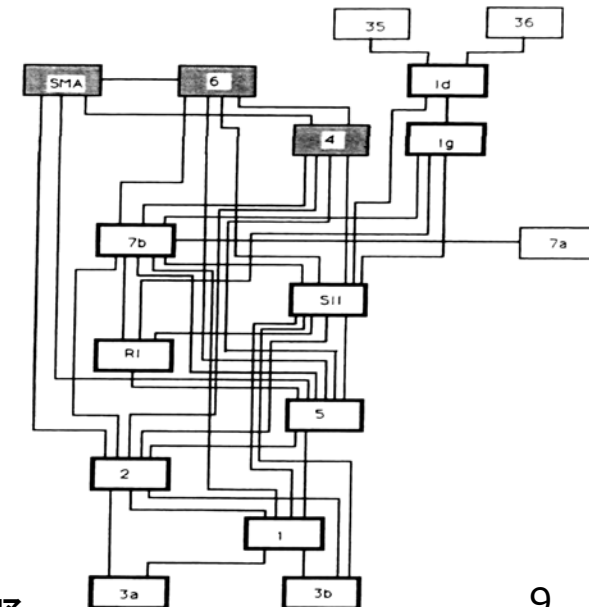
大脳皮質の領野

- 各領野の機能、接続構造はかなり明らかになりつつある。



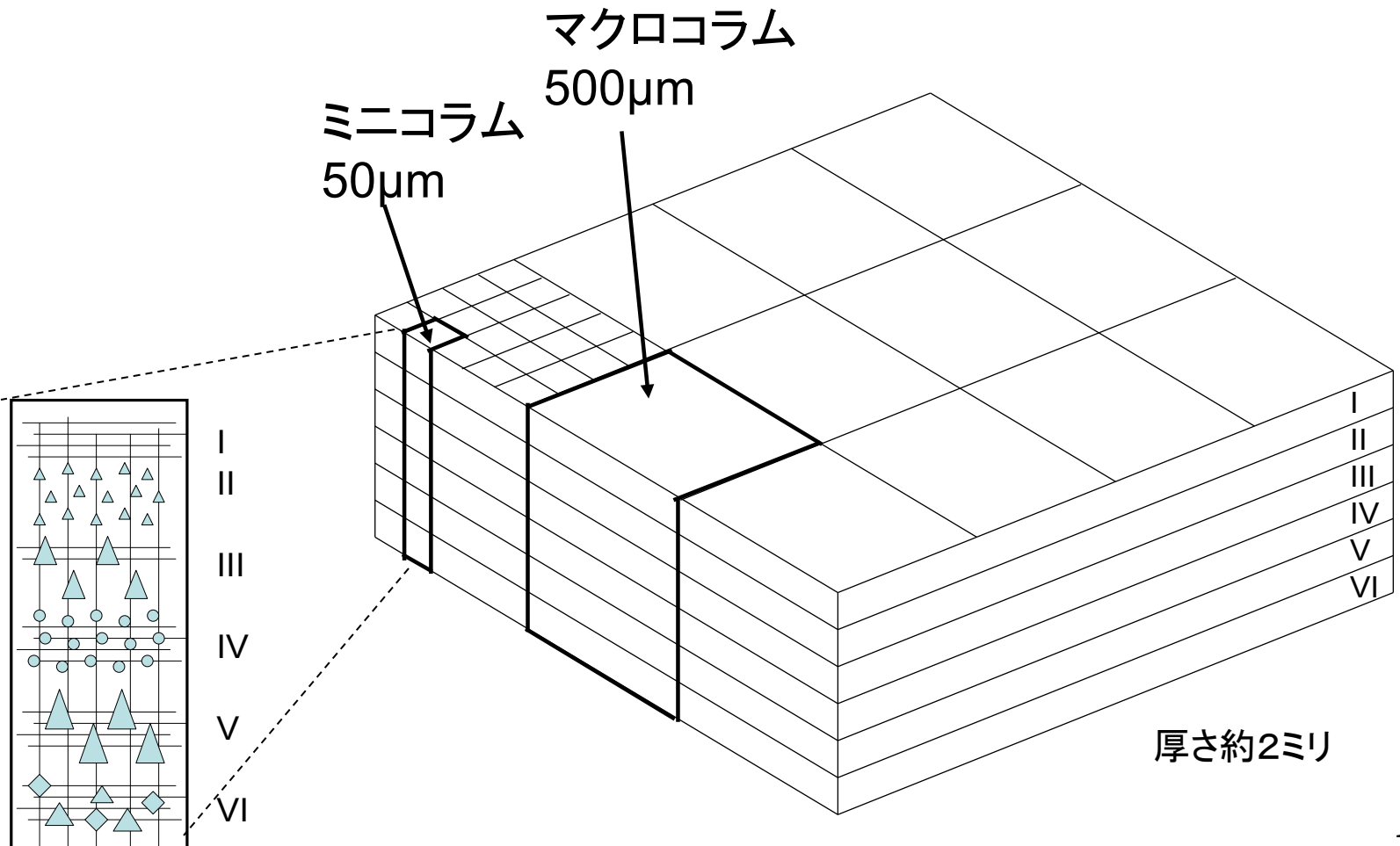
視覚系領野

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



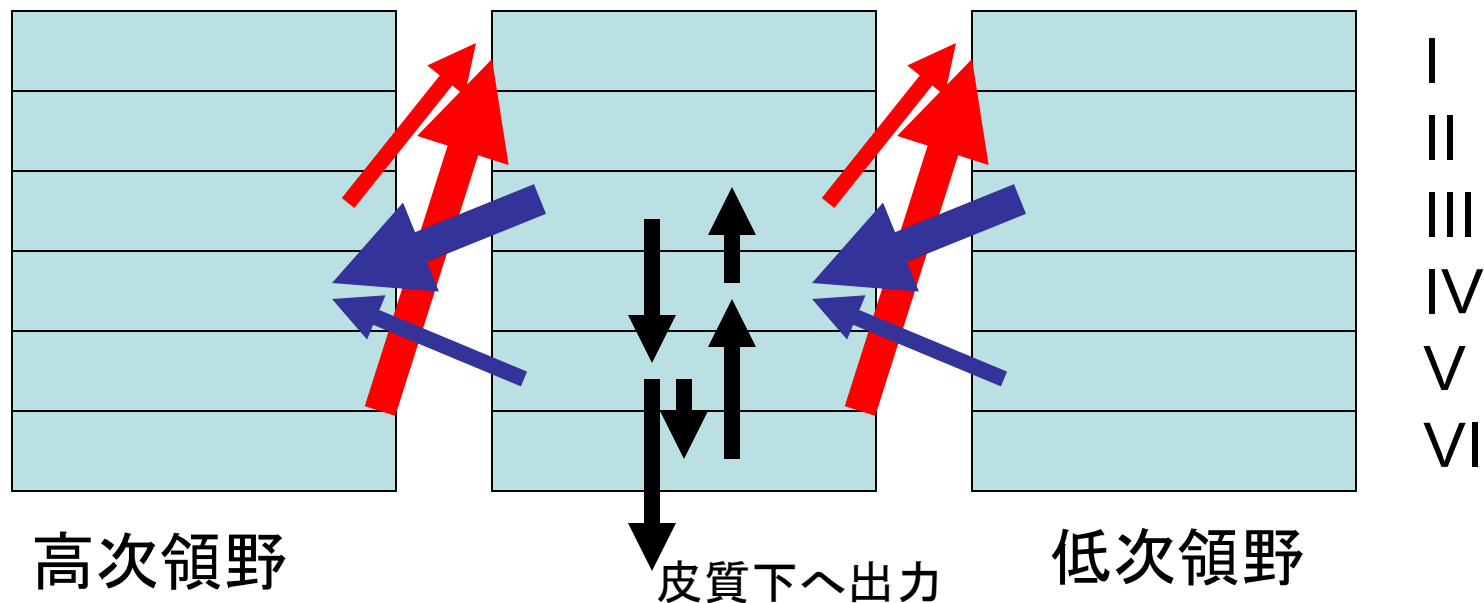
運動野

大脳皮質のコラム構造の模式図



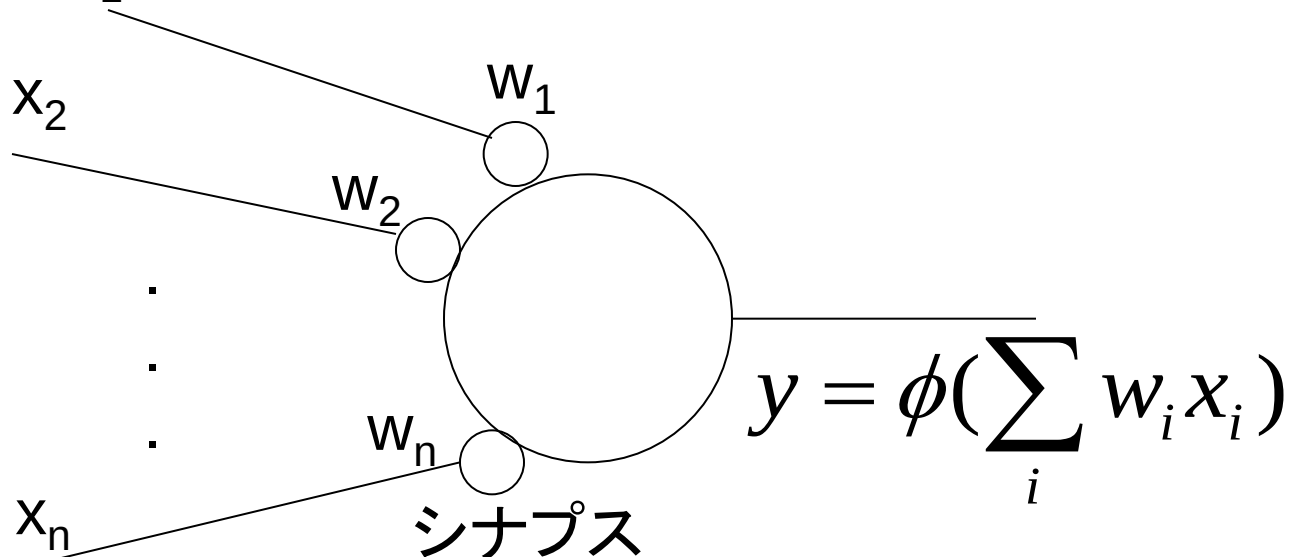
大脳皮質の解剖学的特徴[Pandya and Yeterian 1985][Gilbert 1983]

- 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る。



ニューロン（神経細胞）

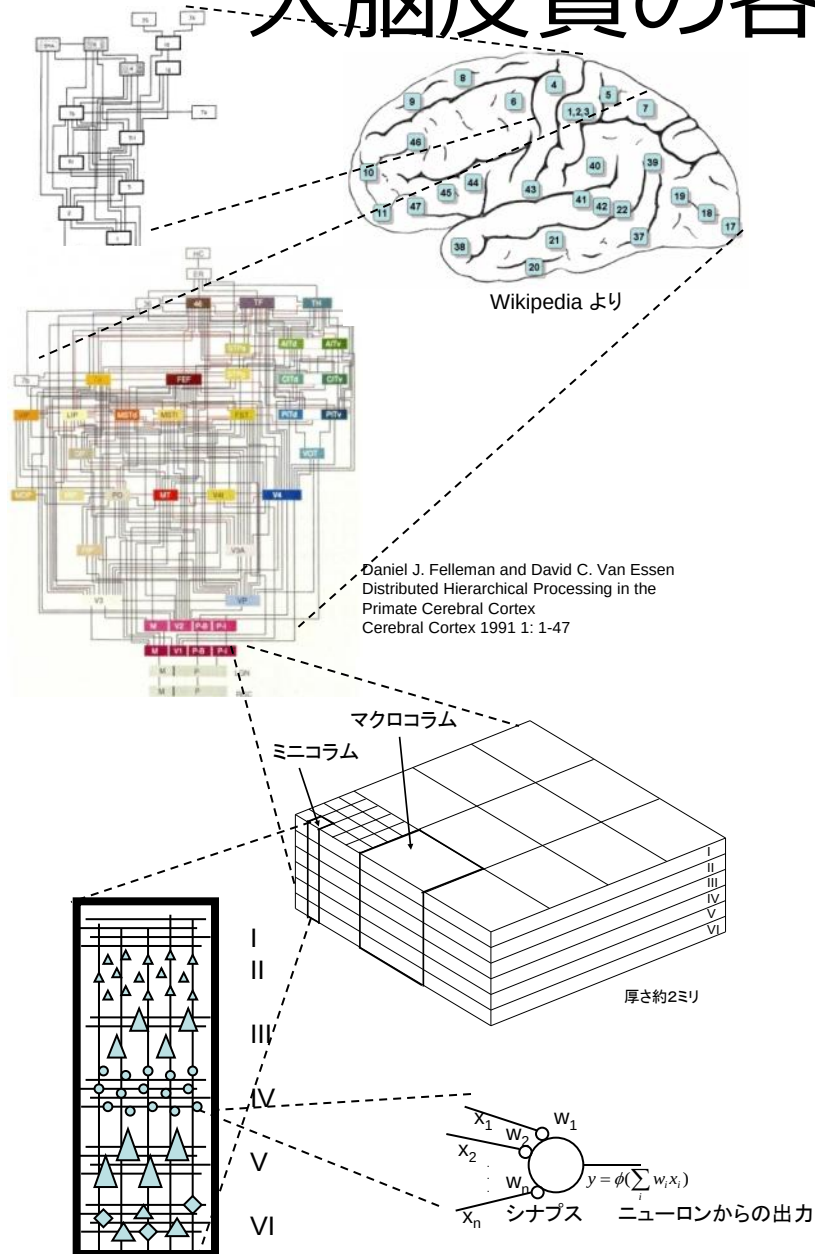
- ヒトの脳で 1 0 0 0 億個。
- 内積計算のような簡単な演算しか行えない。
- シナプスが w という重みを学習。



ニューロンへの入力

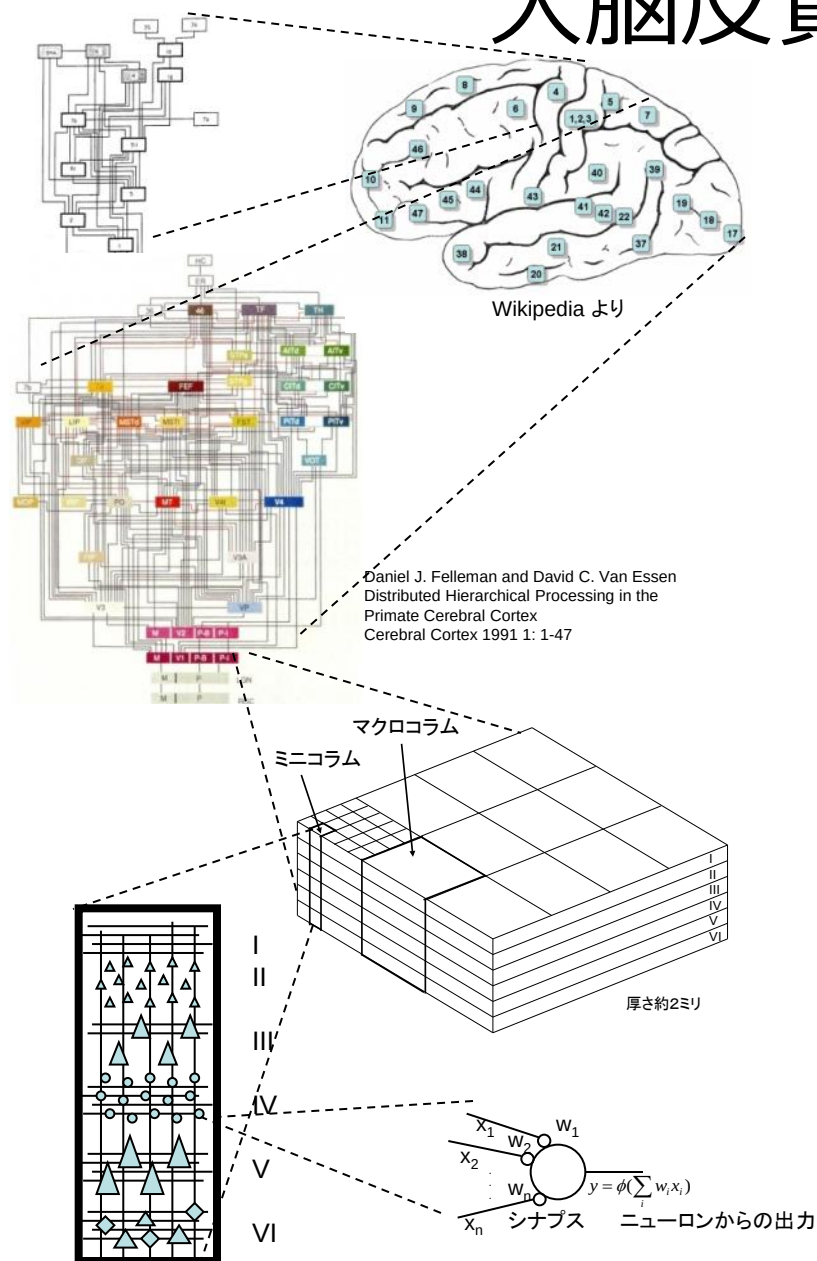
ニューロンからの出力

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

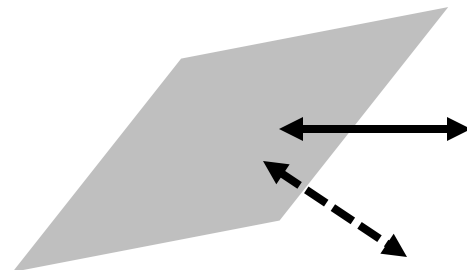


- 脳の様々な高次機能（認識、意思決定、運動制御、思考、推論、言語理解など）が、**たった50個程度**の領野のネットワークで実現されている。

大脳皮質の動作原理解明が最大の課題

脳とベイズ

脳がベイズ統計に従っている証拠の一例 [Weiss et al. 2002]



ひし形の運動方向の
認知に関わる錯視のモデル

Y. Weiss, E. P. Simoncelli, E. H. Adelson,
Motion illusions as optimal percepts. Nat. Neurosci.5, 598
(2002).

doi:10.1038/nn0602-858 pmid:12021763

デモ:「Moving Rhombus Displays by Yair Weiss」

<http://www.cs.huji.ac.il/~yweiss/Rhombus/rhombus.html>

参考: 日本語の解説

柴田和久, 神谷之康 ASCONE 講義-ベイズで読み解く知覚世界,
日本神経回路学会誌, 2007, 14(4):313-8

https://www.jstage.jst.go.jp/browse/jnns/14/4/_contents/-char/ja/

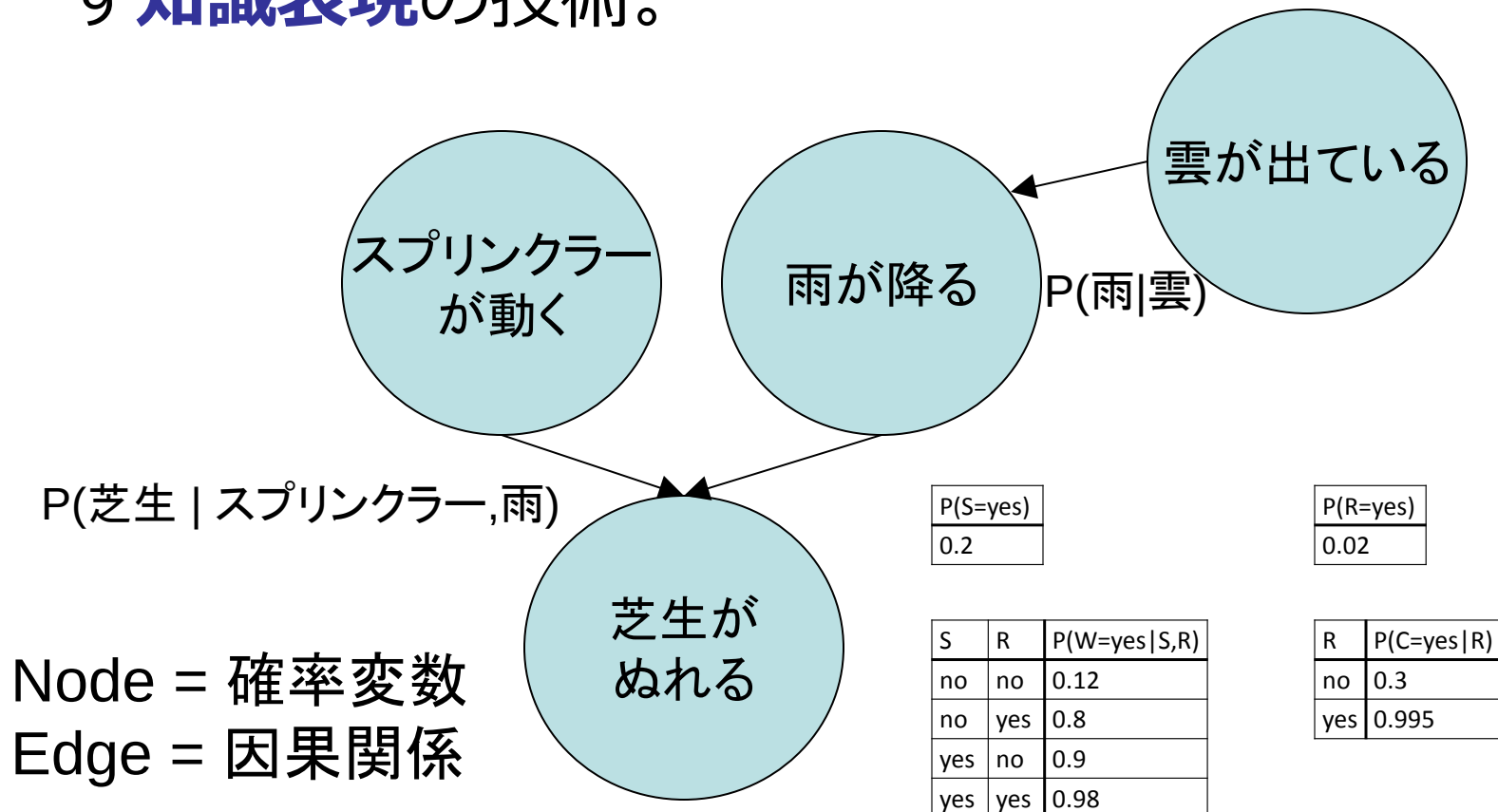
ベイズの効率化技術

- 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

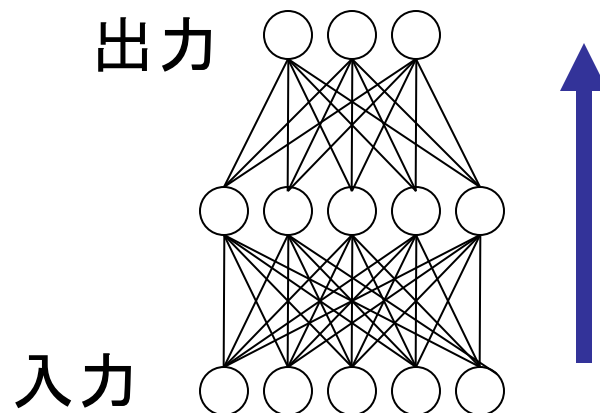
ベイジアンネットワーク [Pearl 1988] とは

- 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。



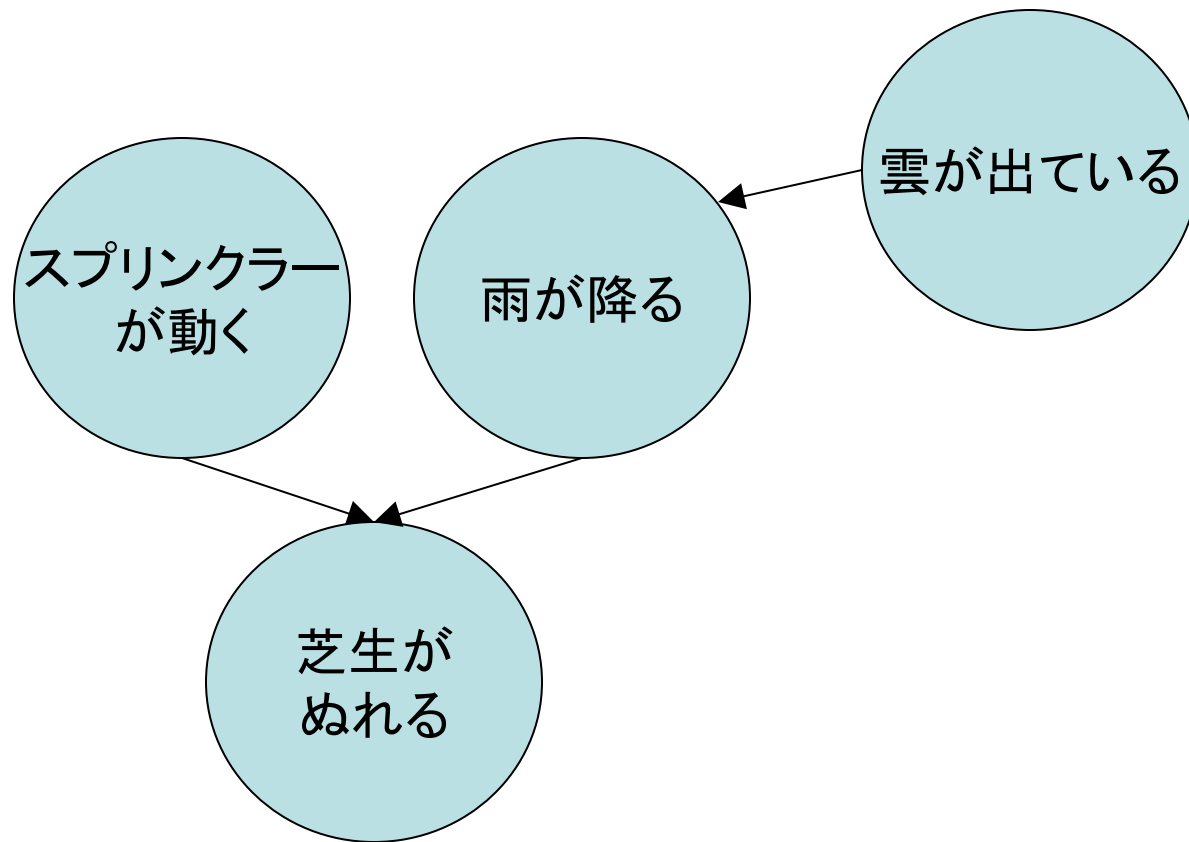
ベイジアンネットと ニューラルネットは違う

- 典型的なニューラルネット（多層パーセプトロン）は、情報が入力から出力への1方向。



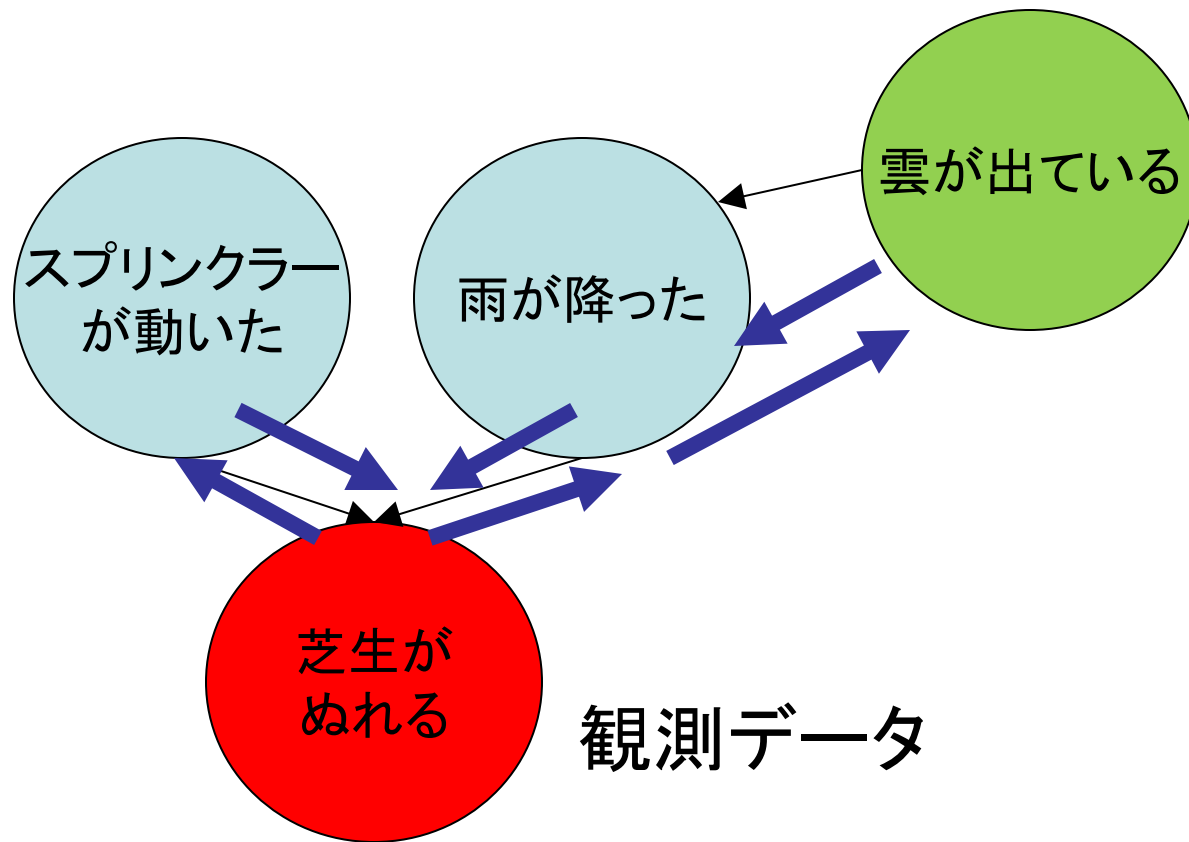
ベイジアンネットの動作

- すべてのノードが入力ノードにも出力ノードにもなり得る。
- 確率を扱う。



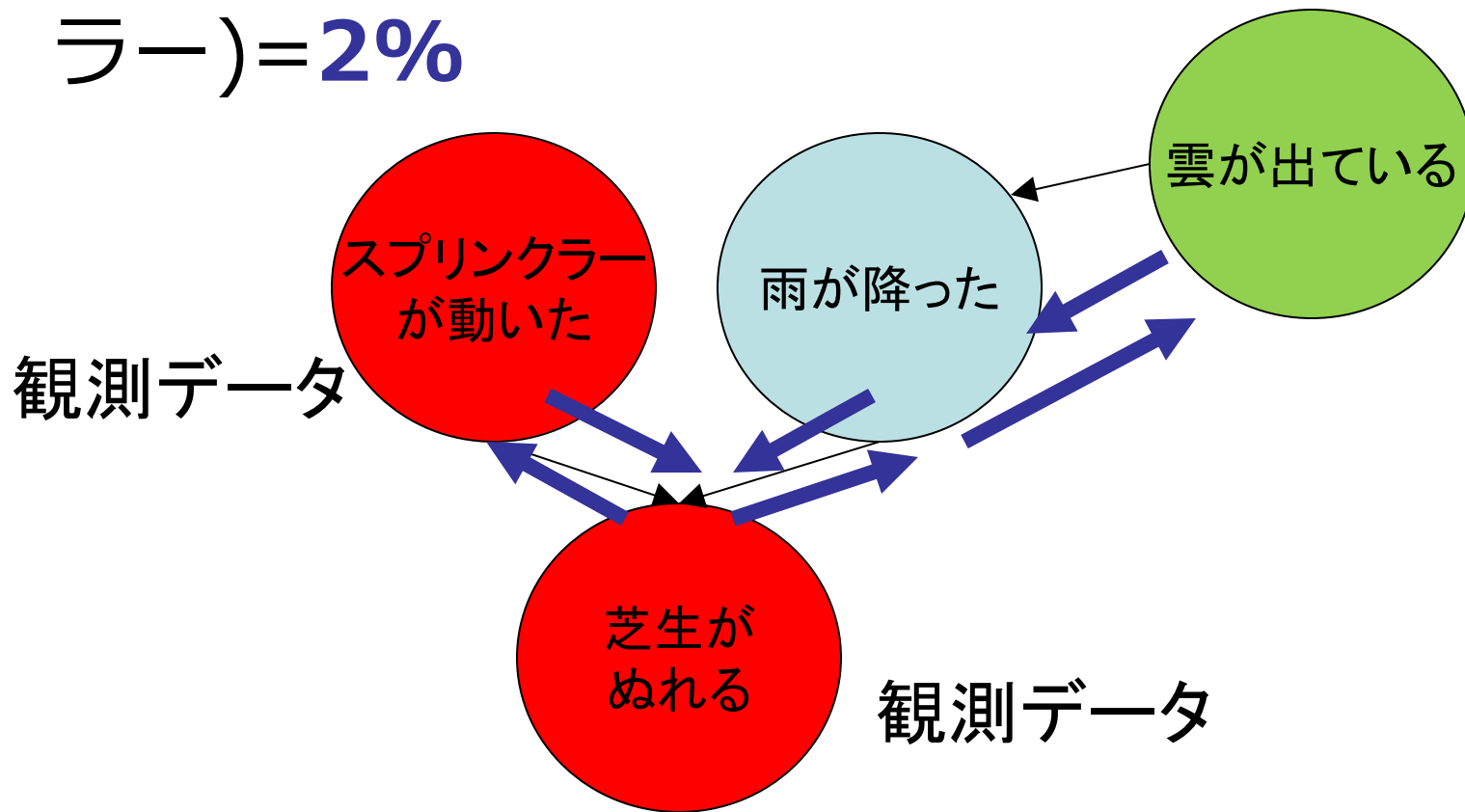
ベイジアンネットの動作

- 芝生が濡れていた時、雲が出ている確率は？
- 推論結果： $P(\text{雲} | \text{芝生}) = 87\%$



ベイジアンネットの動作

- 芝生が濡れていてスプリンクラーも動いている時、雲が出ている確率は？
- 推論結果： $P(\text{雲} | \text{芝生、スプリンクラー}) = 2\%$



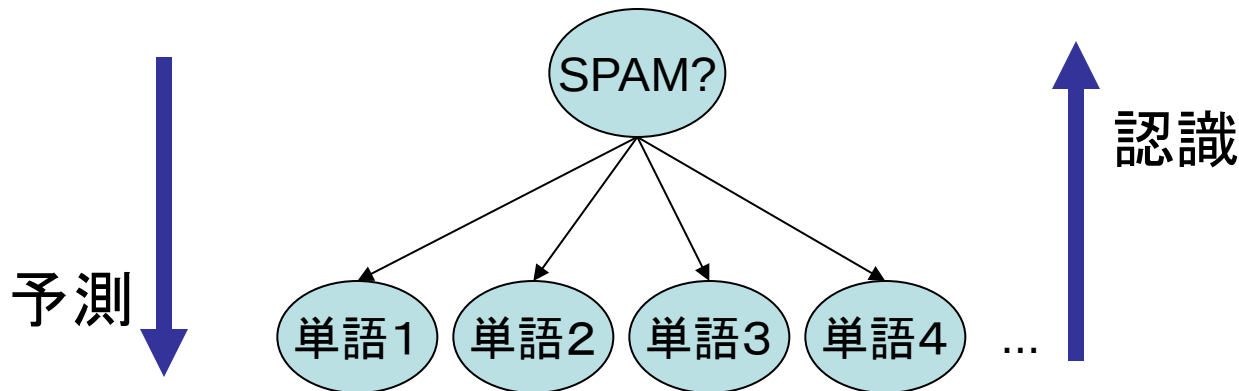
応用例：スパムフィルタ

- ナイーブベイズによる識別の例

$$p(C|F_1, \dots, F_n) = \frac{1}{Z} p(C) \prod_{i=1}^n p(F_i|C)$$

Wikipedia 単純ベイズ分類器

例 スпамメール



スパムメールは高い確率で
単語1を含む。

単語1を含むメールは高い確率で
スパムメールである。

大脳皮質ベイジアンネットモデル

大脳皮質とベイジアンネットの 類似点

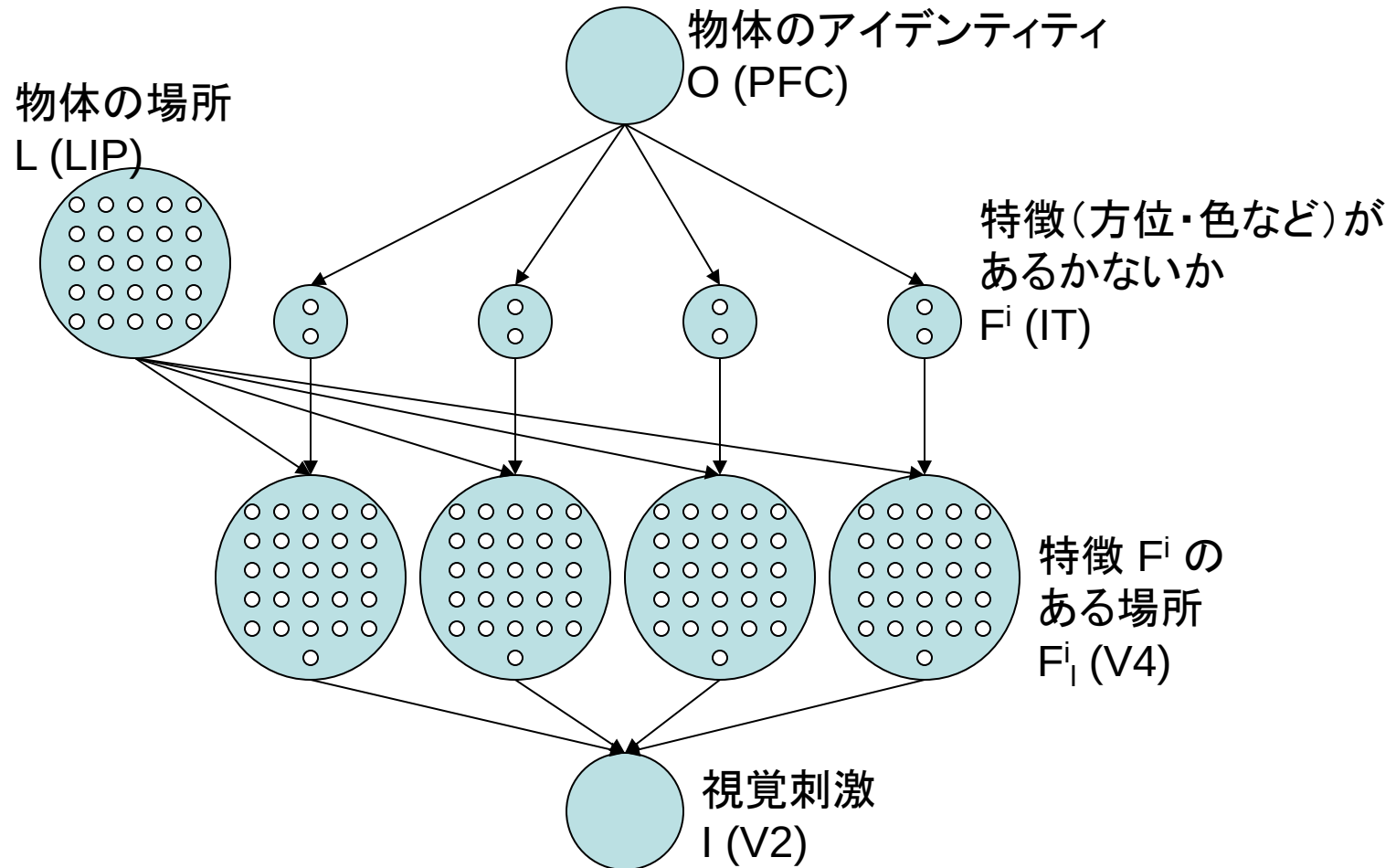
- トップダウンとボトムアップの非対称の接続
- 局所的かつ非同期な情報のやり取りだけで動作
- 値が非負
- 情報が正規化される
- ヘブ則学習
- 文脈や事前知識に依存した認識
- ベイズに基づく動作

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



Chikkerur のモデル まとめ

- 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
 - 心理物理学のモデル
 - 視線がどこに向きやすいか。
 - 注意を向けることで結果認識がどう変わるか。
 - . . .
 - 電気生理学のモデル
 - 注意を向けることでニューロン応答がどう変わるか。

確率伝搬アルゴリズムと 解剖学的構造との対応 [Ichisugi 2007]

確率伝搬アルゴリズム[Pearl 1988]

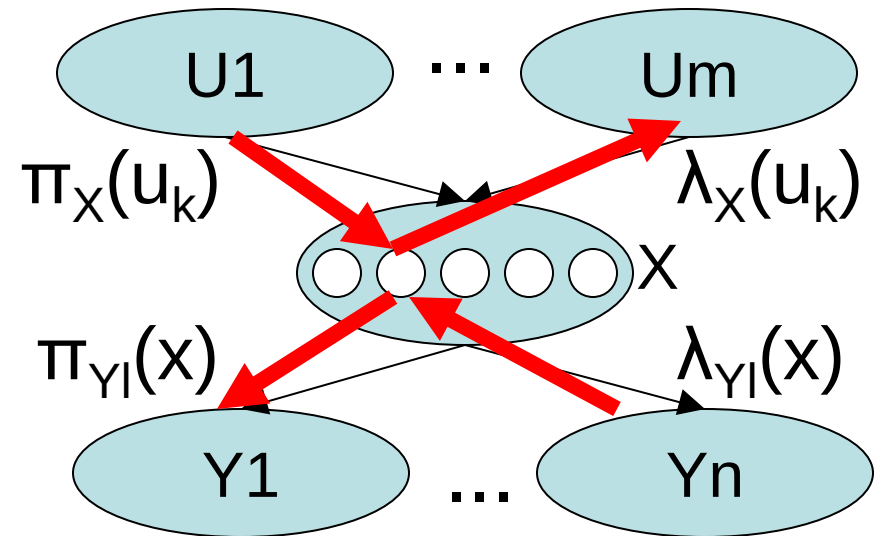
$$BEL(x) = \alpha \lambda(x) \pi(x)$$

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

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

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

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



神経回路で実行するには複雑

近似確率伝搬アルゴリズム [Ichisugi 2007]

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

$$\mathbf{l}_{XY}^{t+1} = \mathbf{z}_Y^t + \mathbf{W}_{XY} \mathbf{o}_Y^t$$

$$\mathbf{o}_X^{t+1} = \bigotimes_{Y \in \text{children}(X)} \mathbf{l}_{XY}^{t+1}$$

$$\mathbf{k}_{UX}^{t+1} = \mathbf{W}_{UX}^T \mathbf{b}_U^t$$

$$\mathbf{p}_X^{t+1} = \sum_{U \in \text{parents}(X)} \mathbf{k}_{UX}^{t+1}$$

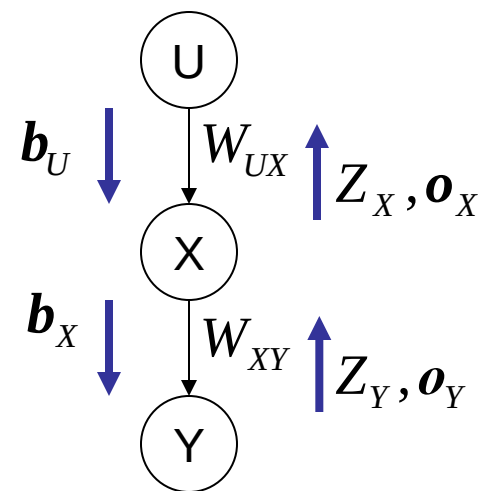
$$\mathbf{r}_X^{t+1} = \mathbf{o}_X^{t+1} \otimes \mathbf{p}_X^{t+1}$$

$$Z_X^{t+1} = \sum_i (\mathbf{r}_X^{t+1})_i \quad (= \|\mathbf{r}_X^{t+1}\|_1 = \mathbf{o}_X^{t+1} \bullet \mathbf{p}_X^{t+1})$$

$$\mathbf{z}_X^{t+1} = (Z_X^{t+1}, Z_X^{t+1}, \dots, Z_X^{t+1})^T$$

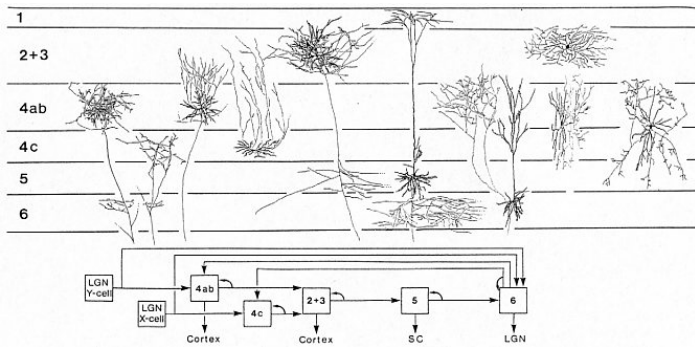
$$\mathbf{b}_X^{t+1} = (1/Z_X^{t+1}) \mathbf{r}_X^{t+1}$$

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

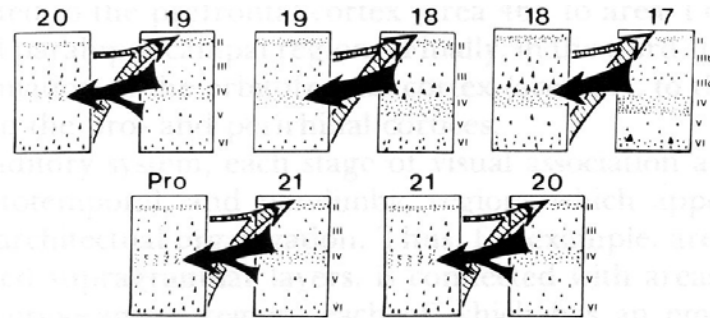


- ・神経回路で実現可能
- ・大規模化可能な計算量・記憶量

コラム構造・6層構造との一致



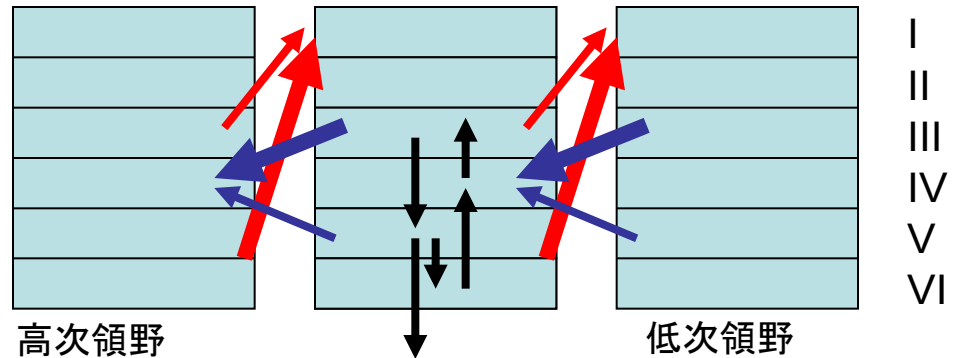
[Gilbert 1983]



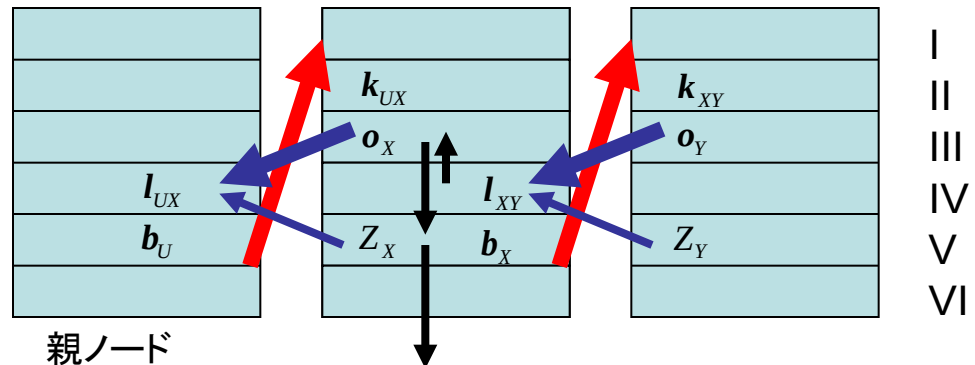
[Pandya and Yeterian 1985]

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.

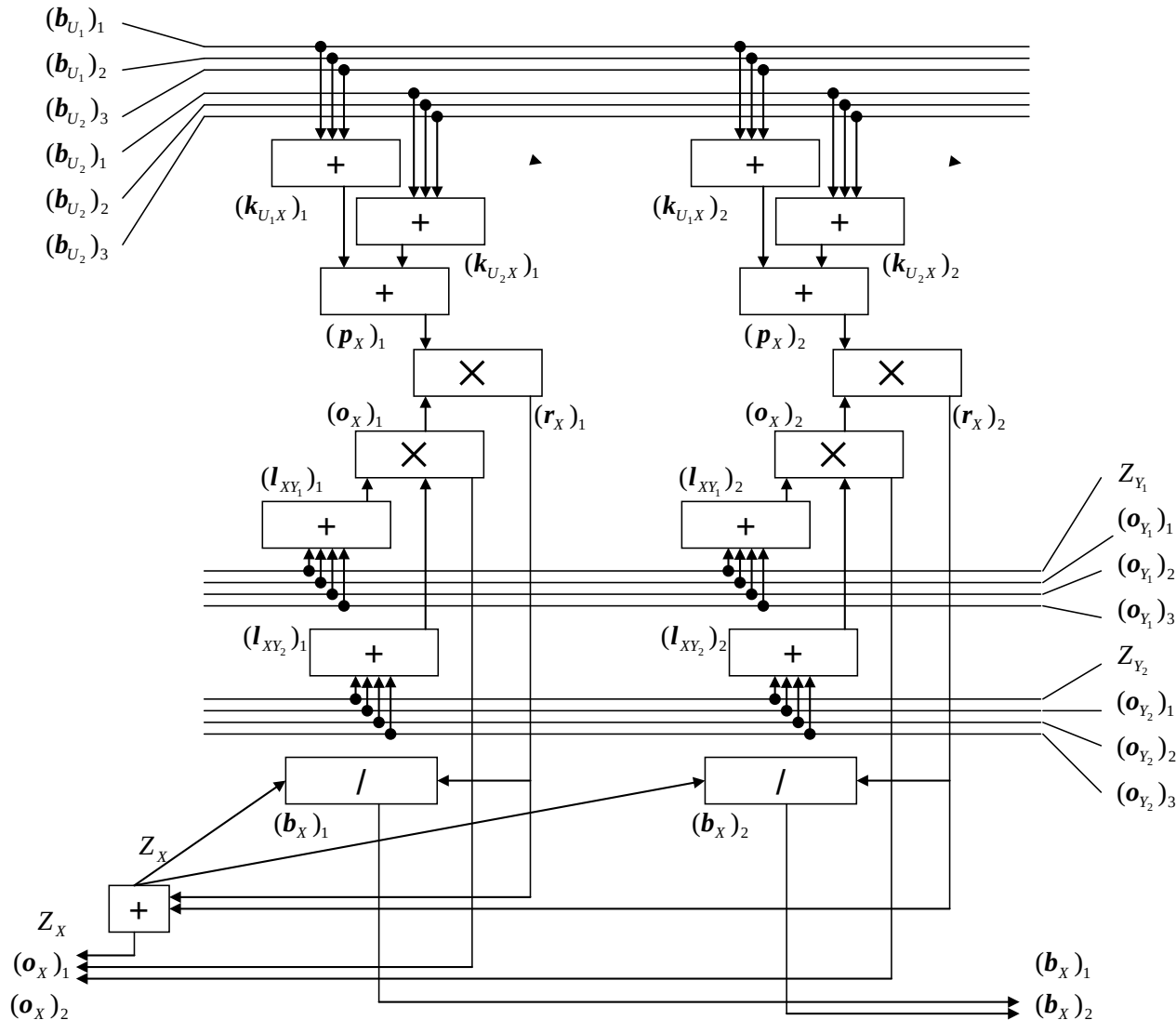


解剖学的構造



モデル

各変数の値を計算する回路



I

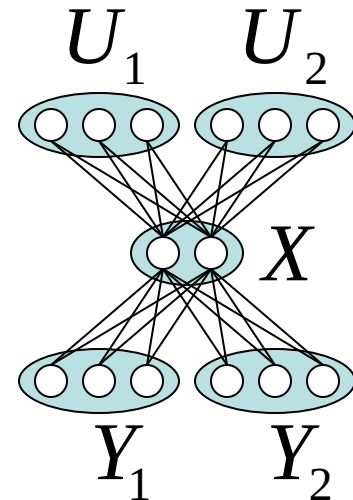
II

III

IV

V

VI

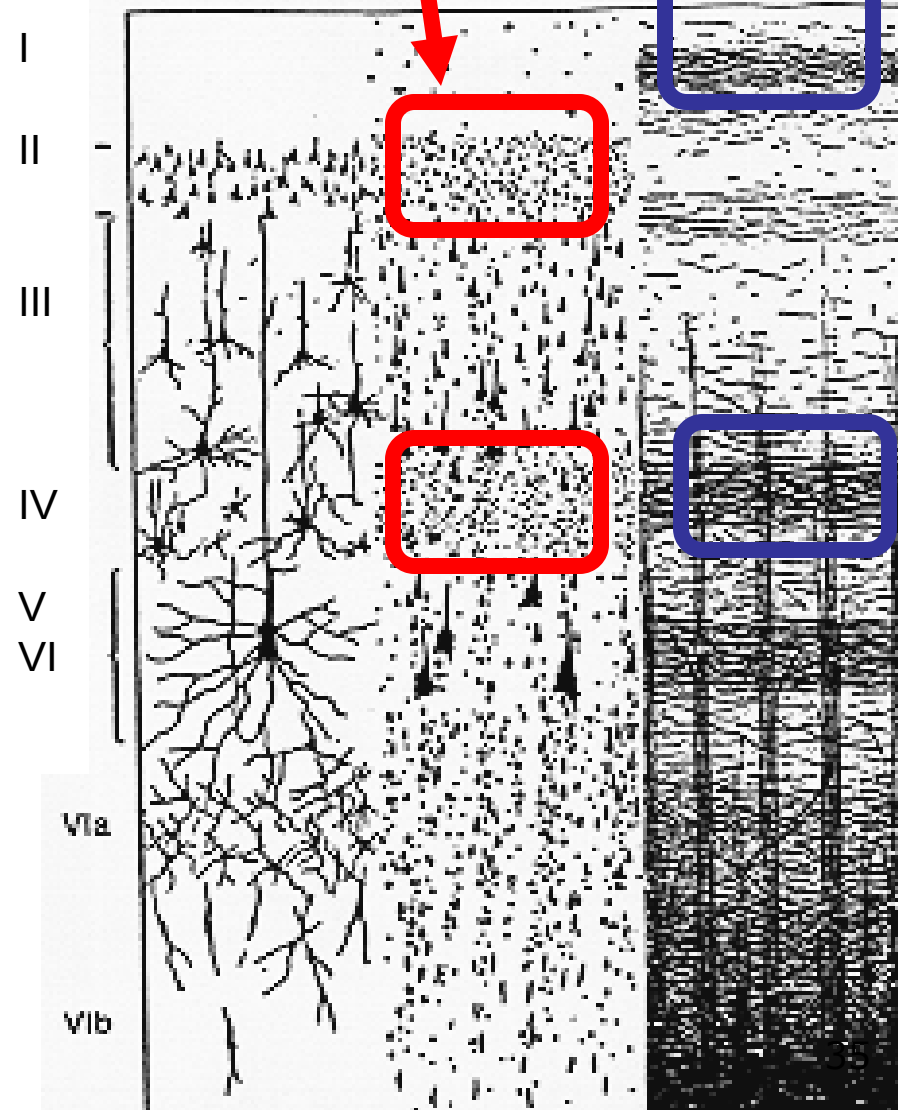
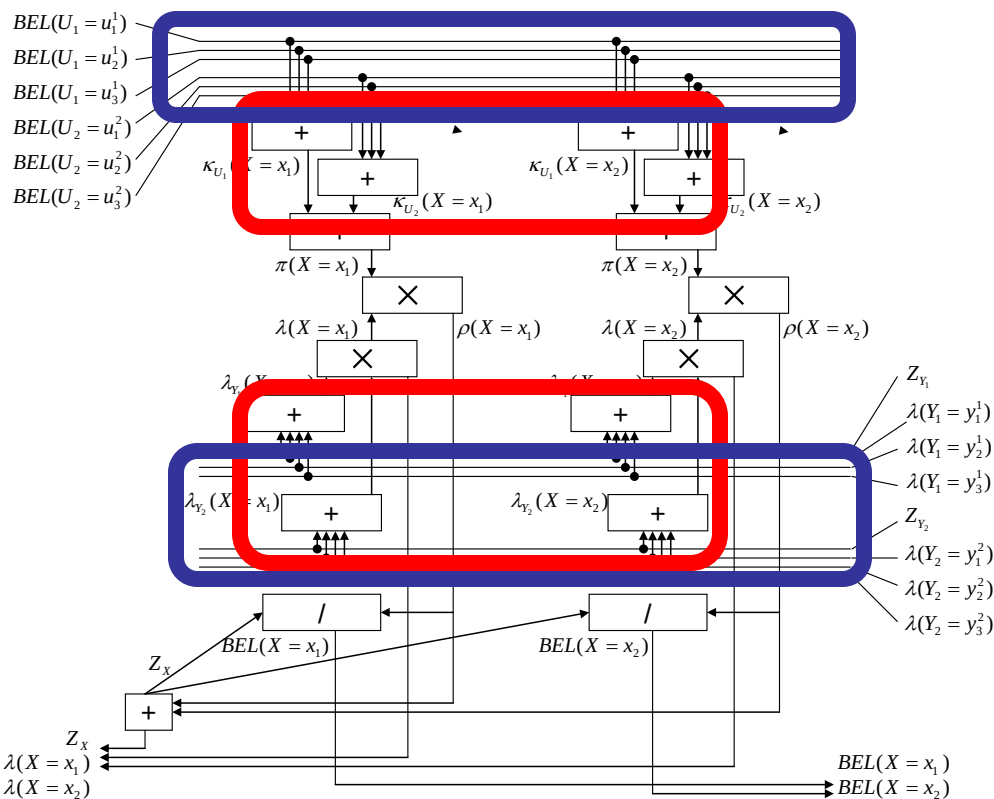


左図は、上記
BESOMネットワ
ークのノードXのユ
ニットの値を計算
する回路

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>

ここまでのまとめ

- 大脳皮質はおそらく、ある種のベイジアンネット。
- 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。

より詳しい情報:「脳とベイジアンネット」

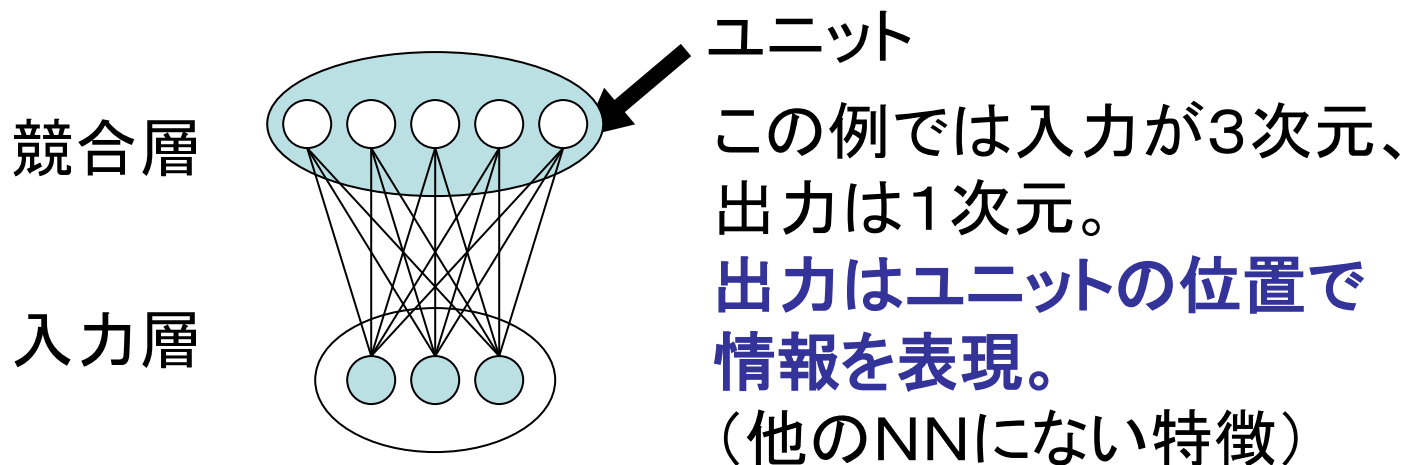
<https://staff.aist.go.jp/y-ichisugi/besom/j-index.html>

BESOMモデル

SOM（自己組織化マップ）

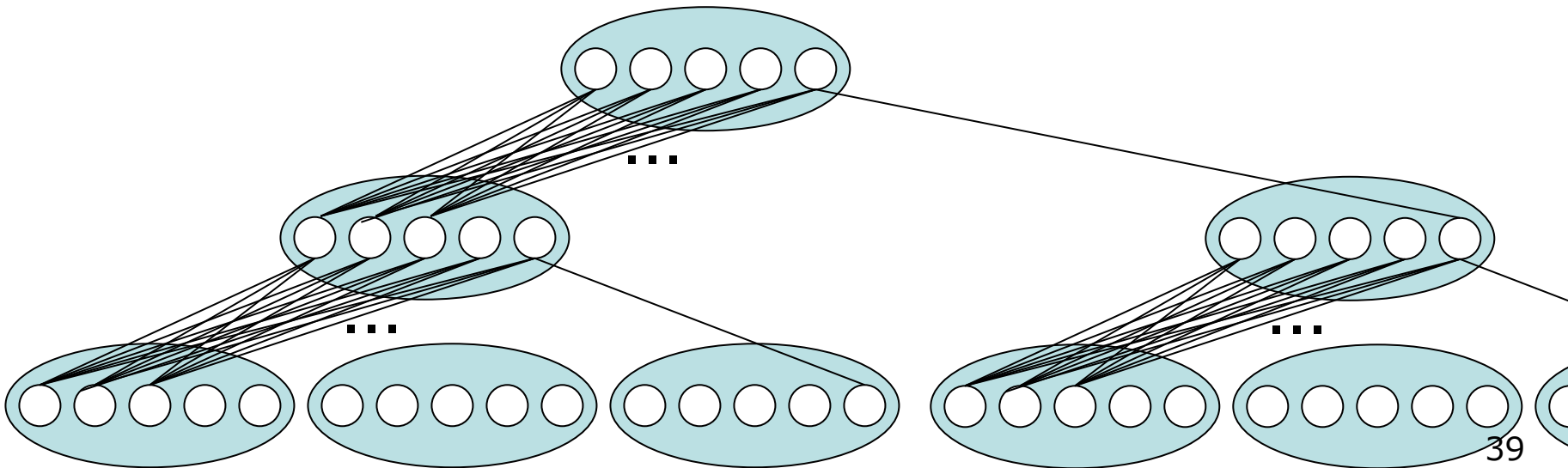
[Kohonen 1995]

- 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。
- 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。



2つの神経科学的知見

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。



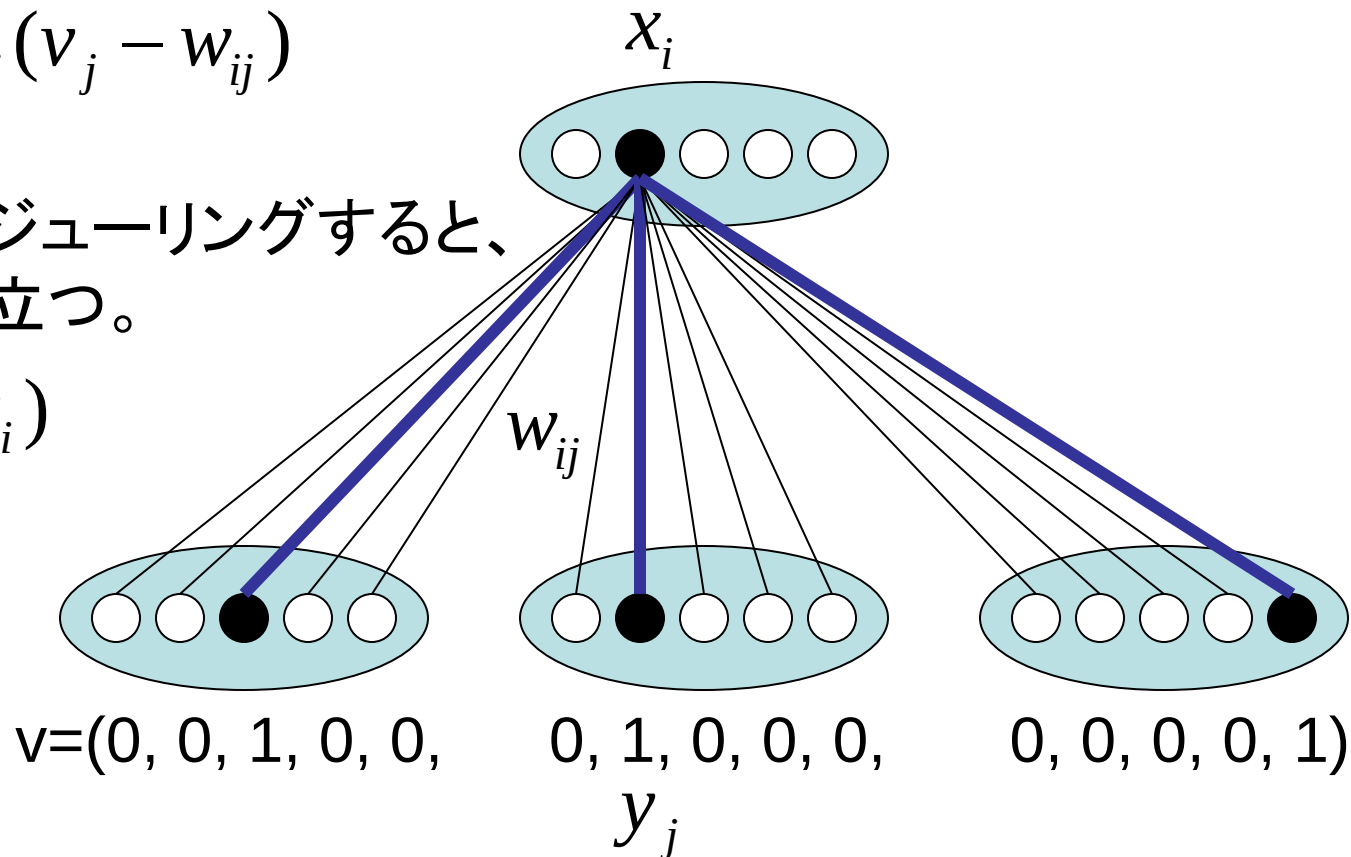
SOMの階層の学習結果が ベイジアンネットになる [Ichisugi 2007]

近傍学習を無視すると学習則は

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

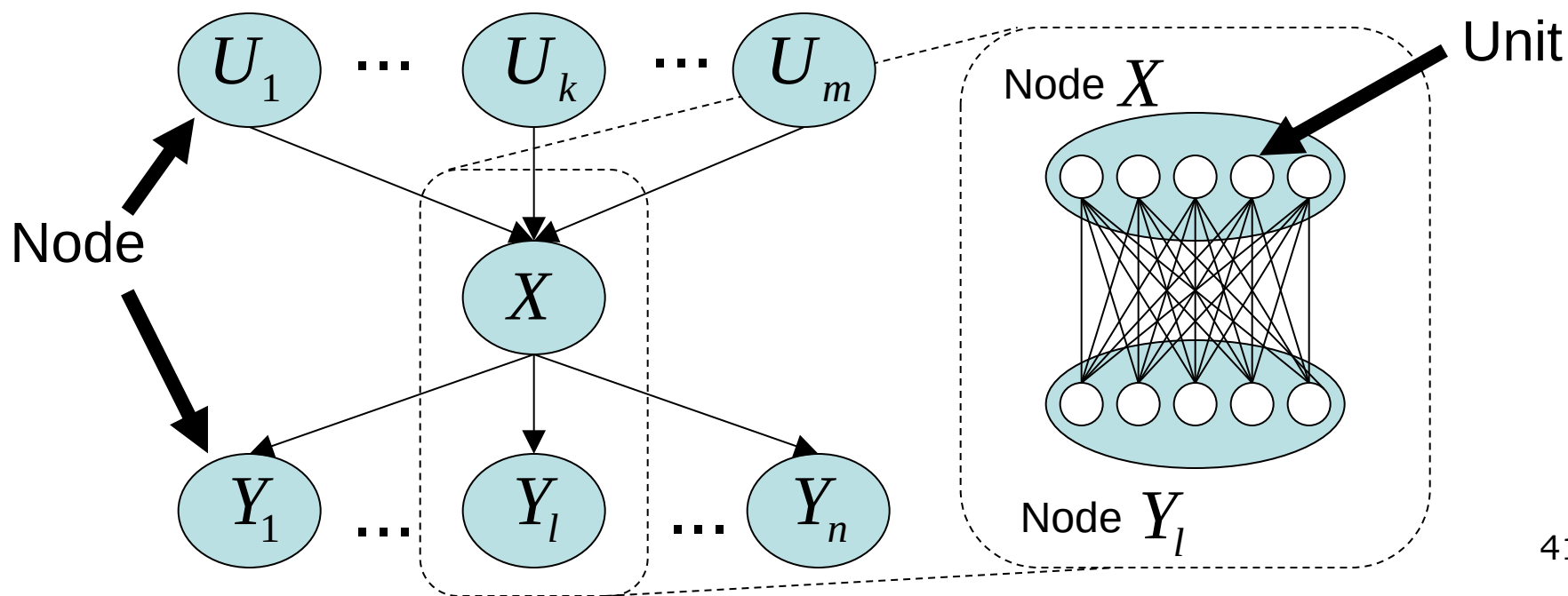
α_i を適切にスケジューリングすると、
下記の式が成り立つ。

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



BESOM, 自己組織化マップ, ベイジアンネット, 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	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



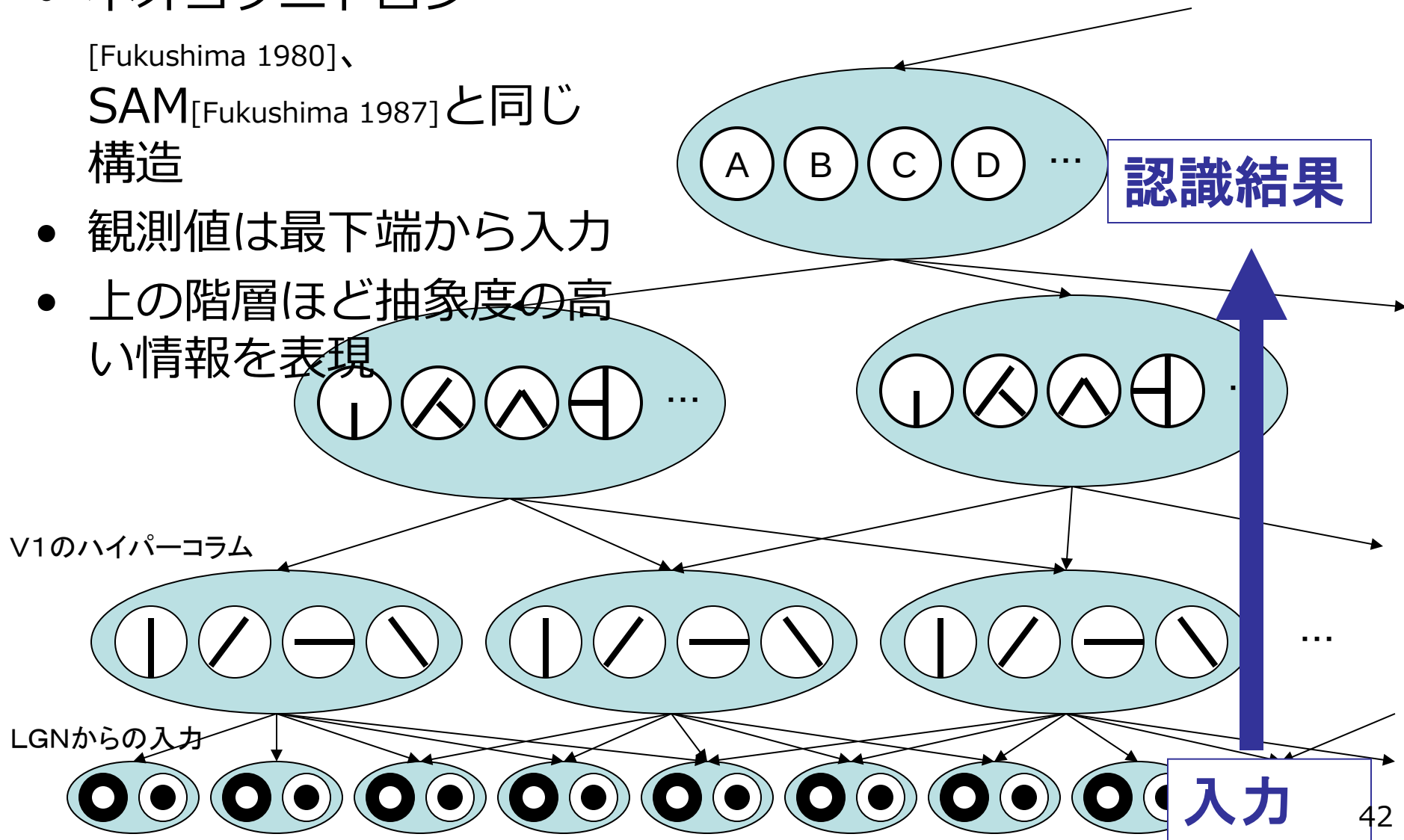
視覚野の模式的説明

- ネオコグニトロン

[Fukushima 1980]、

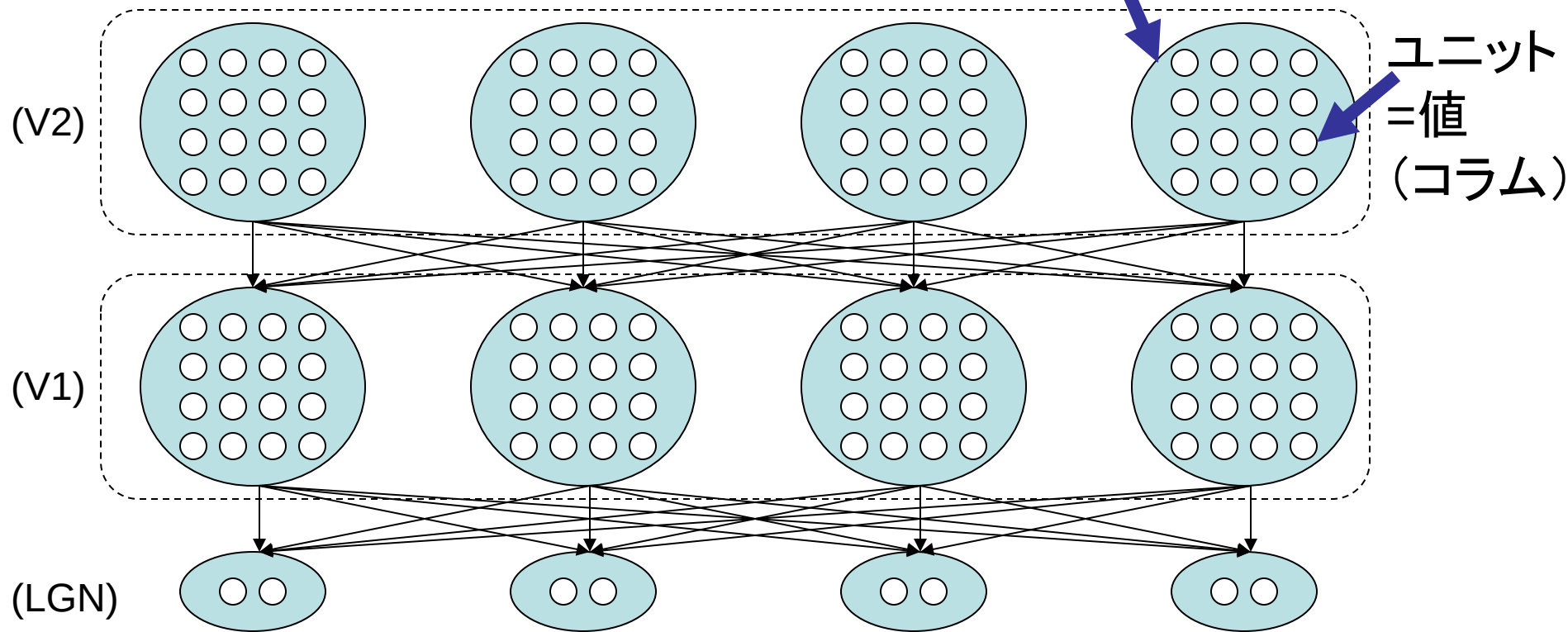
SAM[Fukushima 1987]と同じ
構造

- 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

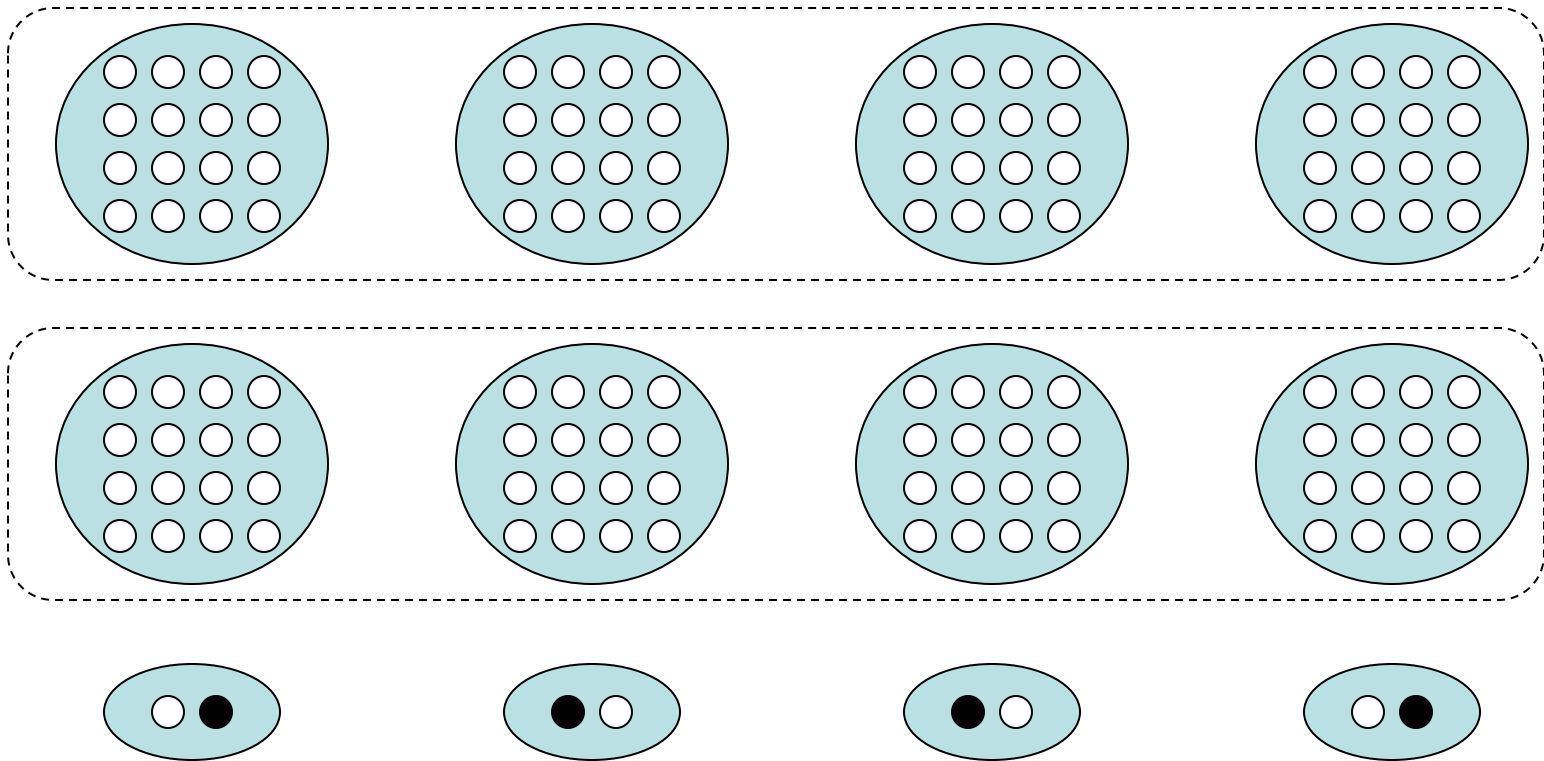
ノード=確率変数=SOM(ハイパーコラム)



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。

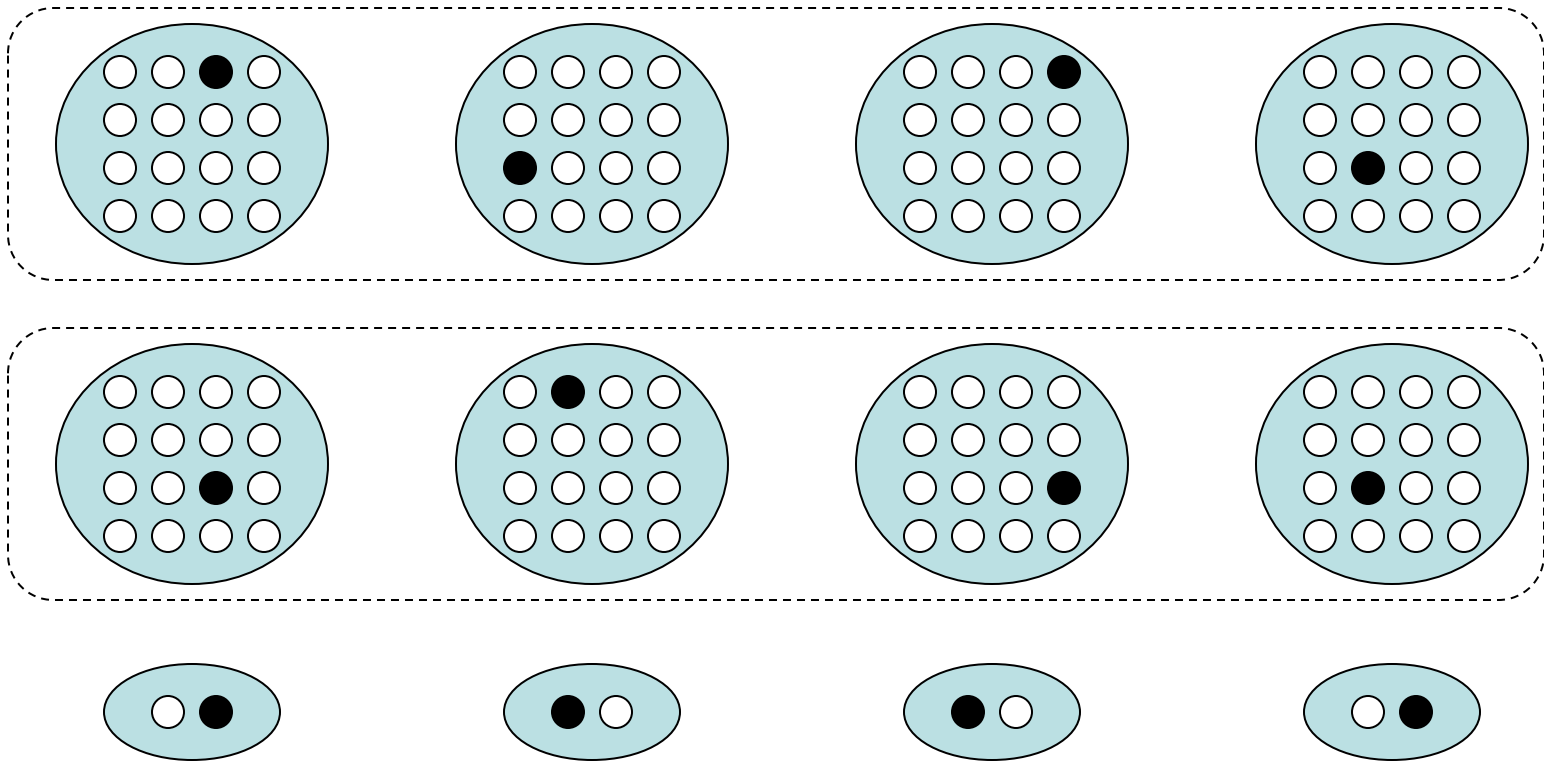
結合の重み
=条件付き確率
(シナプスの重み₄₃)

BESOM の動作



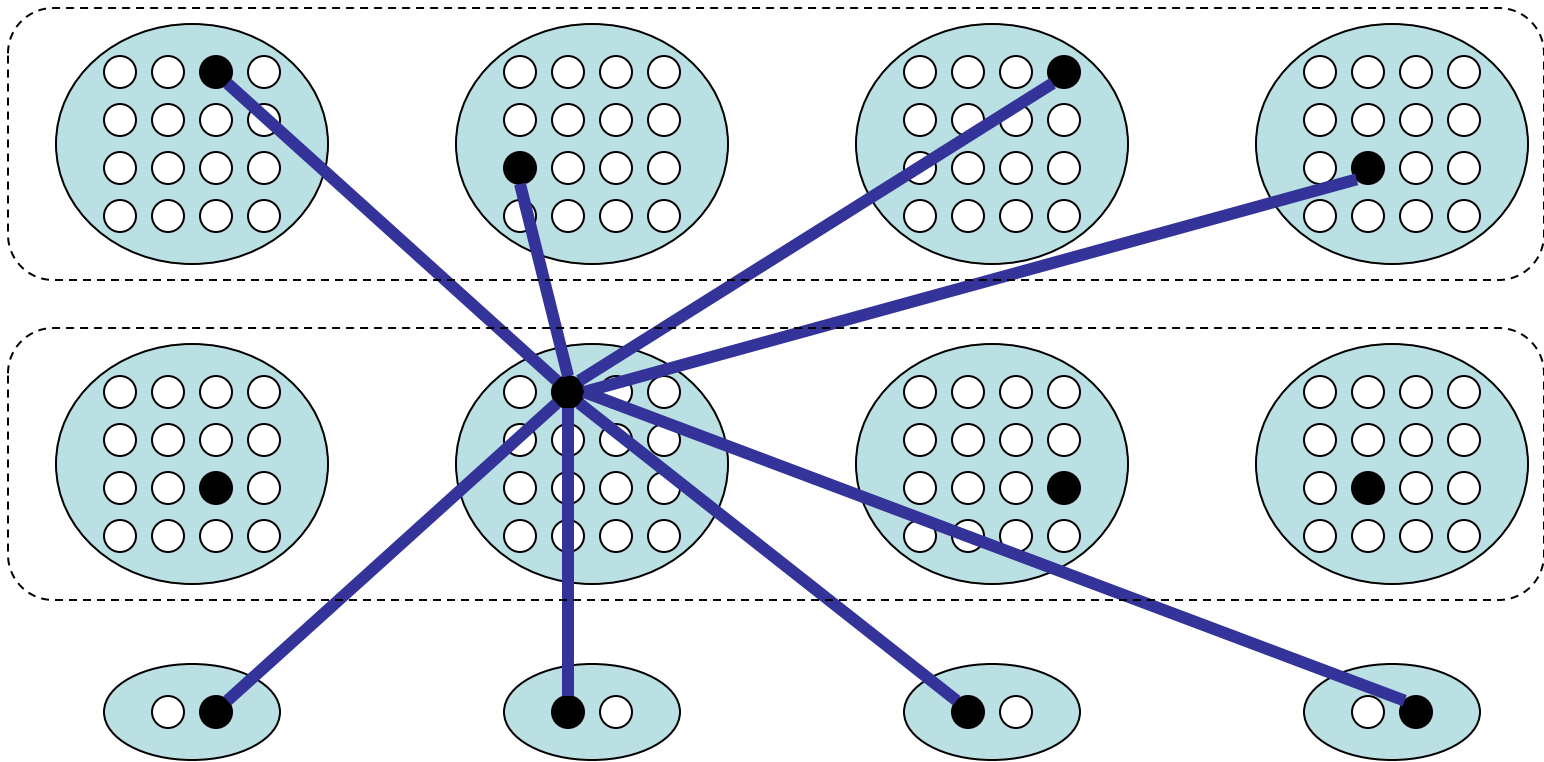
再下端のノードに観測値が与えられる。

BESOM の動作



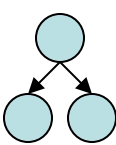
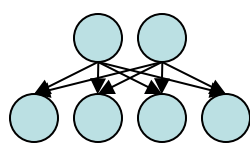
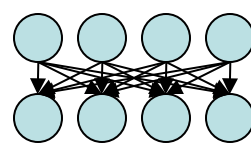
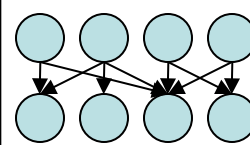
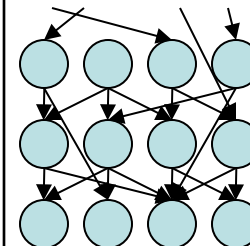
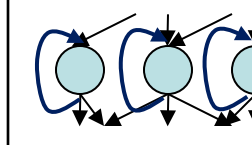
認識：観測データとの同時確率が最大となる値の組み合わせ(MPE: most probable explanation)を求める。

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 ネットのシミュレーションの現状

ネットワークの構造						
機能	近傍学習	非線形ICA	スパース符号化、混合分布	構造学習、特徴選択	多層化	短期記憶、SFA (未発表)
実装	2007年	2009年	2009年	2011年	2012年	プロトタイプ 2013年
隠れノード数	1	2~3	3~10	3~10	10~100	10~100

計算量のオーダー(ノード数 n のとき1入力の認識・学習に必要な計算量)

- ・ 2007年 $O(2^n)$ 全数探索によるMPE計算
- ・ 2009年 $O(n^4)$ 山登り法
- ・ 2012年 ほぼ $O(n)$ 最適化BP、オンラインエッジ選択

認識アルゴリズム OOBP

[Ichisugi, Takahashi 2015]

Belief Propagation [Pearl 1988]

$$\begin{aligned}
 BEL(x) &= \alpha \lambda(x) \pi(x) \\
 \pi(x) &= \sum_{u_1, \dots, u_m} P(x|u_1, \dots, u_m) \prod_k \pi_X(u_k) \\
 \lambda(x) &= \prod_l \lambda_{Y_l}(x) \\
 \pi_{Y_l}(x) &= \beta_1 \pi(x) \prod_{j \neq l} \lambda_{Y_j}(x) \\
 \lambda_X(u_k) &= \beta_2 \sum_x \lambda(x) \\
 &\quad \sum_{u_1, \dots, u_m / u_k} P(x|u_1, \dots, u_m) \prod_{i \neq k} \pi_X(u_i)
 \end{aligned}$$

OOBP

$$\begin{aligned}
 \lambda_{Y_l}^{t+1}(x) &= \beta_2 \sum_{y_l} \lambda^t(y_l) (\pi^t(y_l) - \kappa_X^t(y_l) + w(y_l, x)) \\
 \lambda^{t+1}(x) &= \prod_{l=1}^n \lambda_{Y_l}^{t+1}(x) \\
 \pi_{Y_l}^{t+1}(x) &= \beta_1 \rho^{t+1}(x) / \lambda_{Y_l}^{t+1}(x) \\
 \kappa_{U_k}^{t+1}(x) &= \sum_{u_k} w(x, u_k) \pi_X^t(u_k) \\
 \pi^{t+1}(x) &= \sum_{k=1}^m \kappa_{U_k}^{t+1}(x) \\
 \rho^{t+1}(x) &= \lambda^{t+1}(x) \pi^{t+1}(x) \\
 BEL^{t+1}(x) &= \alpha \rho^{t+1}(x)
 \end{aligned}$$

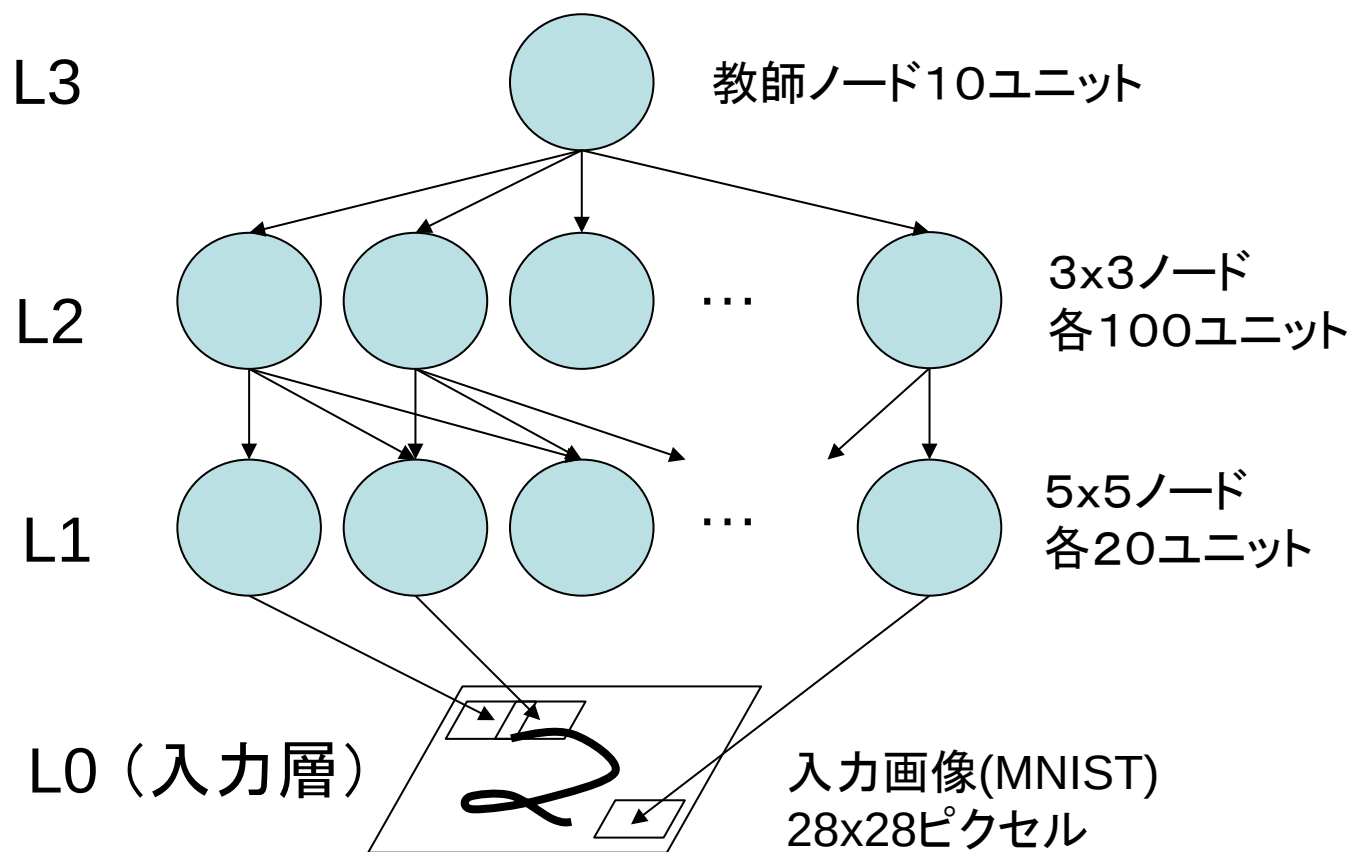
親ノードの数 m に対し
 $O(2^m)$ の計算量

制限された条件付確
率表のモデルを
前提に最適化

$O(m)$ の計算量

$$P(x|u_1, \dots, u_m) = \frac{1}{m} \sum_{k=1}^m P(x|u_k)$$

4層「制限付き」ベイジアンネット 手書き数字認識



$$P(X | U_1, \dots, U_m)$$

$$= \frac{1}{m} \sum_{i=1}^m P(X | U_i)$$

Neocognitron,
Deep Learning
と同じ深い構造の
ベイジアンネット

ノード数にたいして
1入力の処理が
ほぼ線形時間で動作

注: EMを使う場合、現在 $O(n^2)$

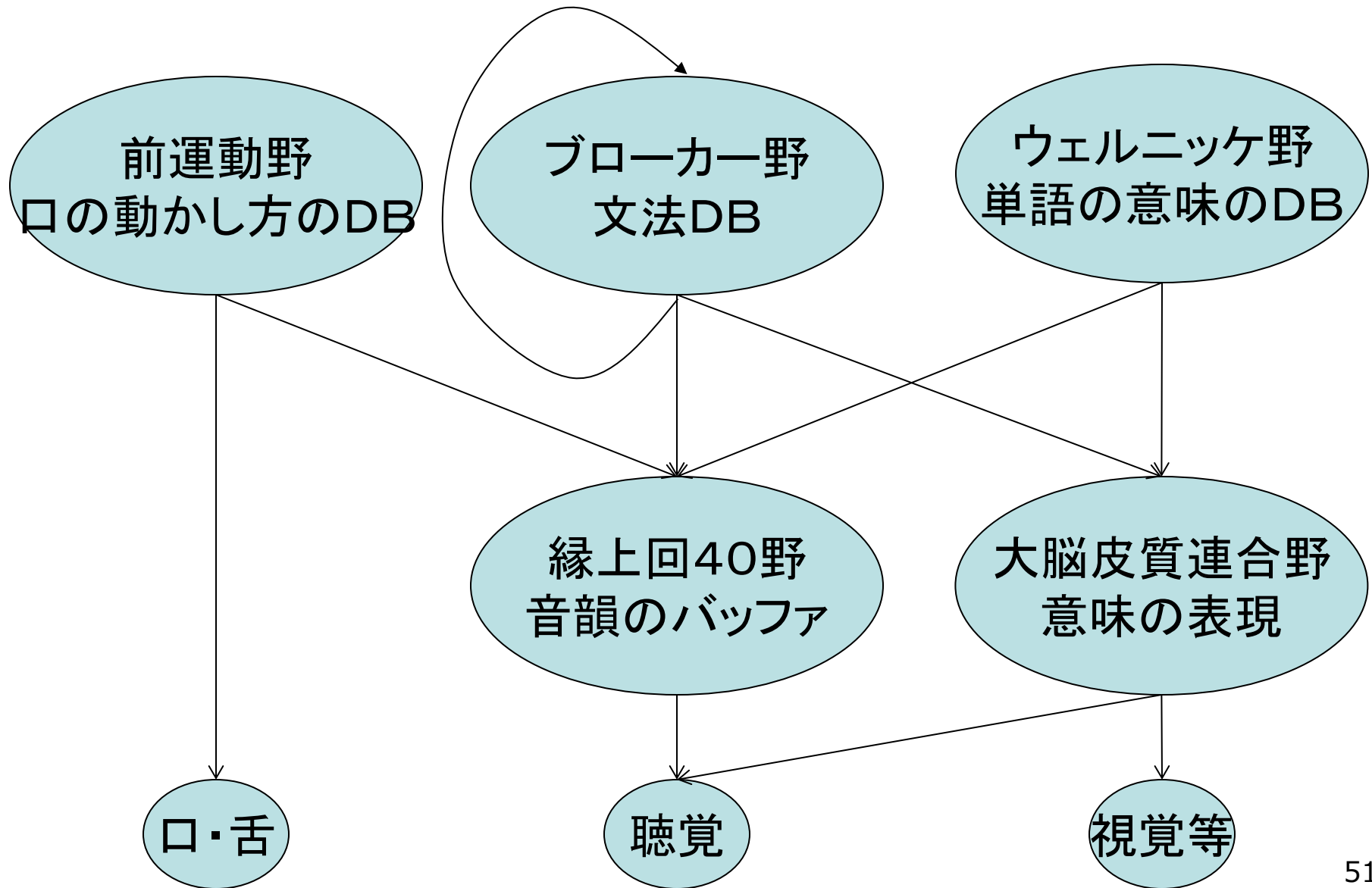
pre-training なし、認識アルゴリズムは
OOBP、EMアルゴリズムで学習
認識率: 92%程度

参考:
linear classifier (1-layer NN) 88%
Ciresan et al. CVPR 2012 99.87%
「MNIST handwritten digit database」
<http://yann.lecun.com/exdb/mnist/>

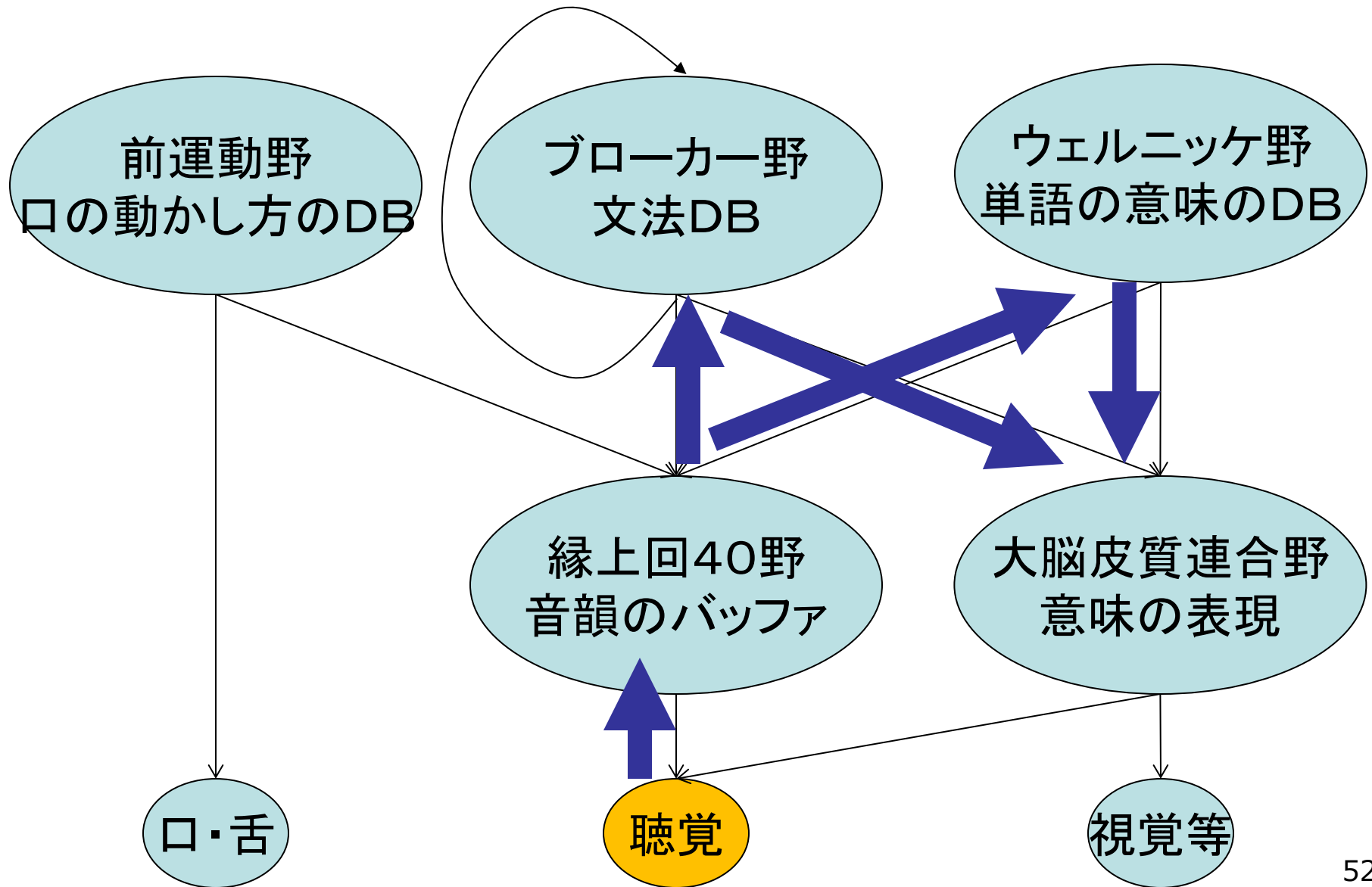
「Deep Learning の構造を持ったベイジアン ネット」が工学的にも有望と考える理由

- スケーラブル：ノード数に対し線形時間で 1 入力を認識・学習。
 - 脳と同様、並列処理によりリアルタイム動作が可能。
- Feed forward NN より高機能。
 - 時間的・空間的文脈の情報が自然に扱える。
 - 半教師あり学習、欠損データの学習が自然に扱える。
- 生成モデルを素直に表現可能。
 - 様々な事前知識を作り込みやすい。
 - 神経科学的知見からも作り込むべき事前知識のヒントが得られる。

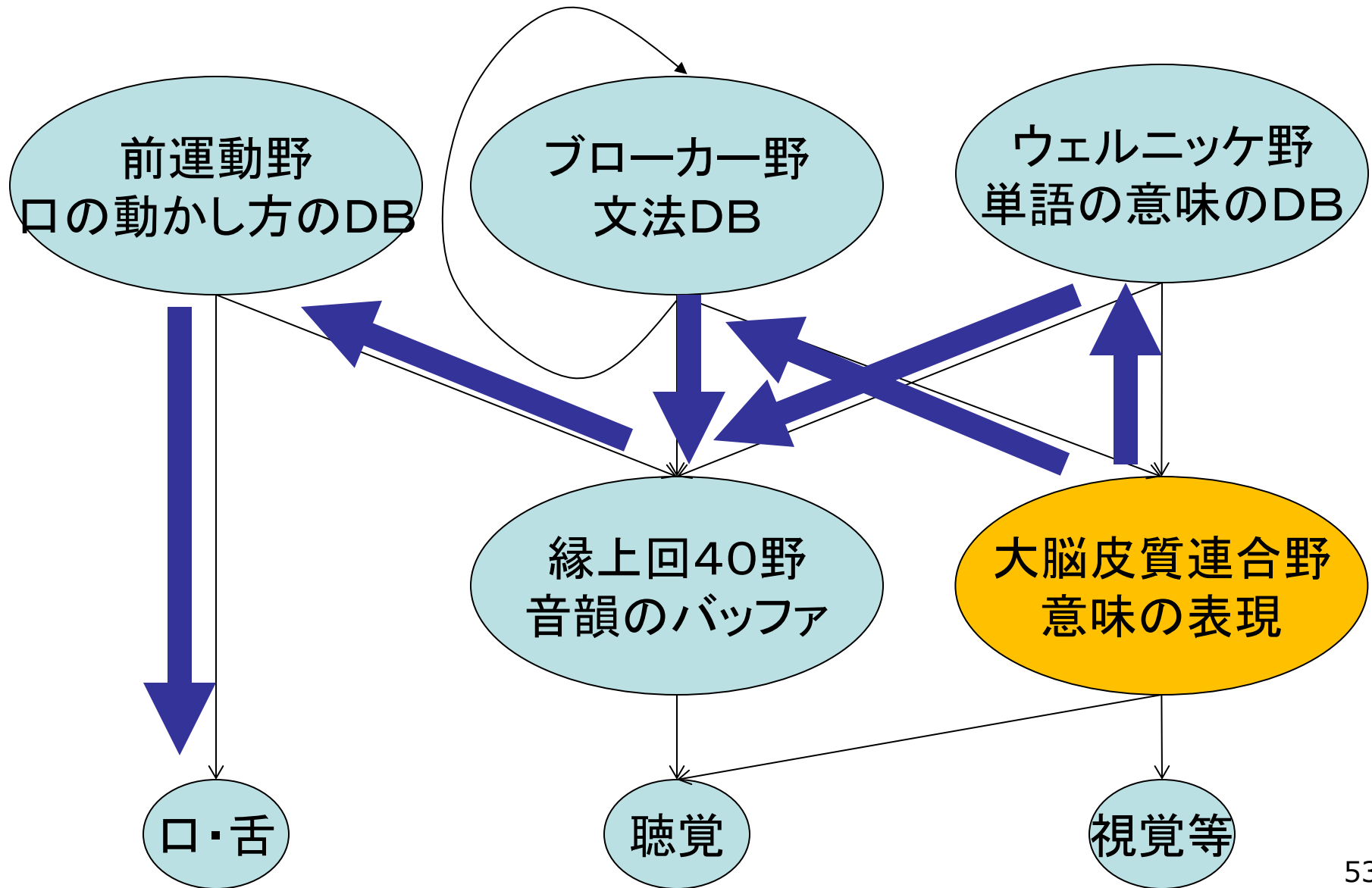
言語野のベイジアンネット？



言語理解



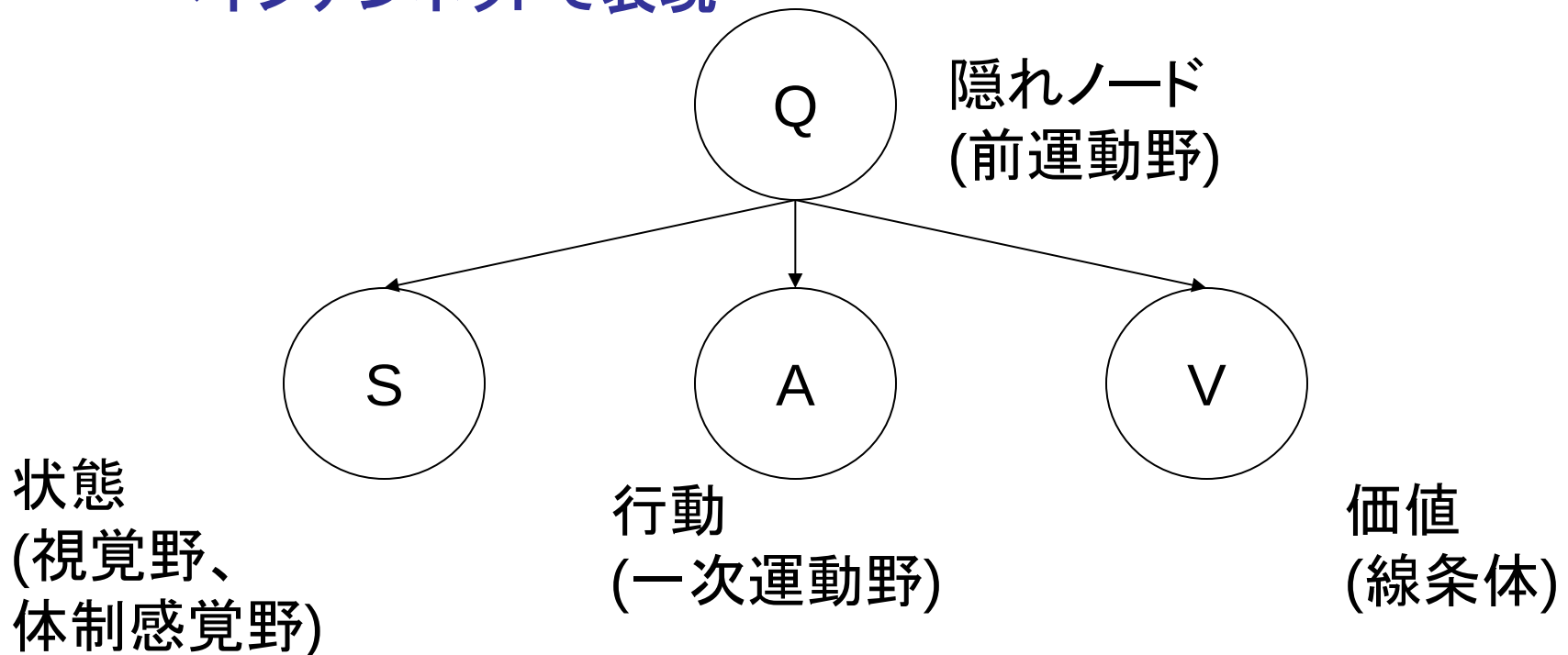
発話



運動野のベイジアンネットモデル

[Ichisugi 2012]

ノードQが競合学習で行動価値関数 $Q(s,a)$ を学習、
ベイジアンネットで表現



$$\begin{aligned} P(q, s, a, V = 1) \\ = P(s | q)P(a | q)P(V = 1 | q)P(q) \end{aligned}$$

関連研究：大脳皮質のアルゴリズムのモデル

	多層	ベイズ	双方向	近傍学習	ICA	強化学習
Neocognitron	○		○*			
SOM				○	○*	○*
sparse-coding		○		○*	○	
HMAX	○					
predictive-coding	○	○	○			
HTM	○	○	○			
DBN, DBM	○	○	○			
BESOM	○	○	○	○	○-	○

大脳皮質の認識・学習の機構を表現し
計算機上で実行できる主なモデル。

○* は後続研究による拡張
○- は英文未発表

まとめ

- バイジアンネットは効率の良い知識表現方法
- 大脳皮質はおそらくある種のバイジアンネット
- SOM の階層はバイジアンネットになる
→ BESOM モデル
- バイジアンネットモデルは脳の高次機能を実現する要素技術として工学的にも有望
 - 人工視覚野、人工言語野、人工運動野