

人工知能最前線

ディープラーニング(深層学習)って何？

2016年7月17日10:00 ~ 11:10

白井英俊（中京大学工学部電気電子工学科）

話の構成

1. 人工知能
2. 深層学習 (Deep Learning)
何ができる？ 何がすごい？
3. 深層学習の基礎：機械学習
ニューラルネット：発想のもととは脳の仕組み
4. 広がる研究領域
時系列データを扱うLSTM
5. 人工知能は人間の敵か味方か？
人工知能は人間の仕事を奪う？強い味方？
6. 深層学習を学ぶには
数学が必要：線形代数、解析（微分積分）、確率統計

1. 人工知能

- 二つの立場

- (1) 人間の知能と全く同等の知能を人工的に作ることを目指す

- (2) 人間の知能と同等の知能を人工的に作ることによって、人間の知能についての科学的理解を深める

- 「人工的に作る」— 計算機やロボットの技術が進歩したことによって可能になりつつある

1990年代、「第5世代コンピュータ」がブーム

2010年以降、「深層学習（ディープラーニング）」による技術革新により応用が広がる

人工知能(AI)は

トヨタ自動車、ソフトバンク、
Google, Facebook,
Microsoft, Baidu, ...

とてもホットなトピック
産業界も注目ー技術者は引く手あまた

人工知能、トップ棋士破る

グーグル開発、囲碁で対戦 人の脳まねた学習威力

【有料会員限定】

「家庭用ロボ、10～15年で一大分野に」トヨタAI子会社CEO

2016/6/20 13:09

小 中 大 保存 印刷 リプリント 共有

トヨタ自動車の人工知能(AI)研究開発子会社、米トヨタ・リサーチ・インスティテュート(TRI)のギル・プラット最高経営責任者(CEO)が20日、日本経済新聞などのインタビューに応じた。プラット氏は家庭用ロボットがAIの有力な応用分野であると指摘し、「10～15年以内にトヨタや産業界にとって一大分野になる」との見方を示した。

AIで運転支援、共同開発

ホンダとソフトバンク

ホンダとソフトバンクは人工知能(AI)を使った自動車の運転支援システムを共同開発する。走行データのほか、表情や声のトーンから感情や嗜好を分析。行動パターンを推測し、運転手が欲

を代行したりできる新たな仕組みを目指す。対象は通信機能を備えた「コネクテッドカー(つながる車)」で、203

人工知能が扱うテーマ

日々集まる大量のデータから構造（因果関係など）を見つける

適用領域：音声認識、画像認識、自然言語処理、ロボット制御（車の自動運転）、医療、商品検索など

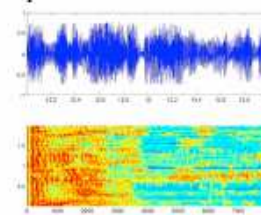
画像や動画



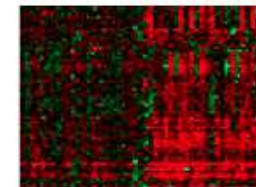
文書や文章



音声・音波



遺伝子情報



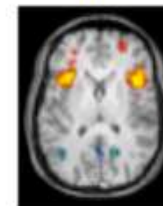
商品お勧め・
感情分析



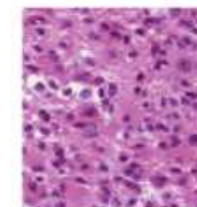
関係データ
SNS



fMRI



癌細胞



ほとんどがラベルなし

話の構成

1. 人工知能
2. 深層学習 (Deep Learning)
何ができる？ 何がすごい？
3. 深層学習の基礎：機械学習
ニューラルネット：発想のもととは脳の仕組み
4. 広がる研究領域
時系列データを扱うLSTM
5. 人工知能は人間の敵か味方か？
人工知能は人間の仕事を奪う？強い味方？
6. 深層学習を学ぶには
数学が必要：線形代数、解析（微分積分）、確率統計

2. 革命的技術：深層学習 (Deep Learning)

日々集まる大量のデータから構造（因果関係など）を見つける

画像や動画



文書や文章



音声・音波



構文解析

深層学習によって、いろいろなレベル
における構造の発見、推論が可能になる

感情分析

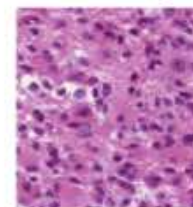
amazon



SNS

facebook

twitter



ほとんどがラベルなし

2. 革命的技術：深層学習 (Deep Learning)

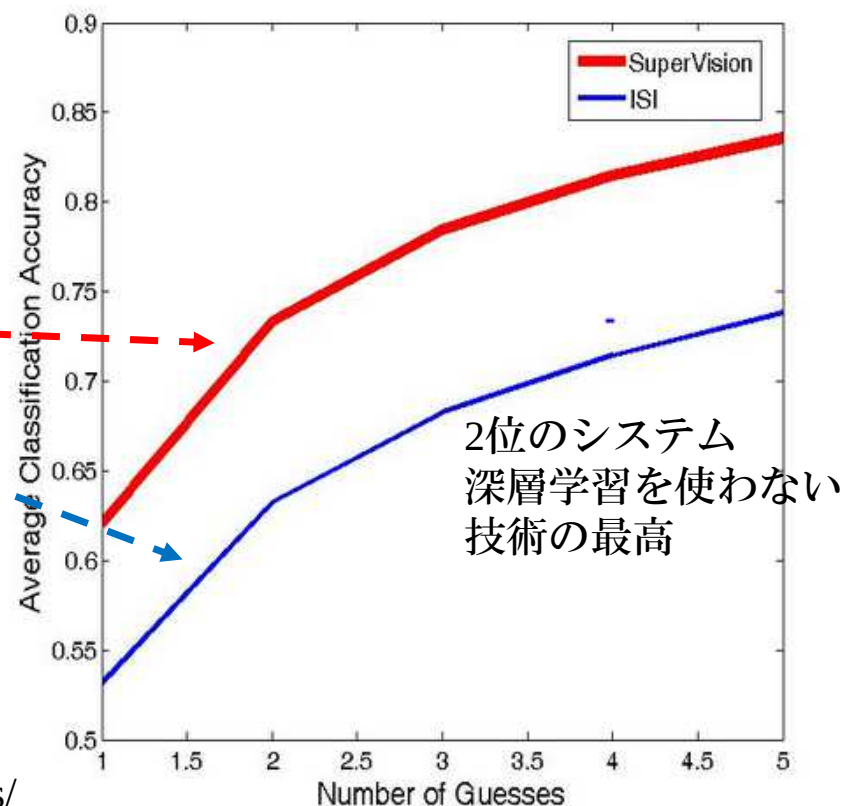
深層学習が知られるようになったきっかけ：2012年9月

一般物体認識のコンテスト ILSVRC
(ImageNet Large Scale Visual
Recognition Challenge)

Krizhevskyらのシステムが大差で1位

誤り率 16% 2位(従来技術)は 26%

2015年では 5 %まで改善



<http://image-net.org/challenges/LSVRC/2012/analysis/>

一般物体認識の例

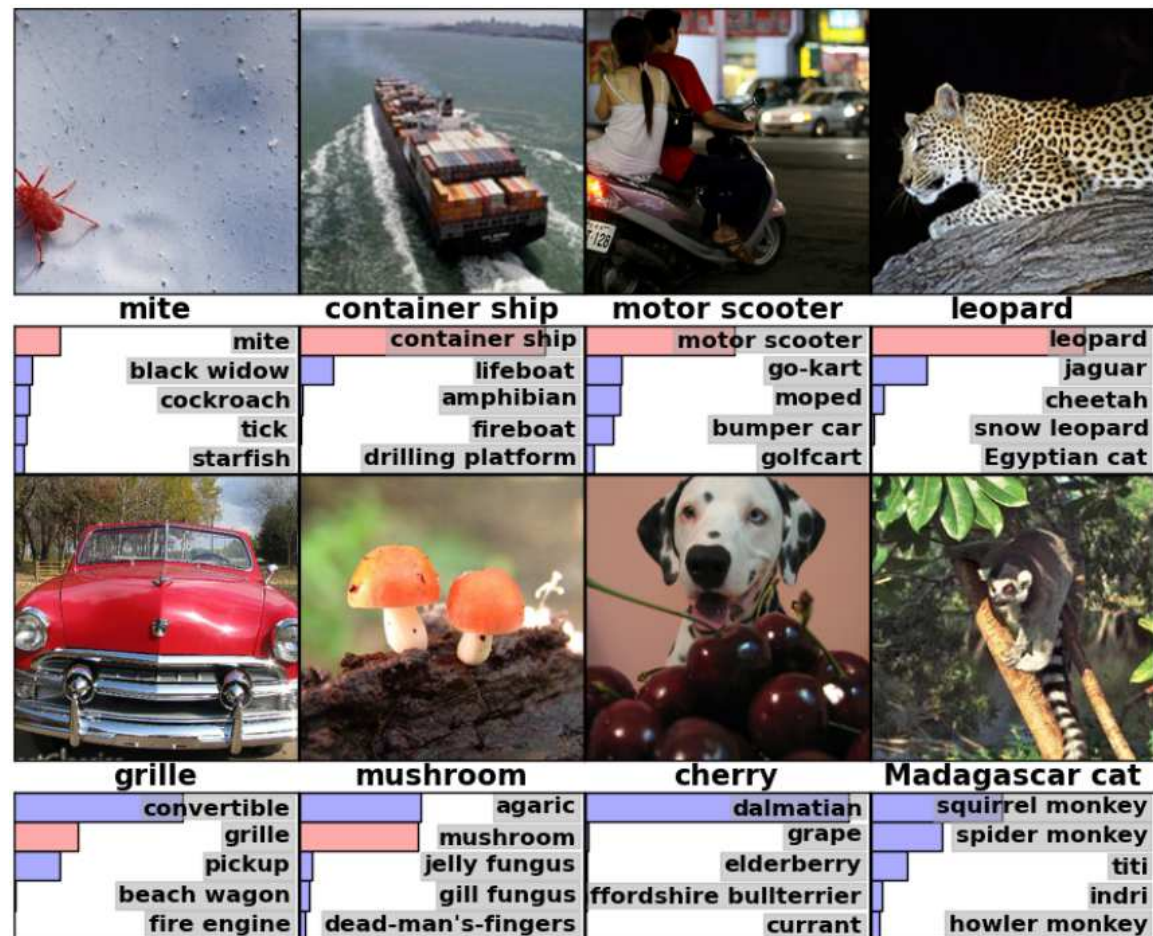
一般物体認識：

画像に写っている
物体を答える

上に画像

下には答えの候補とその確率（可能性）

<http://www.stat.ucla.edu/~junhua.ma/o/m-RNN.html>



正解

不正解

一般物体認識の発展形 (Deep captioning)

画像から説明文生成

(目の不自由な人にとり、目の代わりになる可能性)

窓の前のベンチの上の猫



a cat sitting on a bench in front of a window



a man riding a snowboard down a snow covered slope



a train traveling down a train track next to a forest

画像中の複数の物体を認識
その位置関係から
妥当な文を生成



a man riding a wave on top of a surfboard



a clock tower with a clock on top of it



a person on skis is skiing down a hill

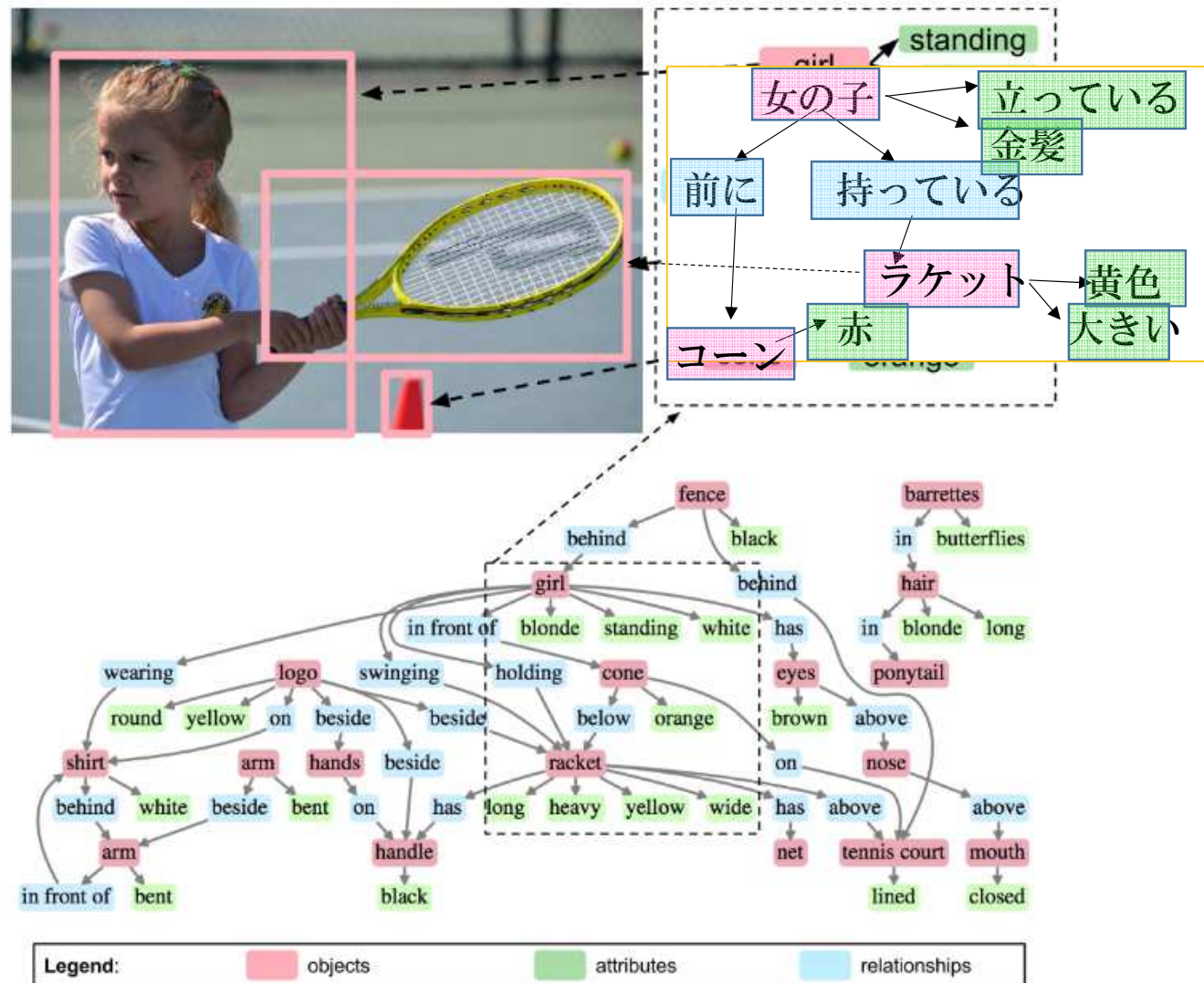
サーフボードで波乗りしている人

<http://www.stat.ucla.edu/~junhua.mao/m-RNN.html>

説明文生成の仕組み

- 画像を物体が写っている領域に分割
- そこから物体を認識
- 物体同士の関係を認識
- 文を組み立て

Image Retrieval using Scene Graphs
 Justin Johnson, Ranjay Krishna, Michael Stark, Li-Jia Li, David Ayman Shamma, Michael Bernstein, Li Fei-Fei
 CVPR 2015



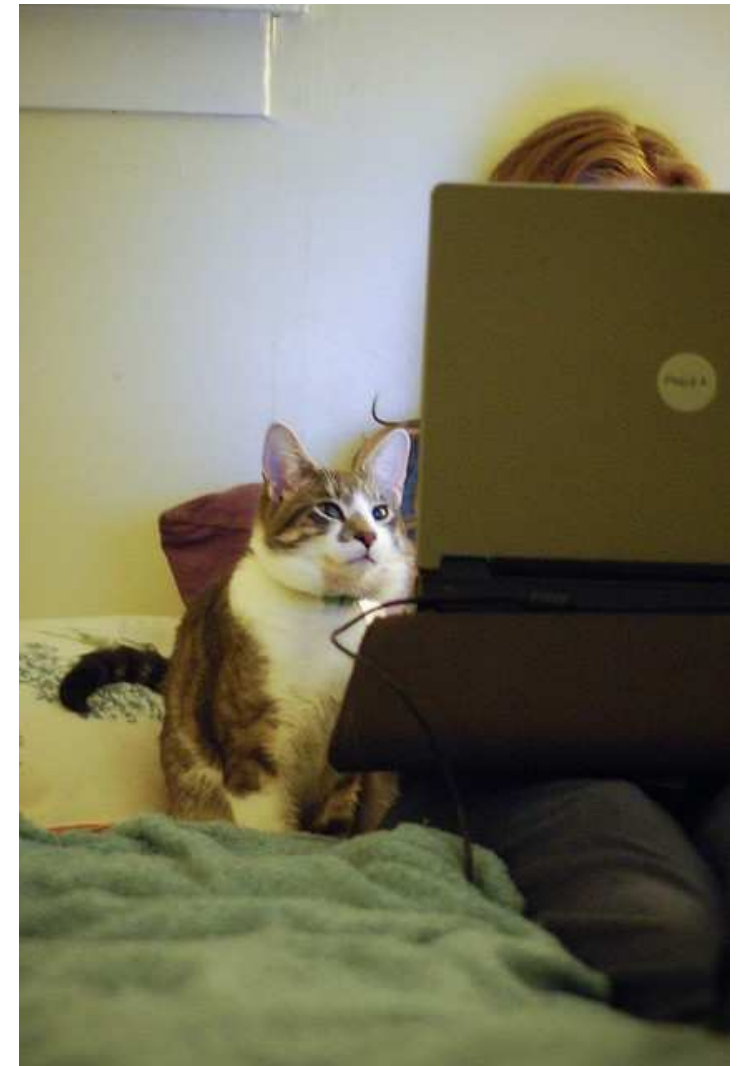
説明文生成の難しさ

候補文1:「猫がベッドの上に座っている」

候補文2:「猫がベッドの上でコンピュータ画面をながめている」

候補文3:「ベッドの上で人が大きなコンピュータ画面の後ろに隠れるように座っており、その横に猫がその画面を座って見ている」

人間の「**注意**(attention)」傾向
情報の有用性
などを考慮して文生成



画像理解：質問応答

これはどこで撮られたシーンか？

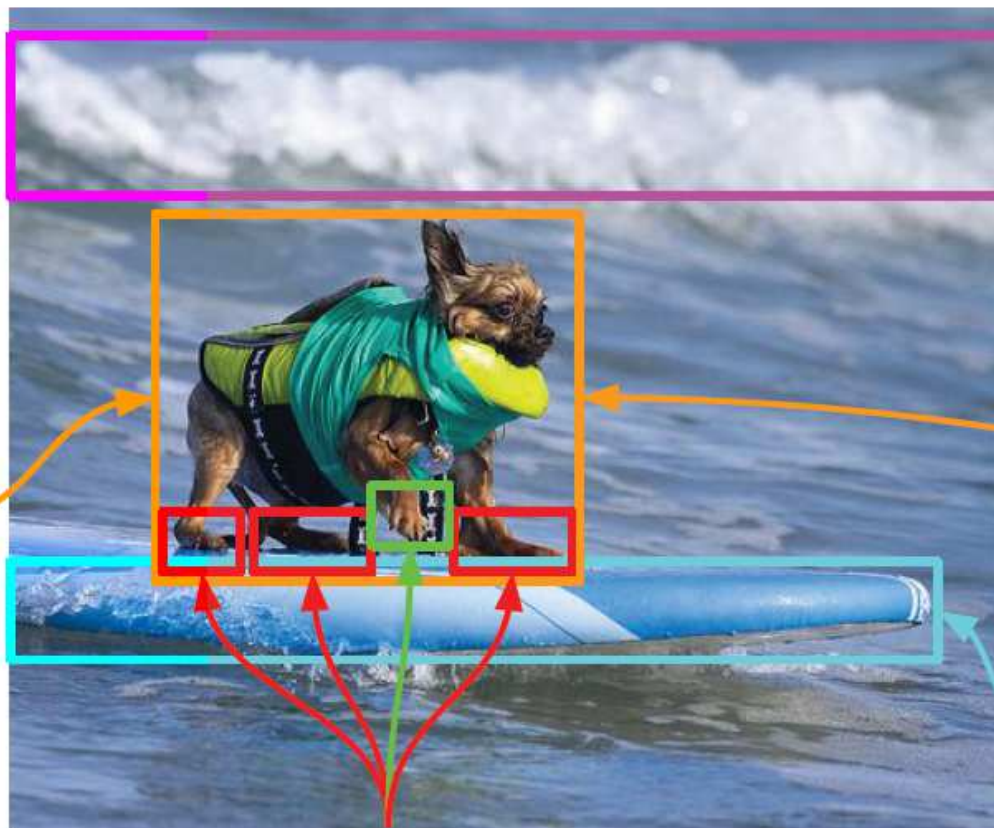
Where does this scene take place?

- A) In the sea. ✓
- B) In the desert.
- C) In the forest.
- D) On a lawn.

この犬は何をしている？

What is the dog doing?

- A) Surfing. ✓
- B) Sleeping.
- C) Running.
- D) Eating.



この泡はなぜ？

Why is there foam?

- A) Because of a wave. ✓
- B) Because of a boat.
- C) Because of a fire.
- D) Because of a leak.

この犬はどこに立っている

What is the dog standing on?

- A) On a surfboard. ✓
- B) On a table.
- C) On a garage.
- D) On a ball.

Which paw is lifted?

音声認識

IBIS2013[音声認識
分野における深層学習
技術の研究動向]

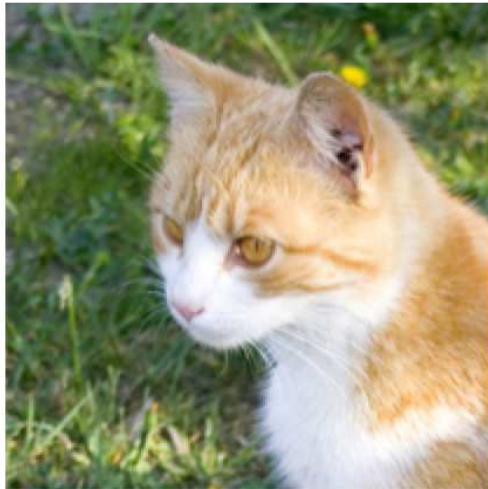


使用データ	研究機関	データ量 (時間)	従来法 エラー率	DNN エラー率	エラー 削減率
CSJ 日本語講義	京大	250h	20.0	17.5	12.5
MIT-OCW 英語講義	NTT	104h	28.2	22.5	20.2
SWBI 英語電話	Microsoft	309h	27.4	18.5	32.5
YouTube 動画音声	Google	1400h	52.3	47.6	9.0
Broadcast 動画音声	IBM	50h	18.8	17.5	6.9

Deep Dream: 芸術への応用

<http://deepdreamgenerator.com/>

<https://research.preferred.jp/2015/>



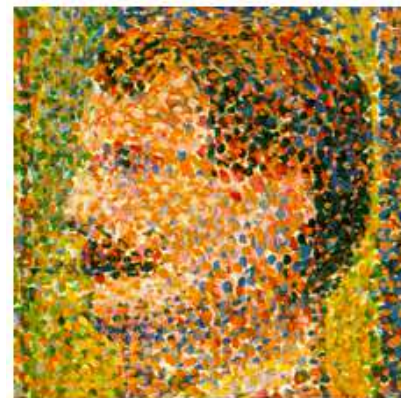
コンテンツ画像



スタイル画像



生成された画像



Deep Q-Network:ゲームへの応用

DeepMind（Googleの子会社）が開発

深層学習と強化学習を組み合わせたアルゴリズムにより動作し、ゲームのルールを教えていない場合でも、どのように操作すれば高得点を目指すことができるのかを判断することができる

DQNはAtari 2600の49種類のゲーム中43種類で従来の人工知能による得点を上回り、29のゲームではプロゲーマーと同等またはそれ以上のパフォーマンスを見せた

(WikiPediaより)



分散深層強化学習によるロボット制御

<https://research.preferred.jp/2015/06/distributed-deep-reinforcement-learning/>

<https://research.preferred.jp/2016/01/ces2016/>



話の流れ

1. 人工知能

2. 深層学習 (Deep Learning) 何ができる？ 何がすごい？

3. 深層学習の基礎：機械学習

ニューラルネット：発想のものは脳の仕組み

4. 広がる研究領域

時系列データを扱うLSTM

5. 人工知能は人間の敵か味方か？

人工知能は人間の仕事を奪う？強い味方？

6. 深層学習を学ぶには

数学が必要：線形代数、解析（微分積分）、確率統計

適用可能な分野の広さ

人間がルールや特徴抽出をしなくとも勝手に学習する

しかも精度が高い（一般物体認識では人間よりも高い）

分野横断的に使える・組み合わせ可能

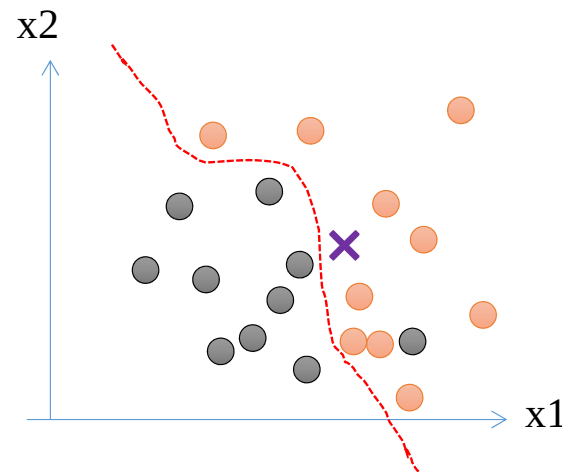
3. 深層学習の基礎：機械学習

機械学習(Machine Learning)

ロボットやコンピュータなどが、経験（学習）をもとにして、行動や予測するためのパラメータを自動的に獲得する手法

その典型例：分類問題

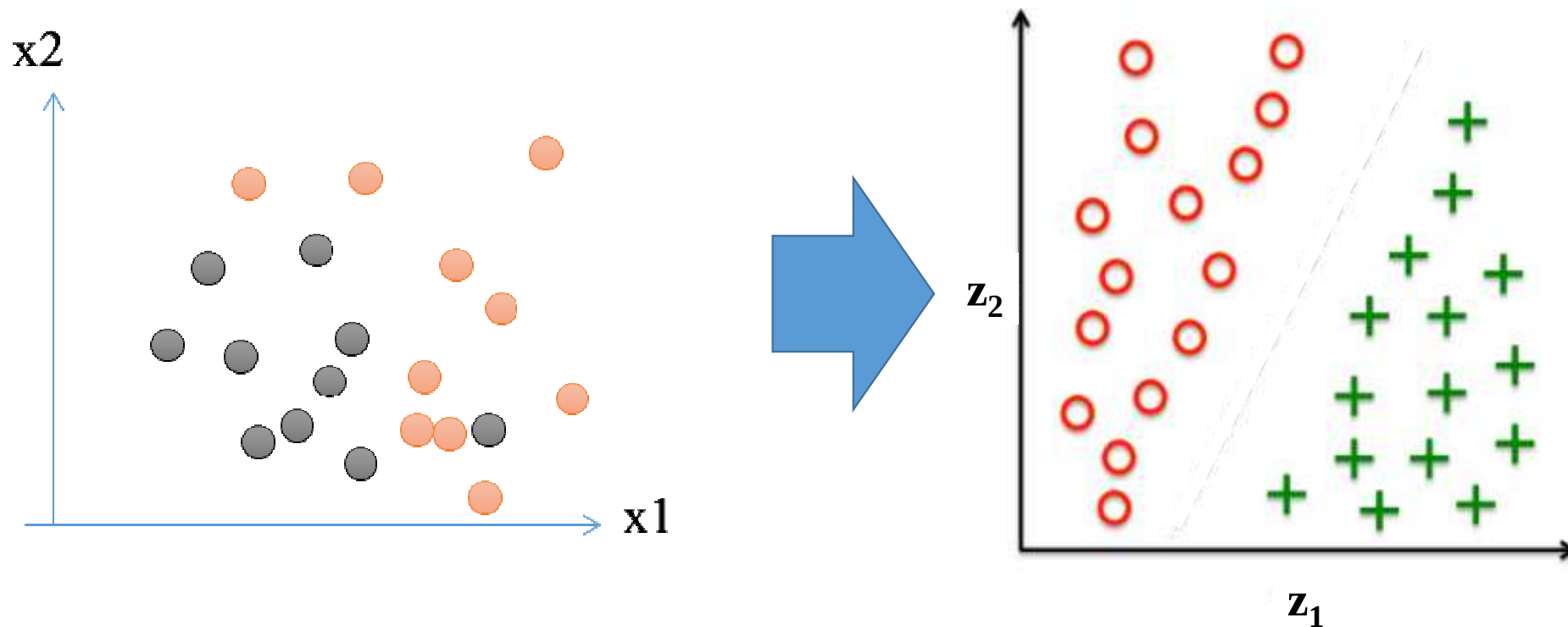
●と●が右図のように分布しているときに、これらを「分ける」方法を求めたい、
もしくは
新たなデータが入ったときに
どちらのグループに属するか
を予測する



特徴量

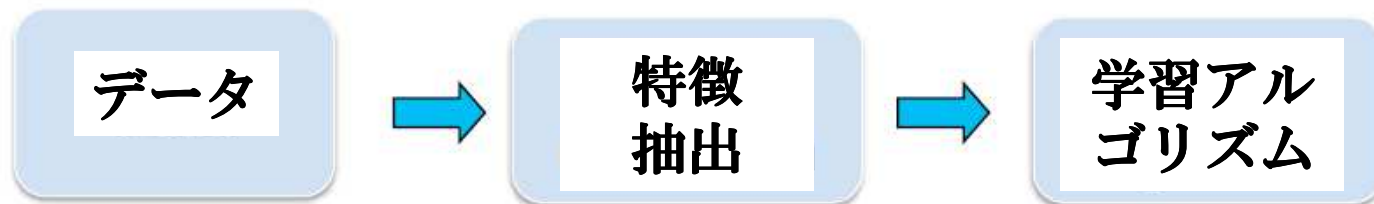
深層学習以前は、
人間が苦勞し、そのような
特徴量を考え出していた

特徴量を工夫する ($x_1, x_2 \Rightarrow z_1, z_2$) ことにより、
分類しにくいデータでも分類しやすくなることもある



深層学習以前の機械学習

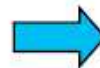
Russ Salakhutdinov (2015) Deep Learning Summer School



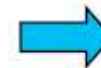
物体
検出



画像



画像特徴

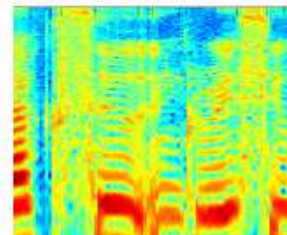
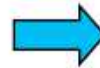


抽出領域

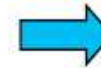
話者
認識



音声



音声特徴

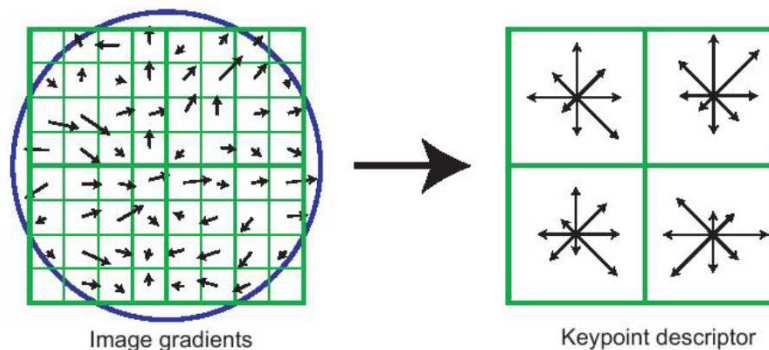


話者特定

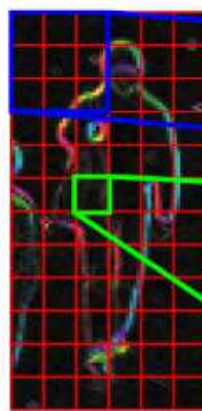
画像特徴の代表例

SIFT (Scale Invariant Feature Transform)

画像の勾配強度と方向算出
勾配方向ヒストグラム作成

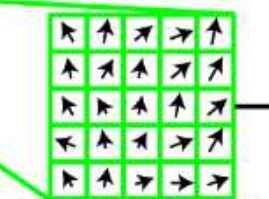


入力画像

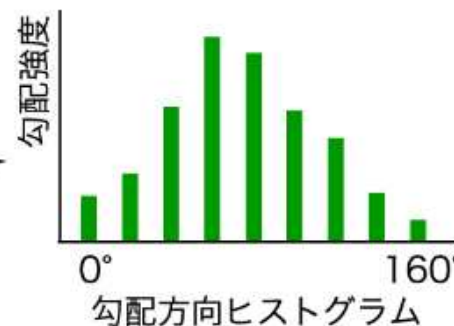


輝度勾配

1ブロック:3x3セル



1セル:5x5ピクセル



HOG (Histogram of Oriented Gradients)

局所領域(セル)の輝度の勾配方向をヒストグラム化

http://www.vision.cs.chubu.ac.jp/joint_hog/pdf/HOG+Boosting_LN.pdf

深層学習による機械学習

適用可能な分野の広さ

人間がルールや特徴抽出をしなくとも勝手に学習する

しかも精度が高い（一般物体認識では人間よりも高い）

分野横断的に使える・組み合わせ可能

物体
検出



画像

話者
認識



音声

深層学習



抽出領域



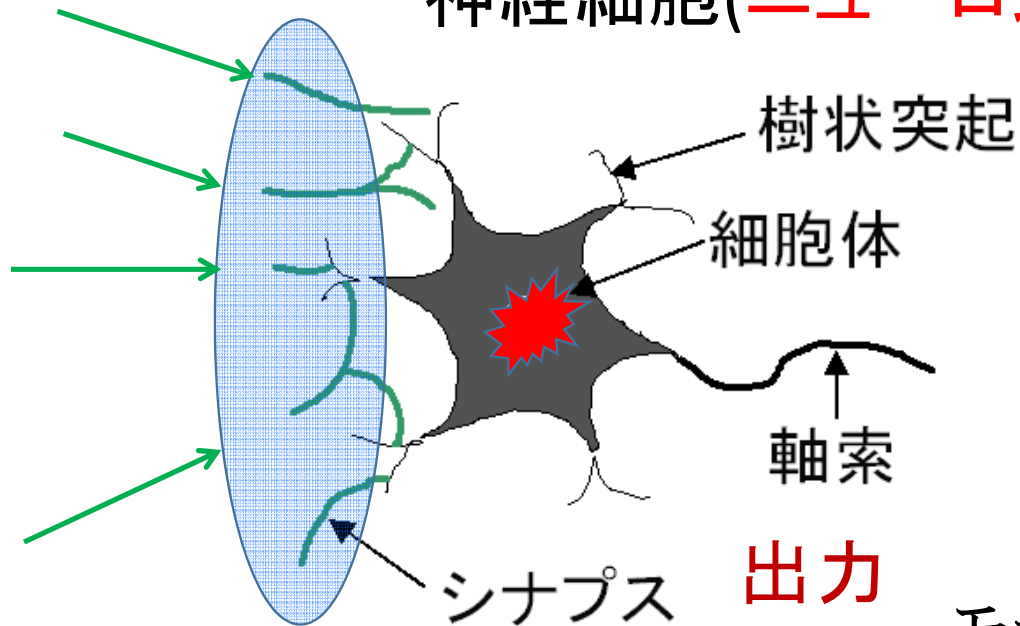
話者特定

話の構成

1. 人工知能
2. 深層学習 (Deep Learning)
何ができる？ 何がすごい？
3. 深層学習の基礎：機械学習
ニューラルネット：発想のもととは脳の仕組み
4. 広がる研究領域
時系列データを扱うLSTM
5. 人工知能は人間の敵か味方か？
人工知能は人間の仕事を奪う？強い味方？
6. 深層学習を学ぶには
数学が必要：線形代数、解析（微分積分）、確率統計

神経回路：脳を構成している回路

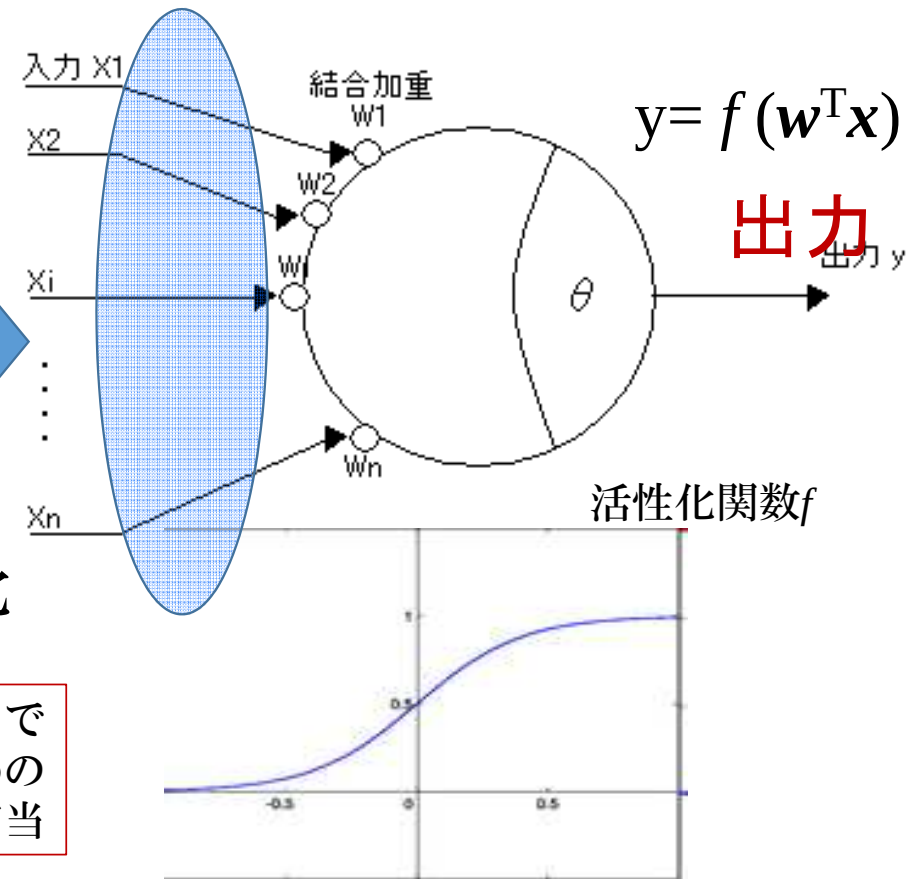
神経細胞(ニューロン)



入力 出力
入力の総和が
閾値を超えたら
発火

モデル化

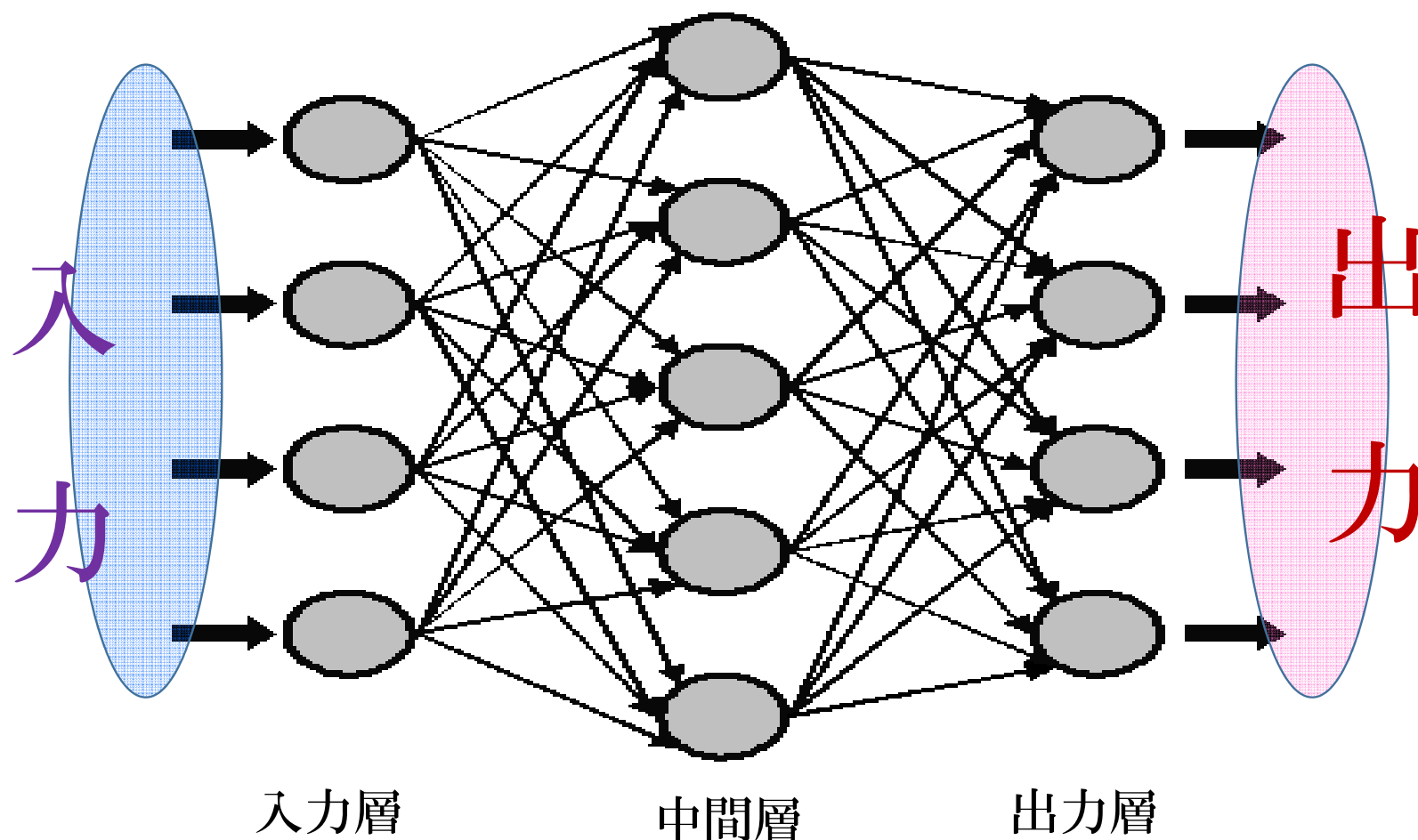
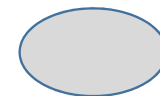
コンピュータで
実現するための
設計図、に相当



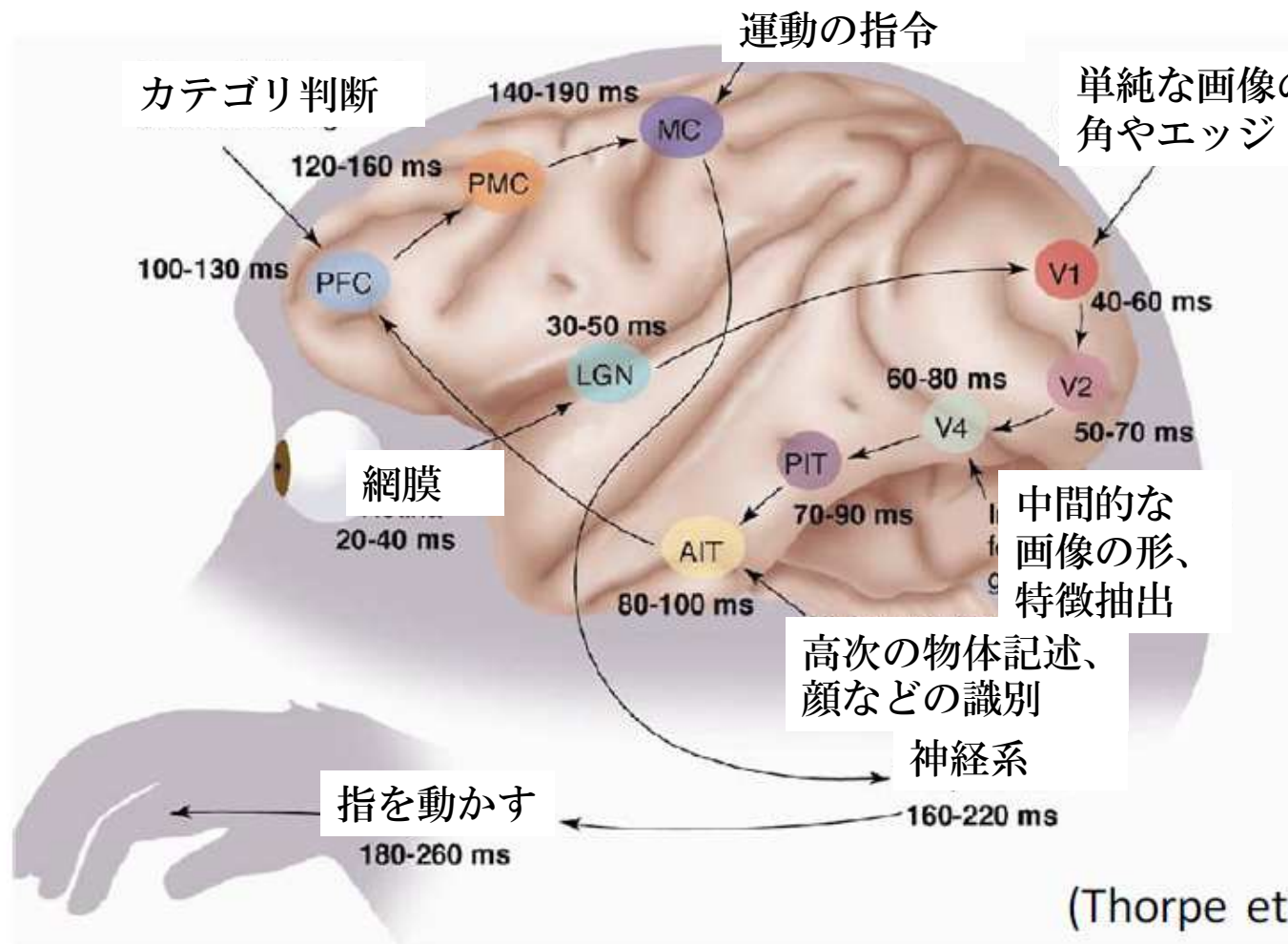
ネット＝ネットワーク、網目状のもの

ニューラルネット：神経回路のモデル化

ニューロン



深層学習：発想は脳の仕組み



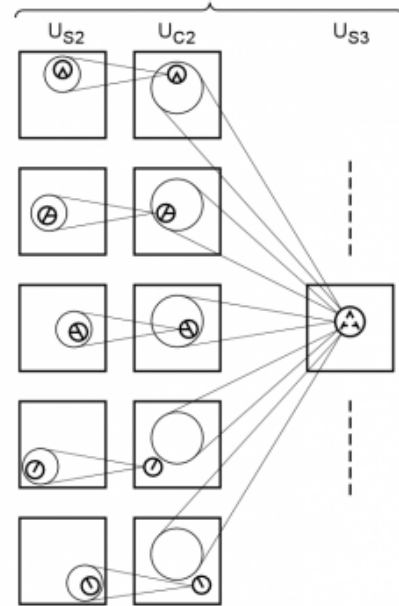
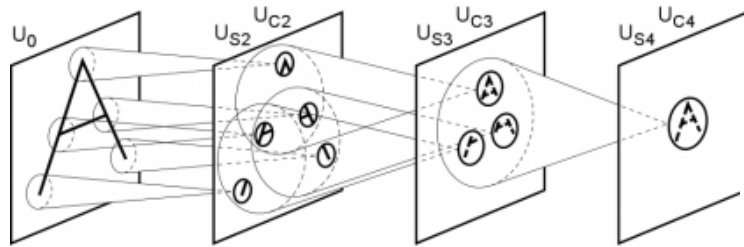
複雑な神経回路網
その中に、何か
規則性（階層構造）
が見られる

(Thorpe et al., 1995-...)

ネオコグニトロン (福島、1979)

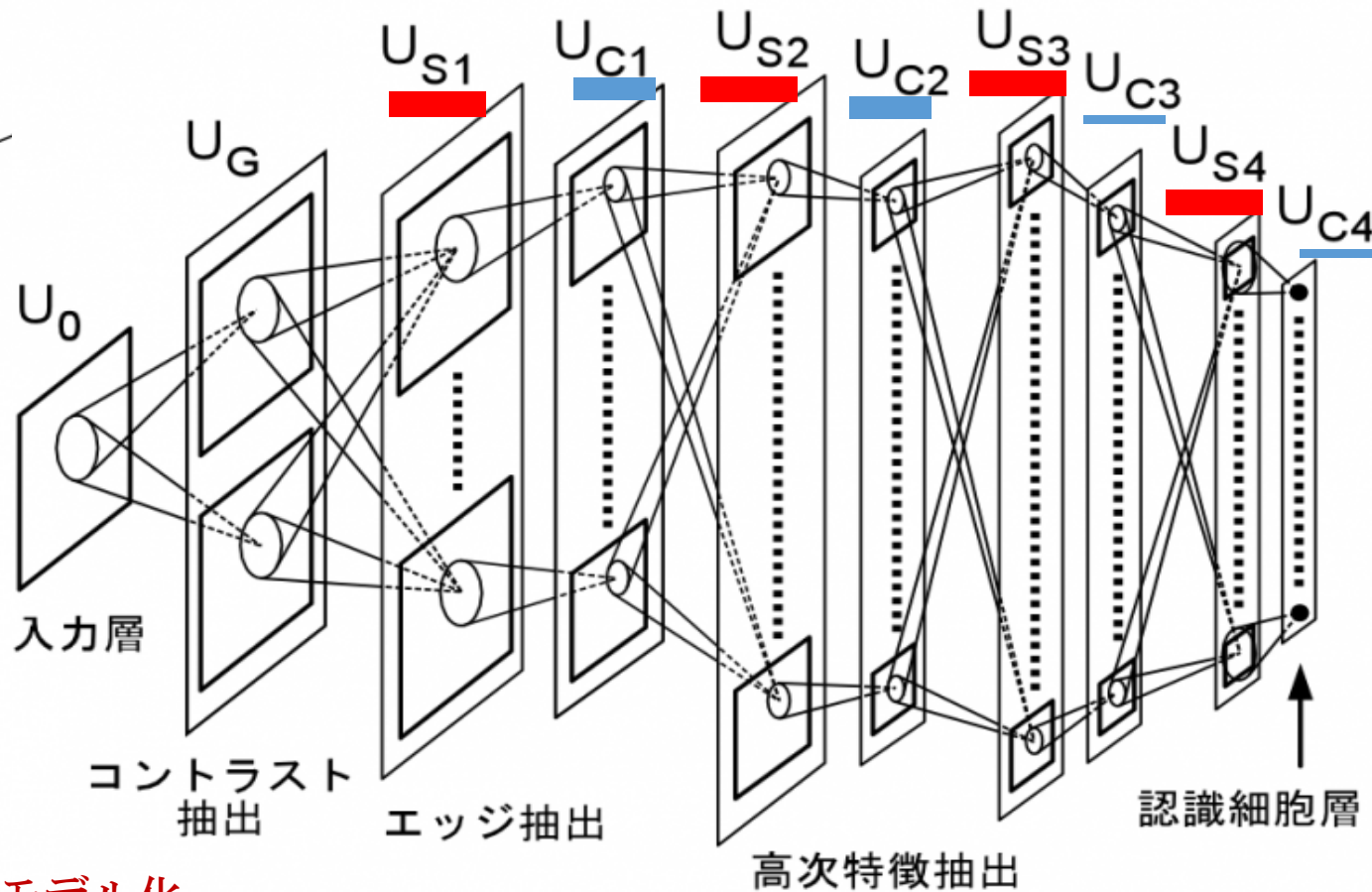
自己組織化機能があり、学習によってパターン認識能力を獲得

S細胞が特徴抽出,
C細胞がぼかし
S細胞とC細胞の組み合わせ



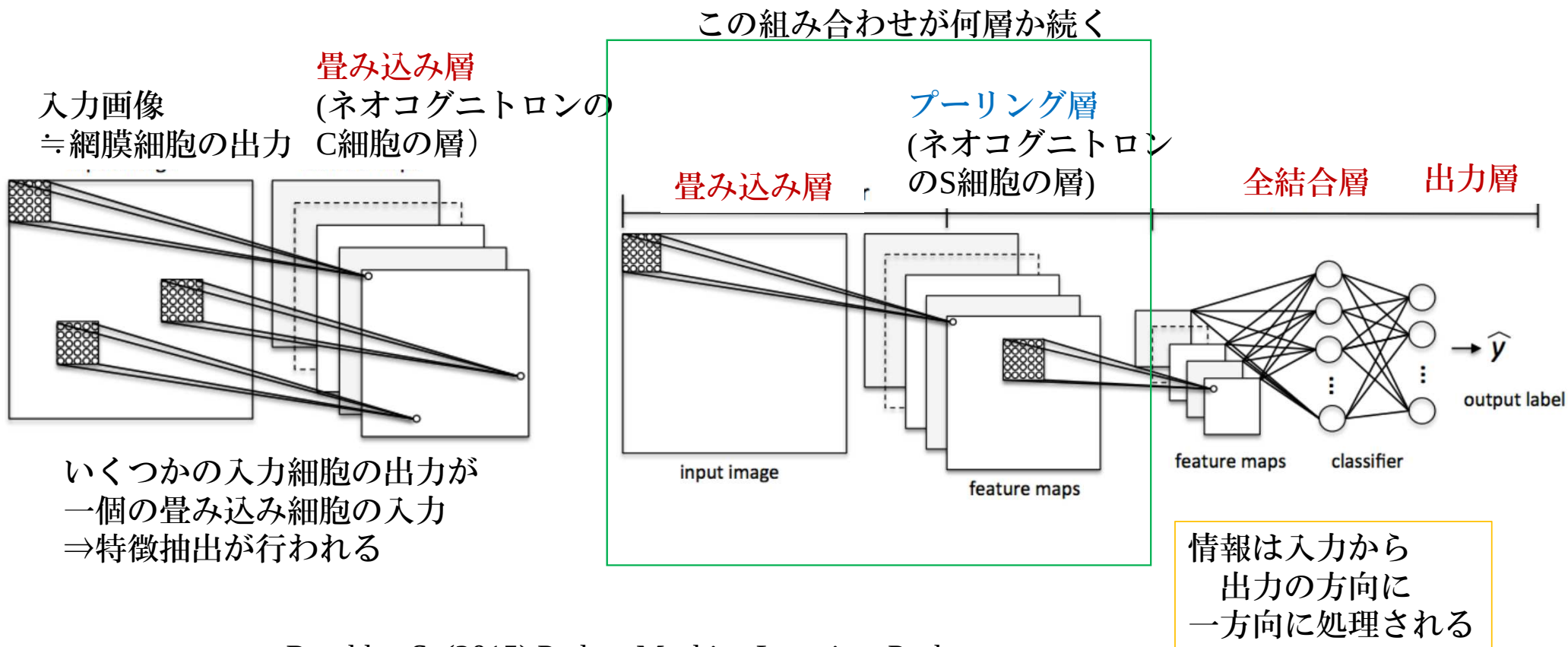
階層的な情報処理

脳の仕組みのモデル化



深層学習の典型モデル：CNN (Convolutional Neural Network)

ネオコグニトロンと
構成がよく似ている

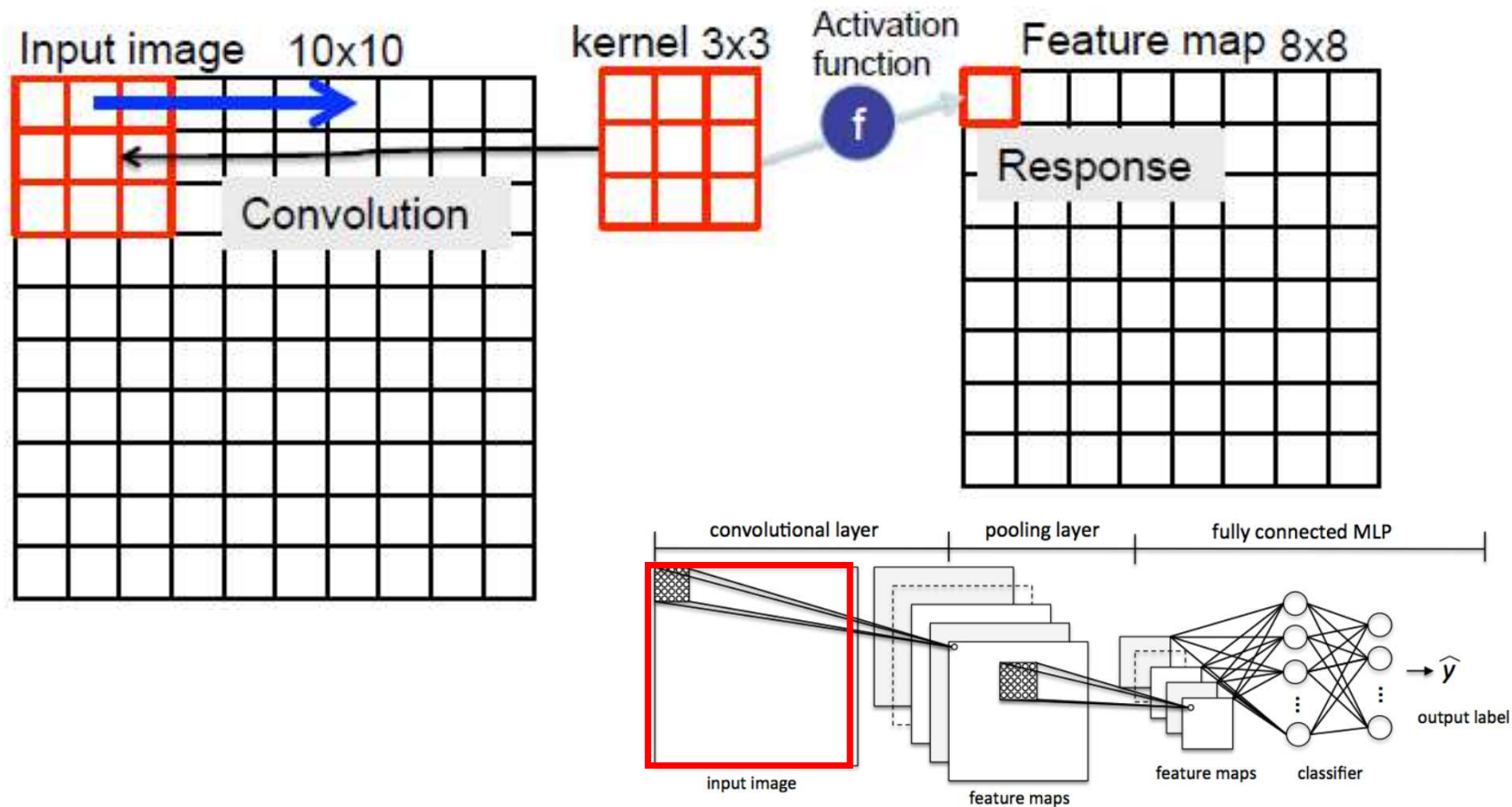


Raschka, S. (2015) Python Machine Learning. Packt

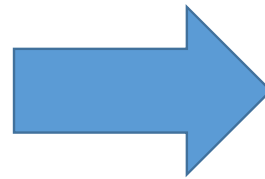
畳み込み層 (ネオコグニトロンのS細胞)

小さな領域にカーネル（フィルタに相当）を適用して畳み込む

⇒ エッジやコーナーなどの特徴を抽出



畳み込みの例(カーネル=フィルタに相当)



グラディエントフィルタ

$[-1, 0, 1],$
 $[-1, 0, 1],$
 $[-1, 0, 1]$

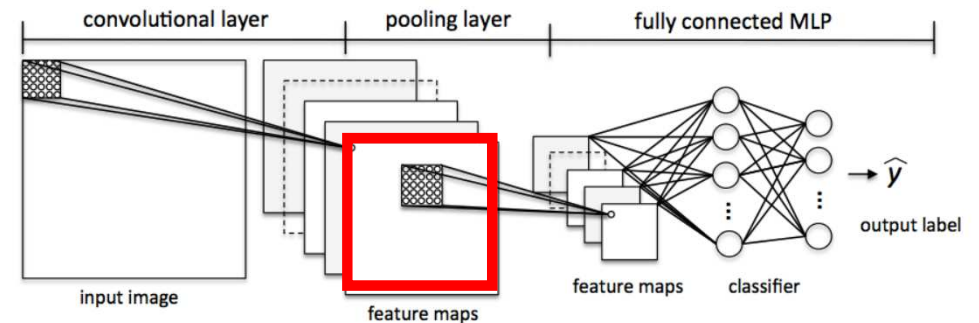
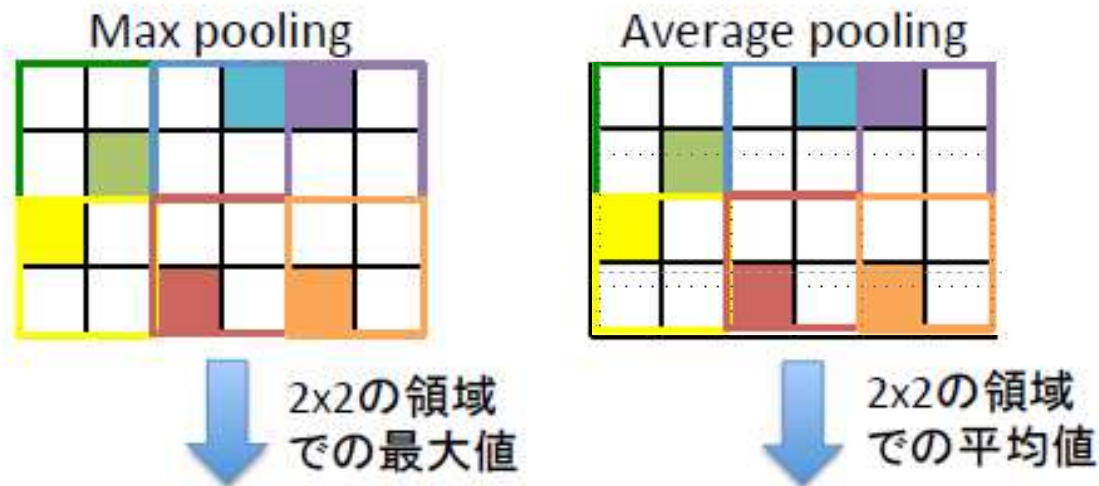
このようなカーネルを複数用意し、
局所領域の特徴抽出を行う
カラーの場合はR,G,Bそれぞれに適用



畳み込み層のカーネルは
学習によって獲得

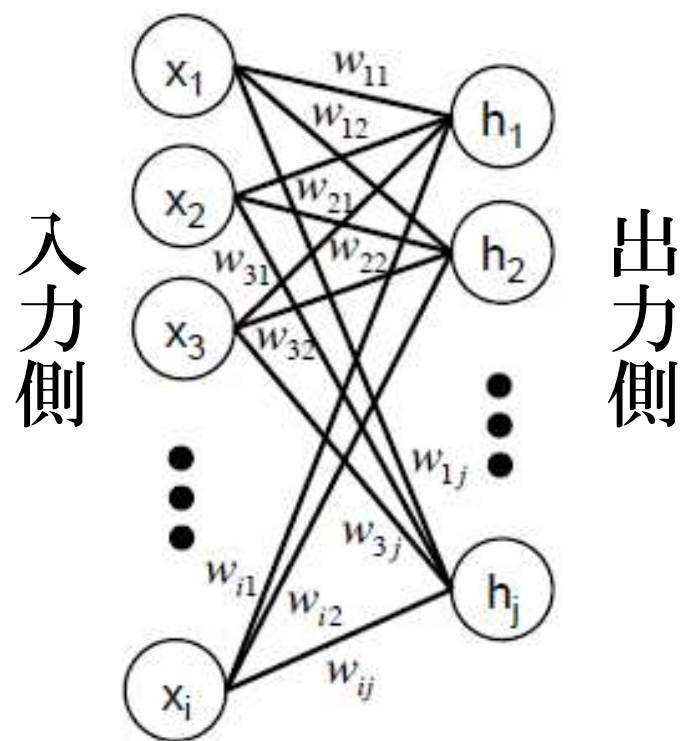
プーリング層 (ネオコグニトロンのS細胞)

畳み込み層の出力を受けサイズを縮小 (大局的な特徴を抽出)



これ以降3枚は山下(2016)「ディープラーニングによる画像認識と応用事例」から図を借用

全結合層

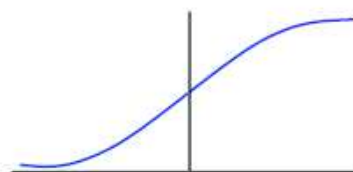


出力

$$h_j = f(\mathbf{w}^T \mathbf{x} + b_j)$$

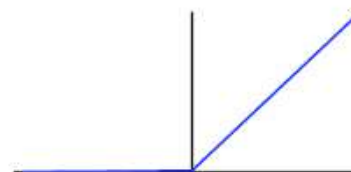
関数 f は活性化関数

シグモイド関数



$$f(x_i) = \frac{1}{1 + e^{-x_j}}$$

Rectified Linear Unit(ReLU)



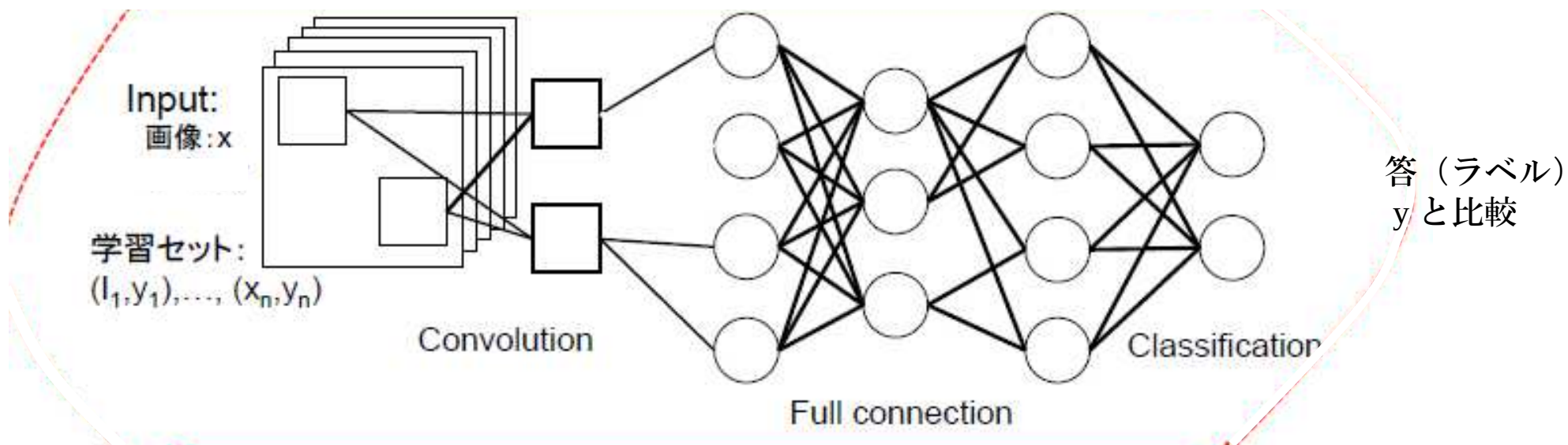
$$f(x_i) = \max(x_j, 0)$$

Maxout



複数のカーネルの畳み込み値の最大を出力

CNNの学習



話の構成

1. 人工知能
2. 深層学習 (Deep Learning)
何ができる？ 何がすごい？
3. 深層学習の基礎：機械学習
ニューラルネット：発想のもととは脳の仕組み
4. 広がる応用領域
時系列データを扱うLSTM
5. 人工知能は人間の敵か味方か？
人工知能は人間の仕事を奪う？強い味方？
6. 深層学習を学ぶには
数学が必要：線形代数、解析（微分積分）、確率統計

広がる応用領域

よくある質問と答：

- 1) ネオコグニトロンが提案され、深層学習のCNNはほとんどそれに近いもののなのに、なぜ今注目されているのか？
- 2) 古くはパーセプトロンや並列分散処理（PDP）というニューラルネットのモデルがあった。それと比べて深層学習はどこが新しい？
- 3) 本質的に深層学習の新しさとは何？

広がる応用領域

よくある質問と答：

1) ネオコグニトロンが提案され、深層学習のCNNはほとんどそれに近いもののなのに、なぜ今注目されているのか？

答：ネオコグニトロンの理論は1979年。

それを計算機で実現するには、計算機パワーもメモリも、ハードディスクも、データもない時代であった。

現代はその時代に比べて1万倍以上の性能が向上

さらにGPUというスーパーコンピュータの性能をもつ機器が100万円程度で購入できる

広がる応用領域

2) 古くはパーセプトロンや並列分散処理（PDP）というニューラルネットのモデルがあった。それと比べて深層学習はどこが新しい？

答：パーセプトロンやPDPモデルと深層学習の違いは、入力層と出力層との間にある中間層の数の多さ。

PDPモデルでは「学習が収束する」実用的なものは、中間層が高々1か2。それ以上だと、学習が収束しないか、過学習になって、実用的でなかった

広がる応用領域

3) 本質的に深層学習の新しさとは何？

答:

- 1) 大量のデータの存在。教師あり学習で学習するには問題と答えのセットが大量になければできない。
- 2) 活性化関数の工夫。学習が継続して行えるようになった
- 3) モデルやデータの公開。誰でもモデルやデータを試せるようになった： caffe, tensorflow, chainer, theano,
- 4) GPUの存在。
- 5) 時系列データを扱うための工夫。RNNやLSTMと呼ばれる
音声や言語など、いろいろな分野に対する応用がやりやすくなった
- 6) ファインチューニング：他のデータで学習したモデルを初期値として使うことで、新たなデータに対しても十分な性能を達成できる場合がある

深層学習の凄さ

- 今までの人工知能の問題
機械学習させるには、役に立つ特徴量が必要
それを今までは人間が設計していた
- 深層学習では
データ（問題）とその望ましい出力（答）のセットから
コンピュータが**特徴量を自動で抽出**
しかも人間が設計したものよりも性能がよい

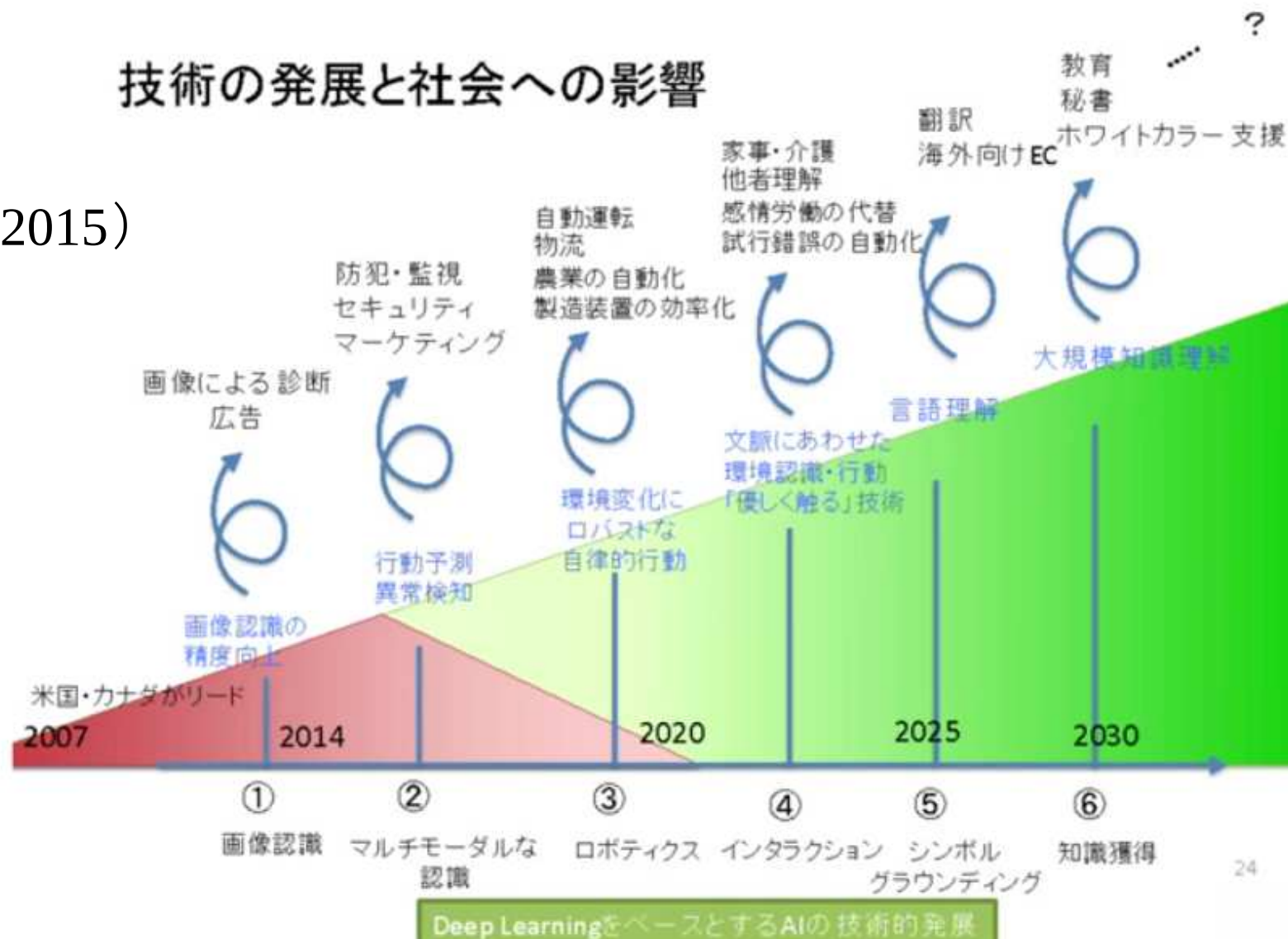
話の構成

1. 人工知能
2. 深層学習 (Deep Learning)
何ができる？ 何がすごい？
3. 深層学習の基礎：機械学習
ニューラルネット：発想のもととは脳の仕組み
4. 広がる応用領域
時系列データを扱うLSTM
5. 人工知能は人間の敵か味方か？
人工知能は人間の仕事を奪う？強い味方？
6. 深層学習を学ぶには
数学が必要：線形代数、解析（微分積分）、確率統計

5

松尾（2015）

技術の発展と社会への影響



考えられるインパクト(松尾2015)

- 機械の動作の飛躍的な向上：製造装置、自動運転、物流
- 犯罪の減少：画像動画による防犯・監視
- 設備保守の自動化：異常検知
- 情報システムのセキュリティ向上：異常検知
- 仮説生成と試行のサイクルの自動化：デザイン、製薬など
- シミュレーション技術の向上：現実の製品化への時間短縮
- 情報システムの連携：人間を介さないシステムの実現

深層学習でできること、できていないこと(現時点)

- 大量の問題と解が与えられれば、その種の問題に性能よく答える
⇒ 本当に「新しい種類の問題（データが十分出回っていない種類の問題）」には対応不十分
答えや評価が与えられないような問題を考えるのも不向き。
- 答えは出せるが、どのようにそれが導かれたかの説明はできない
囲碁の試合でAlphaGoが負けたケース。突然（人間にとって）変な手を打ったが、誰もその原因は説明できなかった
- ある種の論理思考(人間の思考が、視覚や聴覚のような周辺処理系と、思考のような中央処理系に分離されるという学説による)
深層学習は「経験からの学習」で周辺処理系。
人間の思考や論理は「中央処理系」によるもの（だから、記号推論は難しいはず）

論理的思考は可能か？

- 深層学習は、いわば「即物的な思考」

それに対し、幾つかの事実から論理的な帰結を導く、というのは難しいのではないか

例えば：

「太郎は若い」「若い人は夜更かしである」「夜更かしする人は寝坊である」「寝坊すると遅刻する」

から、

「太郎は朝の講義に遅刻しがちである」を導けるか？

人間のことを考えるのはやはり人間

- 計算機は、与えられた問題に対する答えを導こうとする
- ただ、その倫理的な意味、常識的な意味などは考慮されない（そのように学習付けされていない限り）
- SFでは社会的な判断（例えば裁判の判決）を計算機に委ねることの弊害について書かれたものがある

最終的に計算機の判断をうのみにせず、それを材料として判断するのは人間でなければならない（計算機には「責任」をとれない）

参考：自動運転で(死亡)事故があった → 運転手の責任が問われた

話の構成

1. 人工知能
2. 深層学習 (Deep Learning)
何ができる？ 何がすごい？
3. 深層学習の基礎：機械学習
ニューラルネット：発想のもととは脳の仕組み
4. 広がる応用領域
時系列データを扱うLSTM
5. 人工知能は人間の敵か味方か？
人工知能は人間の仕事を奪う？強い味方？
6. 深層学習を学ぶには
数学が必要：線形代数、解析（微分積分）、確率統計

高校で学んでいる数学が基礎

- 機械学習では
データはベクトルとしてあらわされる
データの類似度は ベクトルの内積で求められる
- 深層学習では
情報は入力から出力の方向に処理されるが、
「答えが間違っていたとき」、その情報が出力から入力に逆
向きに流され、ノード間の結合が修正される。そこでは微分
計算が必要
- そもそも現実世界は不確定なもの
それを扱うには、確率統計の理論が必要

この講義からのメッセージ

- 深層学習を代表として、世の中には役に立つ、かつ面白い研究テーマがあふれている
- そのような研究テーマをみつけ、そのような研究をする（それにより世の中に貢献する）ために大学に来てほしい
- それには、（論文を読んだり書いたりするための）国語力と英語力（ほとんどの文献は英語）、
そして、技術的な内容を理解、応用するには数学の力が必要
- 今、高校で学んでいることがいつか役に立つことでしょう

勉強を続けてください。そしてまた中京大学でお会いしましょう