

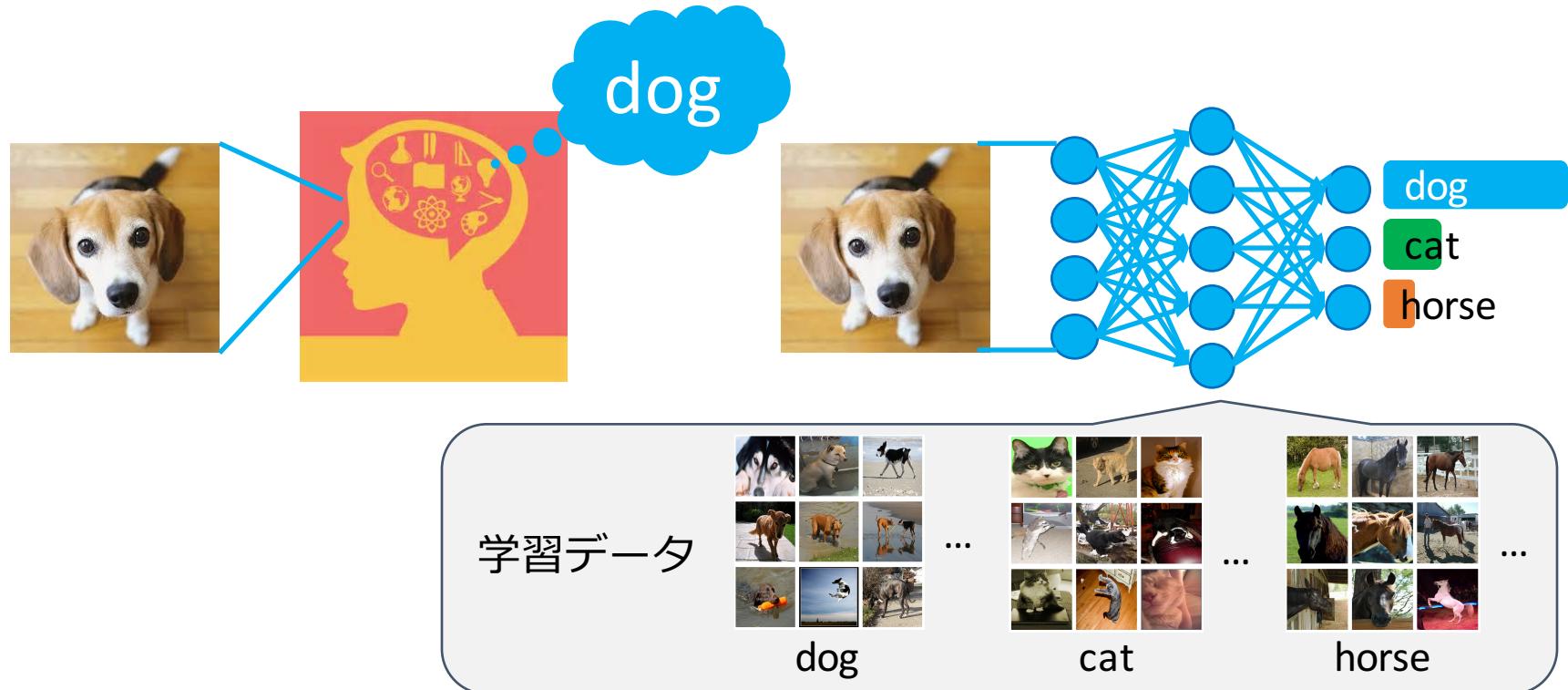
深層学習とその周辺

2016/7/7

ビジネス科学研究科
経営システム科学専攻

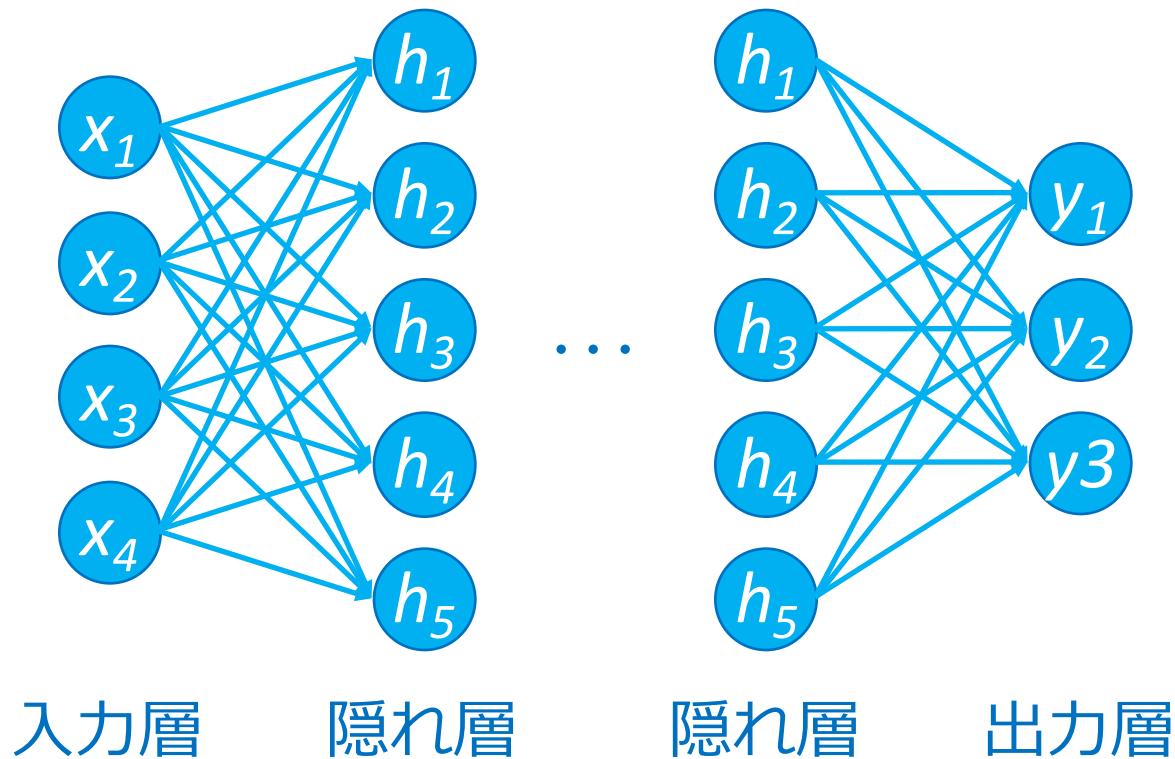
深層学習(Deep Learning)とは

- ニューラルネットワークを用いた機械学習手法
 - 脳の神経細胞(ニューロン)の働きを模した
 - 機械学習とは、データを学習し、パラメータを獲得



ディープニューラルネットワーク

- 隠れ層が2段以上のニューラルネットワークをディープニューラルネットワークという



ニューラルネットワークの歴史

- 黎明～終焉を繰り返し,近年は3度目のブーム

第1期	1940～	• McCullochとPittsが形式ニューロンモデルを発表 [McCulloch-Pitts,43]
	1950～	• Rosenblattがパーセプトロンを発表 [Rosenblatt,57]
	1960～	• MinskyとPapertが単純パーセプトロンの線形分離不可能問題への限界を指摘 [Minsky-Papert,69]
冬	1970～	冬の時代 (階層的構造の学習方法が未解決)
第2期	1980～	• Fukushimaらがネオコグニトロンを提案 [Fukushima,80] • Hopfieldがホップフィールドモデルを提案 [Hopfield,1982] • Rumelhartらが誤差逆伝播法を提案 [Rumelhart+,86] • LeCunらが畳み込みニューラルネットワークを提案 [LeCun,89]
	1990～	冬の時代 (学習時間や過学習に課題, 一方でSVMが流行)
	2000～	• Hintonらが事前学習とオートエンコーダを導入した多層NNを提案 [Hinton+,06]
	2010～	• Seideらが音声認識のベンチマークで圧勝 [Seide+,11] • KrizhevskyらがReLUを提案し画像認識コンペで圧勝 [Krizhevsky,12]

音声認識での圧勝 (2011)

- DNN-HMMを使った手法が、単語誤認識率でが従来法(GMM)に比べ、10%前後の大幅な精度改善
 - 電話での会話音声の標準データセット (Switchedboard)
 - 入力(MFCC)-出力(HMMのステート)関係をDNNで学習

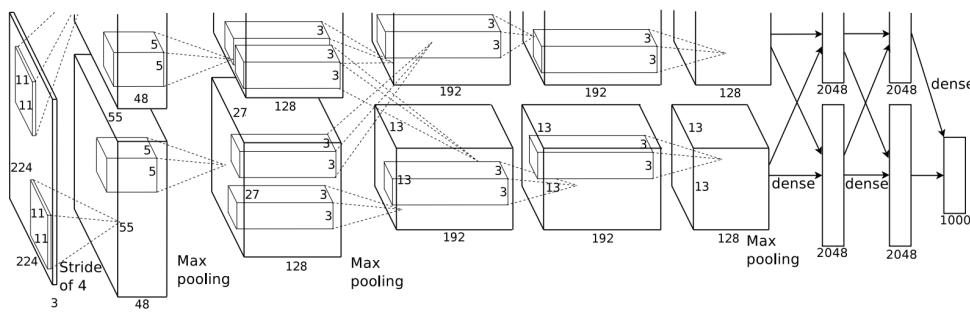
acoustic model & training	recognition mode	RT03S		Hub5'00 SWB	voicemails		tele- conf
		FSH	SW		MS	LDC	
GMM 40-mix, ML, SWB 309h	single-pass SI	30.2	40.9	26.5	45.0	33.5	35.2
GMM 40-mix, BMMI, SWB 309h	single-pass SI	27.4	37.6	23.6	42.4	30.8	33.9
CD-DNN 7 layers x 2048, SWB 309h, this paper (rel. change GMM BMMI → CD-DNN)	single-pass SI	18.5 (-33%)	27.5 (-27%)	16.1 (-32%)	32.9 (-22%)	22.9 (-26%)	24.4 (-28%)

Seide, Frank, Gang Li, and Dong Yu. "Conversational Speech Transcription Using Context-Dependent Deep Neural Networks." *Interspeech*. 2011.

画像認識での圧勝 (2012)

・一般物体認識 (Hintonのグループ)

- ImageNet Large-scale Visual Recognition Challenge 2012
 - 1000カテゴリ, 1カテゴリ当たり 約1000枚の訓練画像
 - 置込み層5, 全結合層3, 2つのGPUで別々に学習 (AlexNet)
 - 誤識別率が10%以上減少 (過去数年間での向上は1~2%) ※



順位	チーム	誤認識率
1	SuperVision	0.15315
2	ISI	0.26172
3	OXFORD_VGG	0.26979
4	XRCE/INRIA	0.27058
5	University of Amsterdam	0.29576
6	LEAR-XRCE	0.34464

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems*. 2012.

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

※ → 0.064 (Google, 2014)

ImageNet Large-scale Visual Recognition Challenge (IRSVRC)

- ImageNet のデータの一部を用いたフラッグシップコンペティション (2010年より開催)
 - ImageNet [Deng et al., 2009]
 - クラウドソーシングにより構築中の大規模画像データセット
 - 1400万枚, 2万2千カテゴリ (WordNet に従って構築)

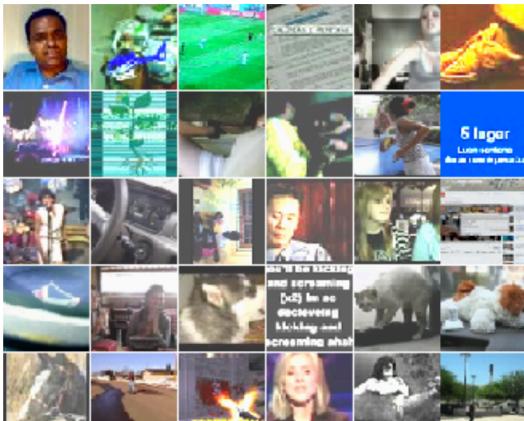


- コンペでのタスク
 - 1000クラスの物体カテゴリ分類
 - 学習データ120万枚, 検証用データ5万枚, テストデータ10万枚
 - 200クラスの物体検出
 - 学習データ45万枚, 検証用データ2万枚, テストデータ4万枚

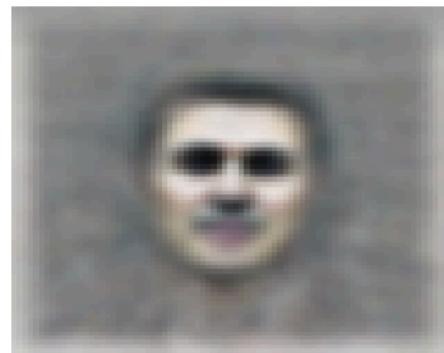
中山英樹. "招待講演深層畳み込みニューラルネットによる画像特徴抽出と転移学習 (音声)." 電子情報通信学会技術研究報告 = IEICE technical report: 信学技報115.146 (2015): 55-59.

画像の教師なし学習 (2012)

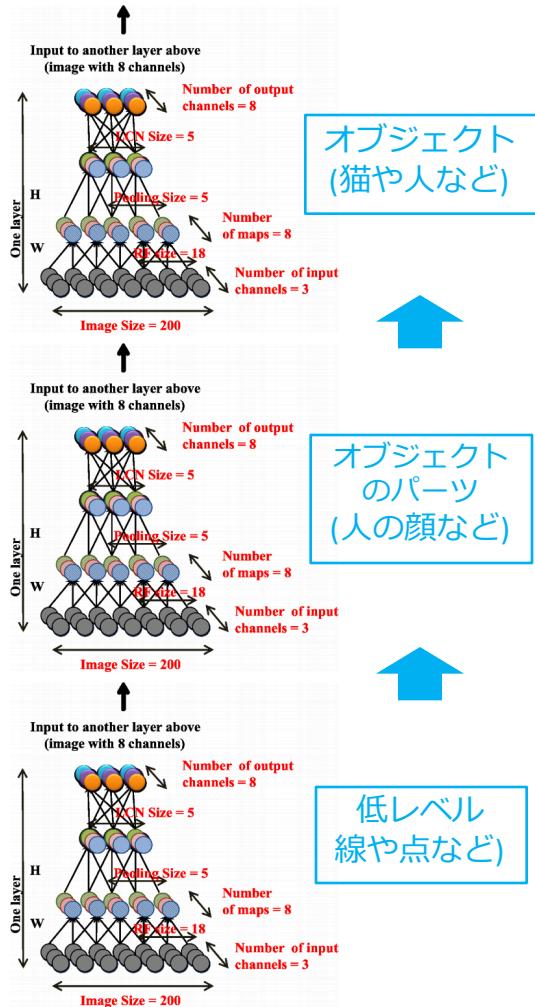
- 4層x3段=12層の autoencoder
 - パラメータ数10億個 (<人)
 - 16コアのPC 1000台 ×3日間
 - YouTubeの画像1000万枚 (顔画像は3%)



学習データ



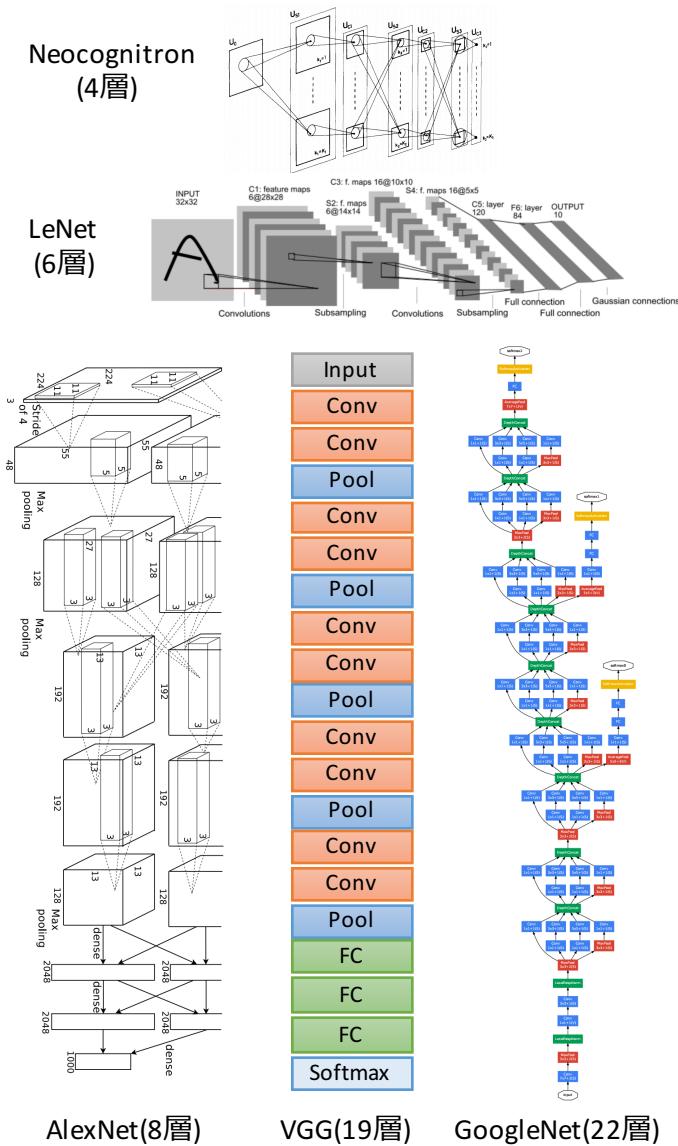
顔ニューロンに対して
最適化された刺激



Le, Quoc V. "Building high-level features using large scale unsupervised learning." In *ICML*, 2012.

ネットワークの変遷(画像系)

畳み込み ニューラル ネットワー クの発見	Neocognitron (1980) [1]	<ul style="list-style-type: none"> 生物の覚神経路を模倣 特徴抽出と情報集約の処理を繰り返す
	LeNet (1998) [2]	<ul style="list-style-type: none"> 畳み込みとプーリングの導入 誤差逆伝搬法によって学習した
プーリング, 活性化関数, 正則化	AlexNet (2012) [6]	<ul style="list-style-type: none"> Ave./Max Pooling, Normalization ,ReLU[4], Dropout[5]の導入 8層
	Network in Network (2013) [7]	<ul style="list-style-type: none"> 畳み込み層に非線型性を導入
	VGG Net (2014) [8]	<ul style="list-style-type: none"> 小さい畳み込みサイズ(3x3)を多段にした 19層
	GoogLeNet / Inception (2014-2015) [9,10]	<ul style="list-style-type: none"> 複数の畳み込みフィルタを並列に用いる 22層
アーキテク チャの多様 化	SPP Net (2014) [11]	<ul style="list-style-type: none"> 様々なサイズの画像を許容 CNN 時のリサイズを回避
	All Convolutional Net, guided BP (2014) [12]	<ul style="list-style-type: none"> プーリングをストライド2の畳み込みに置き換え guided BPによる超高次層の特徴可視化
学習法の多 様化	Exemplar CNN (2014) [13]	<ul style="list-style-type: none"> Data Augmentation を利用して教師なし表現学習
	Triplet Network (2014) [14]	<ul style="list-style-type: none"> ユーフリッド空間上でCNN上の特徴同士が同じクラスなら近くなるように,別クラスなら遠くなるようにした
超多層ア ーキテクチャ	Batch Normalization (2015) [15]	<ul style="list-style-type: none"> パラメータ付き正規化処理 複雑なアーキテクチャをスクラッチで学習させる必須技術
	Residual Network; ResNet (2015) [16]	<ul style="list-style-type: none"> 152層からなる超多層 途中の特徴マップを何層か前にバイパス



- [1] K. Fukushima. Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics* 36, 1980.
- [2] Y LeCun, L Bottou, Y Bengio, P Haffner. Gradient based learning applied to document recognition. *Proceedings of the IEEE* 86, 1998.
- [3] K. Jarrett, K. Kavukcuoglu, M. Ranzato, Y. LeCun. What is the best multi stage architecture for object recognition?. *CVPR*, 2009.
- [4] X. Glorot, A. Bordes, Y. Bengio. Deep Sparse Rectifier Neural Networks. *AISTATS* 11, 2011.
- [5] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov. Improving neural networks by preventing co adaptation of feature detectors. *arXiv: 1207.0580*, 2012.
- [6] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. *NIPS*, 2012.
- [7] M. Lin, Q. Chen, S. Yan. Network In Network. *arXiv: 1312.4400*, 2013.
- [8] K. Simonyan, A. Zisserman. Very Deep Convolutional Networks for Large Scale Visual Recognition. *arXiv: 1409.1556*, 2014.
- [9] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich. Going deeper with convolutions. *arXiv: 1409.4842*, 2014.
- [10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna. Rethinking the Inception Architecture for Computer Vision. *arXiv: 1512.00567*, 2015.
- [11] K. He, X. Zhang, S. Ren, J. Sun. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *arXiv: 1406.4729*, 2014.
- [12] J. T. Springenberg, A. Dosovitskiy, T. Brox, M. Riedmiller. Striving for Simplicity: The All Convolutional Net. *arXiv: 1412.6806*, 2014.
- [13] A. Dosovitskiy, P. Fischer, J. T. Springenberg, M. Riedmiller, T. Brox. Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks. *arXiv: 1406.6909*, 2014.
- [14] E. Hoffer, N. Ailon. Deep metric learning using Triplet network. *arXiv: 1412.6622*, 2014.
- [15] S. Ioffe, C. Szegedy. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. *arXiv: 1502.03167*, 2015.
- [16] K. He, X. Zhang, S. Ren, J. Sun. Deep Residual Learning for Image Recognition. *arXiv: 1512.03385*, 2015.

ライブラリ(フレームワーク)

- 論文で多く使われている Caffe や Torch
 - 国内は Chainer と TensorFlow がシェア拡大中

ライブラリ	開発	特徴	言語	ライセンス	GPU
Caffe	BVLC (UC Berkley)	<ul style="list-style-type: none">C++で高速 (Python, MATLABバイディング)画像処理のモデルが充実 (ModelZoo)	C++	BSD	<input type="radio"/> CUDA, cuDNN
Theano	Univ. Montreal	<ul style="list-style-type: none">Pythonベース, Numpyとの高い親和性独自のPython→C変換で高速DLライブラリの草分け	Python	BSD	<input type="radio"/> CUDA, cuDNN
Torch7	Facebook AI, DeepMind, Twitter	<ul style="list-style-type: none">独自言語 Lua (MATLABに似た実装)Lua→C変換で高速 (C/C++バイディング)拡張性が高い, ライブラリが充実	Lua	BSD	<input type="radio"/> CUDA, cuDNN
Keras	Univ. NY	<ul style="list-style-type: none">Theanoを隠蔽ライブラリが充実 → LSTM, RNN などにも対応	Python	MIT	<input type="radio"/> Theano と同じ
Chainer	PFI/PFN	<ul style="list-style-type: none">Pythonベース, Numpyとの高い親和性導入が簡単, Caffeモデルの読み込み可Python上でデバッグできる	Python	MIT	<input type="radio"/> CUDA, cuDNN
TensorFlow	Google	<ul style="list-style-type: none">Pythonベース, Numpyとの高い親和性分散環境で並列処理が可能 (Distributed Tensorflow)	C++/ Python	Apache 2.0	<input type="radio"/> CUDA, cuDNN

ModelZoo

- Caffe (BSD 2-Cluse license)
 - 画像系フレームワークとしては, デファクトスタンダードで, トップクラスの高速性
 - ドキュメントが充実 → <http://caffe.berkeleyvision.org/>
 - ImageNet 等の結果を再現可能
 - IPython notebook によるコード実例多数
- ModelZoo
 - 研究者の作った学習済みネットワークを共有
 - AlexNet, VGG, GoogLeNetなども
 - 最新の成果を極めて容易に試せる

CNNを用いた転移学習

- 転移学習
 - あるドメイン(データセット)で学習した識別器(特徴抽出器)を他ドメインでの識別器作成に役立てる
- 学習済みモデルを利用
 - 転用先のタスクと何らかの関係がある(と期待できる)十分に**大規模な**データセットで学習したネットワーク
- 2つのアプローチ
 - 特徴抽出器として利用 (Pre-trained feature)
 - 学習済NWを特徴抽出器とし,中間層の出力をを利用して識別器を作成
 - Fine-tuning
 - 学習済NWを初期値とし,適用先データセットでさらに学習
 - 所望のタスクを内包するものでなければ効果が薄い (むしろ悪化)

Pre-trained feature 実験

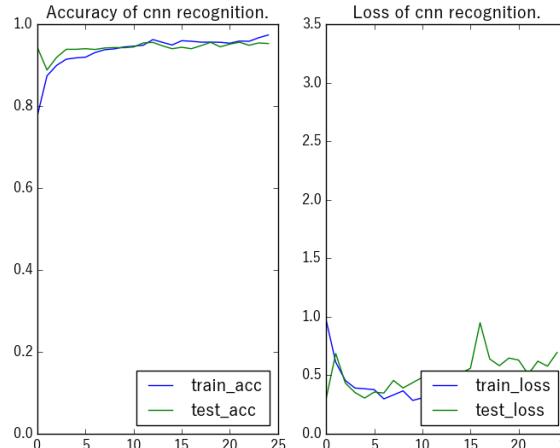
学習済みモデルからの転移学習がうまくいかないので,学習済みモデルを特徴量抽出器として使ってみた (1と2)

1. train_pre-traind-svm.py
2. train_pre-traind-mlp.py --batchsize 25 –plot
3. train_pre-traind-train.py --batchsize 25 –plot

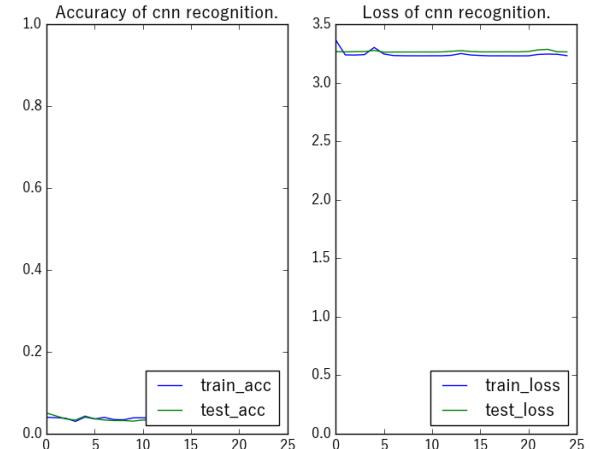
(1. SVM結果)

```
Loading Caffe model file ...
load training dataset
train: 4480 test: 498 accuracy: 0.847389558233
train: 4480 test: 498 accuracy: 0.901606425703
train: 4480 test: 498 accuracy: 0.815261044177
train: 4480 test: 498 accuracy: 0.895582329317
train: 4480 test: 498 accuracy: 0.849397590361
train: 4480 test: 498 accuracy: 0.714859437751
train: 4480 test: 498 accuracy: 0.849397590361
train: 4480 test: 498 accuracy: 0.833333333333
train: 4481 test: 497 accuracy: 0.835010060362
train: 4481 test: 497 accuracy: 0.766599597586
accuracy: 83.08%
('time spent:', 2448.2420089244843)
```

(2. MLP結果)



(3. fine-tuning)



Pre-trained feature 実験

学習済みの VGG 19-layer をから推定した特徴量で分類器を作成 (1)

1. train_pre-traind-mlp.py --batchsize 25 --plot
2. test_pre-traind-mlp.py --sample shiba.jpeg

shiba_inu



beagle



saint_bernard



1 100.00% shiba_inu
2 0.00% chihuahua
3 0.00% miniature_pinscher
4 0.00% basset_hound
5 0.00% pomeranian

1 100.00% beagle
2 0.00% american_bulldog
3 0.00% boxer
4 0.00% german_shorthaired
5 0.00% english_setter

1 100.00% saint_bernard
2 0.00% yorkshire_terrier
3 0.00% japanese_chin
4 0.00% american_pit_bull_terrier
5 0.00% basset_hound

海外企業の動向

- ・大手IT企業によるトップ人材の囲い込みが活発化

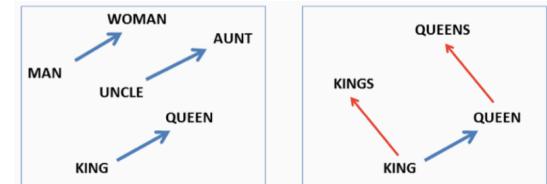
Google	<ul style="list-style-type: none">• Toronto大 Hinton 教授と学生(Alex Krizhevsky と Ilya Sutskever) の会社 DNNresearch を買収(2013)• Deep Learning の英国会社 Deep Mind Technologies を4億ドル(約420億円)で買収(2014)
Baidu	<ul style="list-style-type: none">• シリコンバレーに Institute of Deep Learning (IDL) を設立(2013)• Stanford大 Andrew Ng 教授を Deep Learning の研究所所長• 300億円を研究予算として投資(2014)
Facebook	<ul style="list-style-type: none">• 人工知能研究所 Facebook AI Lab 設立, New York大の Yann LeCun 教授を所長に招く(2013)• 人工知能の新興企業 Vicarious 社への4,000万ドルの投資ラウンドに参加(2014)
Yahoo!	<ul style="list-style-type: none">• 画像のDeep Learningのために LookFlow という会社を買収(2013)
Open AI (非営利団体)	<ul style="list-style-type: none">• Elon Musk, Sam Altman などが AI研究の非営利団体 OpenAI を設立(2016)
Toyota	<ul style="list-style-type: none">• Toyota Research Institute, Inc. 5年間で10億ドルを人工知能研究へ(2016)

自然言語処理への適用メリット

- データから自動的に特徴量を学習できる
 - 単語の表現方法(単語分散表現)
- 従来よりも広い文脈情報が扱える
 - 言語モデル(RNN言語モデル)
 - 機械翻訳(Seq-to-Seqモデル)
- 画像や音声などを扱うモデルとの親和性が高く、マルチモーダルなモデル構築が可能
 - 音声認識(DNN音響モデル+RNN言語モデル)
 - 画像キャプション(CNN+RNN言語モデル)

自然言語処理への適用

- word2vec [Mikolov+, 2013] (単語ベクトル)
 - 単語の持つ言語的な性質を含み,類似する単語が類似した固定長ベクトルを持つよう学習
 - 語彙数が多くなっても効率的に学習
 - 著者がツールを公開→世界中で注目



上: 意味ベクトル演算の代表例

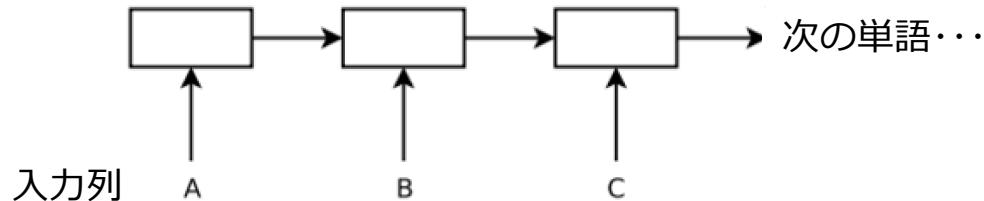
例: word2vec で各コーパスを学習し, 「パンク」検索した → コーパスやドメインによる差が出る

Wiki 日本語 2016/02 (6,235,945文)		Yahoo! 知恵袋:車 (1,439,341文)		教えてgoo:車 (391,912文)		みんなQA:QA (874,457文)	
word	distance	word	distance	word	distance	word	distance
パンク	1.000000	パンク	1.000000	パンク	1.000000	パンク	1.000000
パンク・ロック	0.741283	バースト	0.671790	バースト	0.614949	バースト	0.620422
ロック	0.738832	釘	0.656777	チューブレス	0.529546	応急	0.585563
エモ	0.723946	チューブレスタイヤ	0.610192	破裂	0.512507	テンパータイヤ	0.558305

※ word2vec は隠れ層を持たないのでDeepではないが, DLの事前学習に用いられるので DLの技術のひとつ

自然言語処理への適用

- Recurrent NN言語モデル [Mikolov+,2010]
 - 言語モデルとは文の尤もらしさをスコア化するモデル
 - 過去の履歴を考慮して,現在の単語を予測
 - RNN (with LSTM, GRU) が主流



例: RNNにより文字単位で「吾輩は猫である」を学習し,文生成した

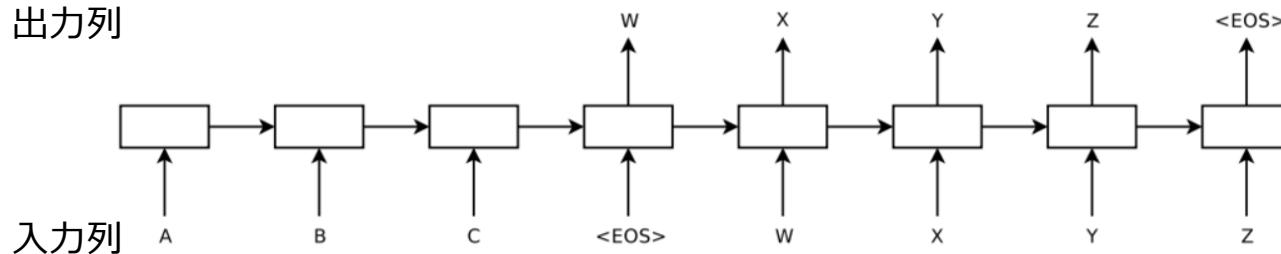
吾輩は猫である。鍋の中がちょっと立派な事はないから、そうしておやじや君は何と云う。甘らんですか。そりや、それですばからかったのは、つまり葉中だけ上げて冷淡に描き出した。今日は日曜だつてもしかしたに違ない。

自然言語処理への適用

- Sequence to Sequence [Sutskever+, 2014]

- LSTM (Long-Short Term Memory) を2つ用い, 入力系列を固定長ベクトルに変換し, そのベクトルから出力系列を生成
- 機械翻訳も系列→系列を生成するモデルとして扱える

出力列



“ABC”という単語列から“WXYZ”という単語列への翻訳

例: Seq-to-Seq により文字単位で百人一首を文字単位で学習し, 下の句を生成した

これやこの行くも帰るも別れては

知るも知らぬぬ逢坂の関

これやこの行くも帰るも別れては

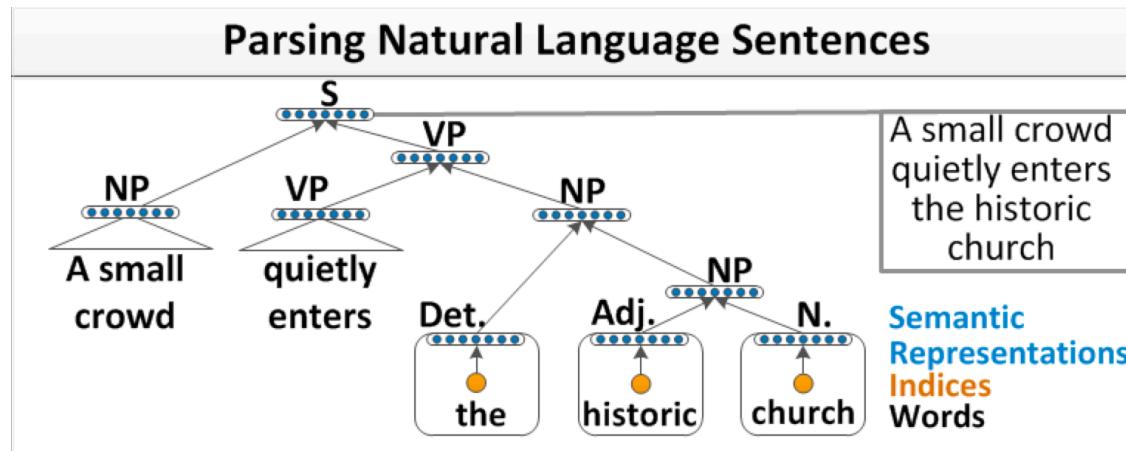
知るも知らぬ恋逢坂の思し

これやこの行くも帰るも別れては

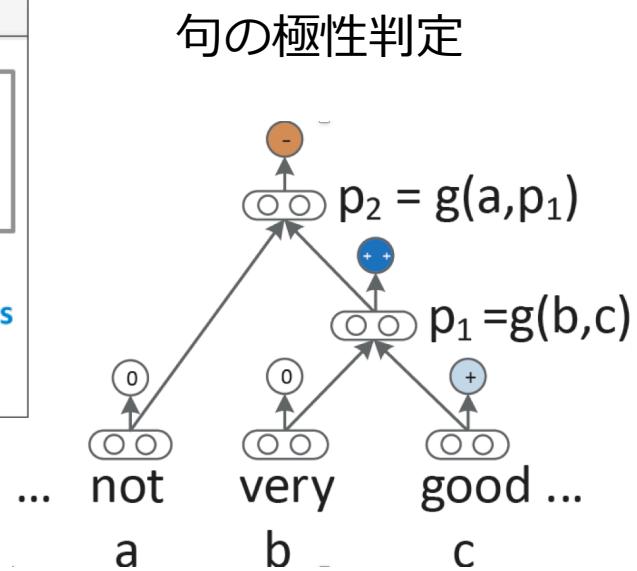
知るも知らぬも逢坂の関

自然言語処理への適用

- Recursive Autoencoder [Socher+, 2011]
 - 句や文の意味をそれを構成する単語の合成により得る
 - 単語ベクトルから文の意味を再帰的に計算
 - 句の感情極性分類も実現 [Socher+, 2013]

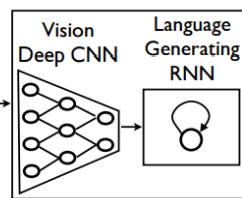


構文構造を再帰的にNNで学習

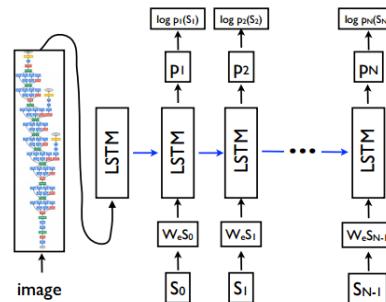


マルチモーダル

- Google NIC [Vinyals+, 2015]
 - 画像だけからキャプションを生成
 - 画像表現CNN + LSTM(文生成, 翻訳)を直列に接続することで、マルチモーダルな入出力を自然に統合



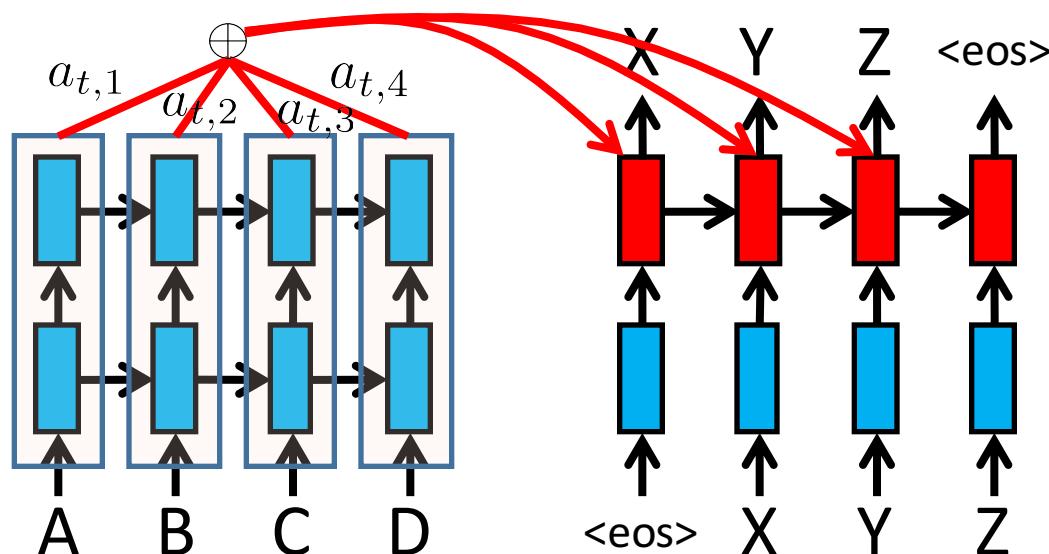
A group of people shopping at an outdoor market.
There are many vegetables at the fruit stand.



Describes without errors	Describes with minor errors	Somewhat related to the image	Unrelated to the image
 A person riding a motorcycle on a dirt road.	 Two dogs play in the grass.	 A skateboarder does a trick on a ramp.	 A dog is jumping to catch a frisbee.

アテンション

- RNN の標準的なテクニックになりつつある
 - RNN翻訳 [Bahdanau+, 2015]
 - どこに「注意」して翻訳するかを学習
 - アテンションのモデル化によって,アライメントを明示的に与える必要がなくなった



深層学習(Deep Learning)まとめ

- “Deep learning” という言葉 (2006～)
 - [Hinton and Salakhutdinov, Science, 2006]
 - 従来より,さらに深く,幅広いニューラルネット
 - 2010年頃で7～8層程度, 現在は20層以上も
- 2012年の大ブレーク以来,アカデミア,産業界で利用
 - 2014～2015中に出された関連論文数は 1500 超える
- 音声認識,画像認識の分野で劇的な精度向上
 - 多くがすでに実用化
- 自然言語処理の分野でも,従来法の置き換えが加速
 - word2vec, RNN言語モデルなどで実用化
- 生データから目的変数まで end-to-end で構造学習
 - 従来の特徴量エンジニアリングを最小限にできる