

名古屋大学特別講義 2016年6月29日(水)

ディープラーニングによる 自然言語処理 (概要編)

日本アイ・ビー・エム株式会社

東京基礎研究所

坪井 祐太

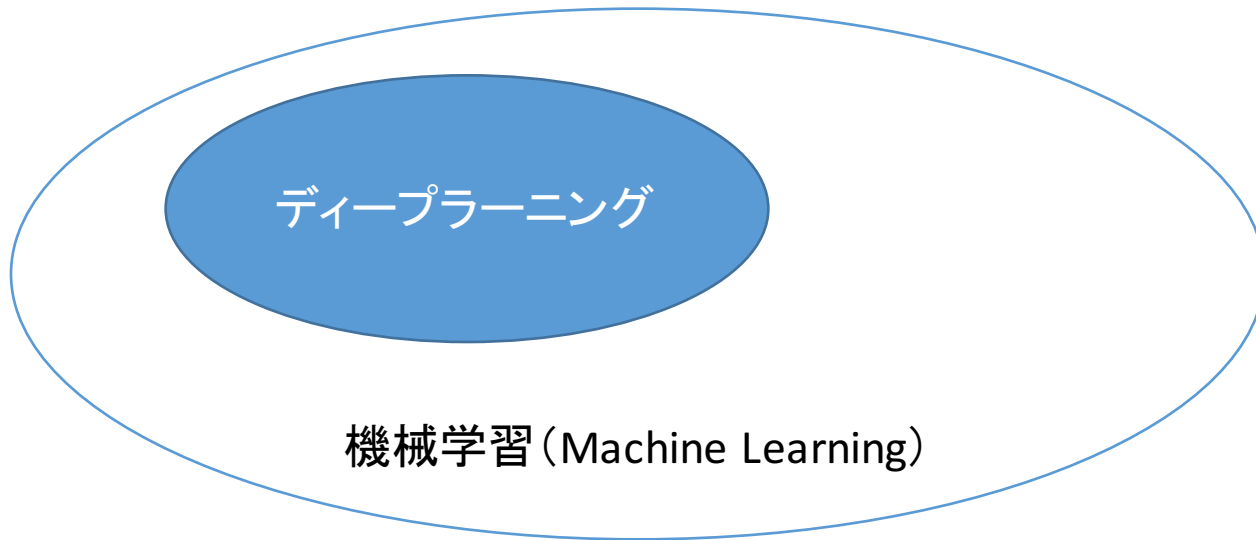
yutat@jp.ibm.com

講義の概要

- 概要編
 - ディープラーニング概要
 - 自然言語処理におけるニューラルネットワークの分類
- 技術編
 - ニューラルネットワーク技術詳細

ディープラーニングとは？

- 機械学習の一分野



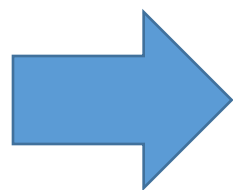
機械学習101 (教師あり学習) 事例からの学習

訓練データ (正解)



Dog

学習



学習結果

$f(\text{image})$

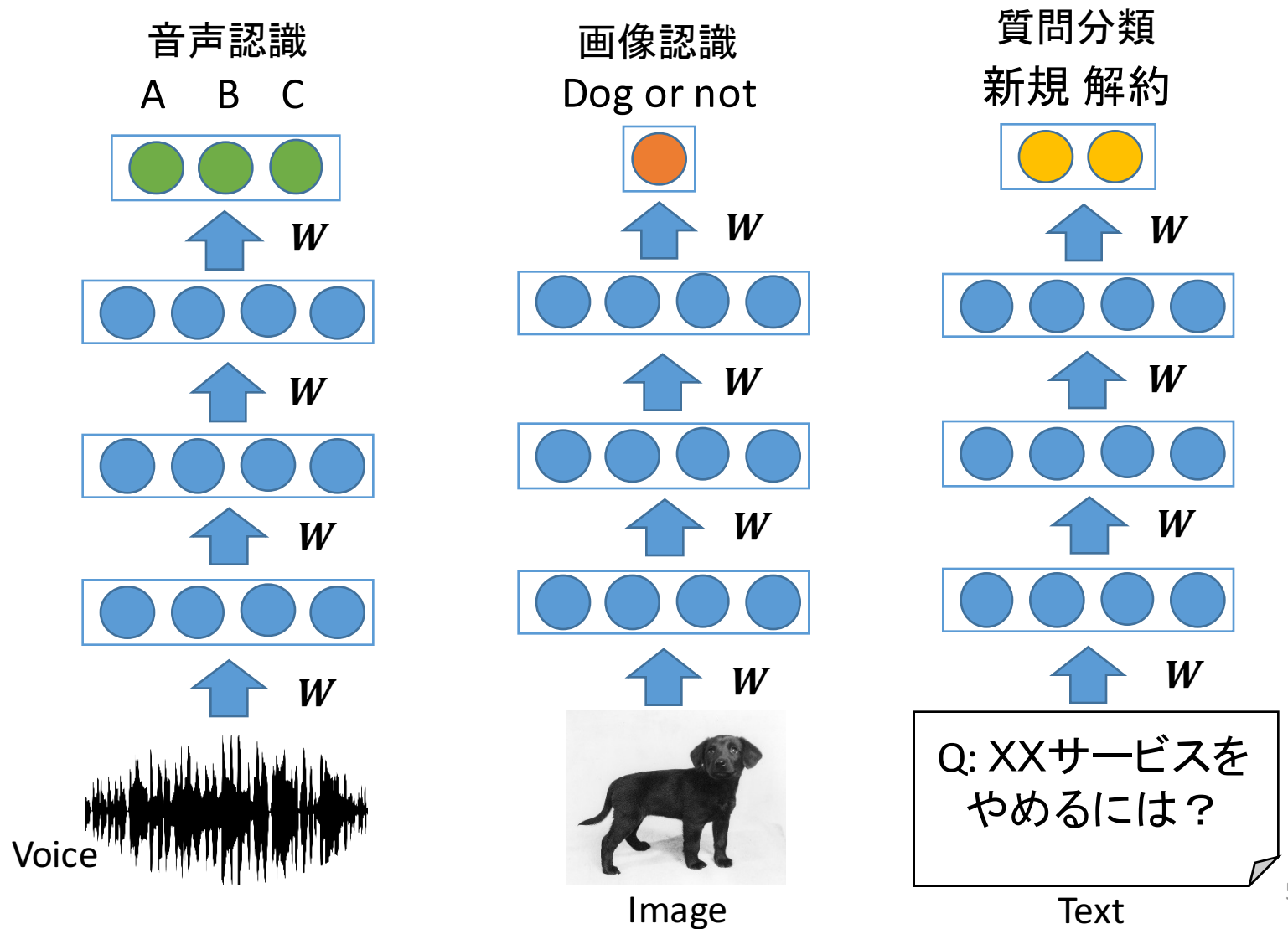
$\rightarrow \{\text{Dog, Not dog}\}$



Not dog

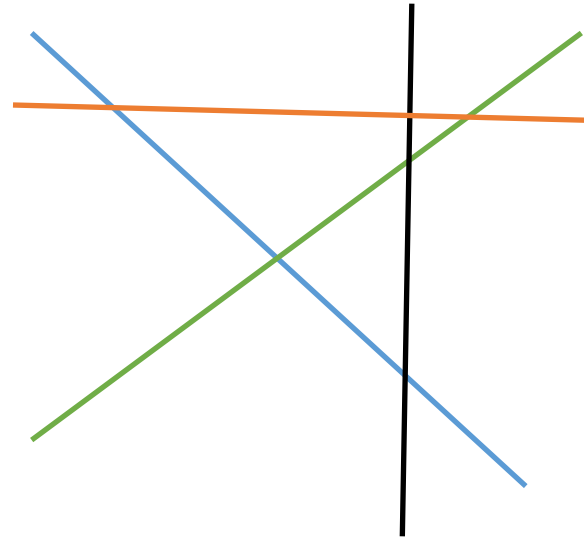
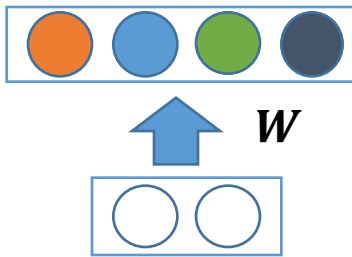
⋮

ディープラーニングとは→深い階層を持つ人工ニューラルネットワークの学習



ニューラルネットの表現能力

- 2次元入力、4隠れ変数の例→4本の直線で2次元空間を分割



→パラメータ数に対して指数的な領域を表現可能

ディープラーニングの工学的な利点

- これまでの機械学習の課題: 入力の実現方法
 - ベクトル表現する方法(特徴抽出)をうまく選ぶ必要がある
 - データごとに特徴抽出のパラメータをチューニングする必要性
→特徴抽出のチューニング次第で精度は大きく異なる
 - 代表的な特徴抽出手法
 - 画像: SIFT, HoG, Textons, Spin image, RIFT, GLOH, ...
 - 音声: Spectrogram, MFCC, Flux, ZCR, Rolloff, ...
 - テキスト: TF/IDF, N-gram, POS Tagging, Syntactic/Semantic parsing, ...
- 利点: タスク解決と特徴抽出の同時学習
 - 素の入力形式 (RGB, Sound wave, Character sequence, etc.)
 - タスクを解くために必要な特徴量抽出も同時学習
 - 特徴抽出個別のチューニングがニューラルネットのチューニングと一体化

学習された階層的な特徴抽出（画像）

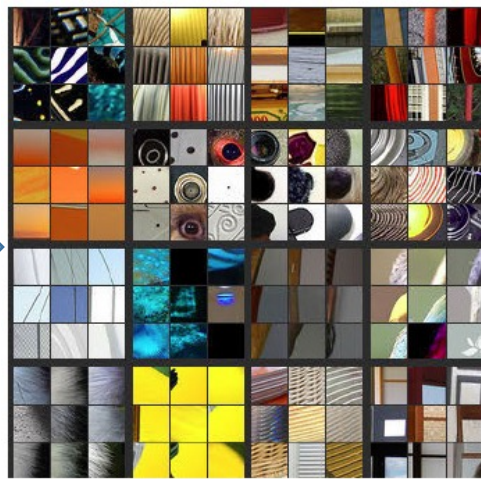
上層になるほど抽象的な特徴が学習されている

- 図は M.D. Zeiler, R. Fergus, “Visualizing and Understanding Convolutional Networks” ECCV 2014
より引用

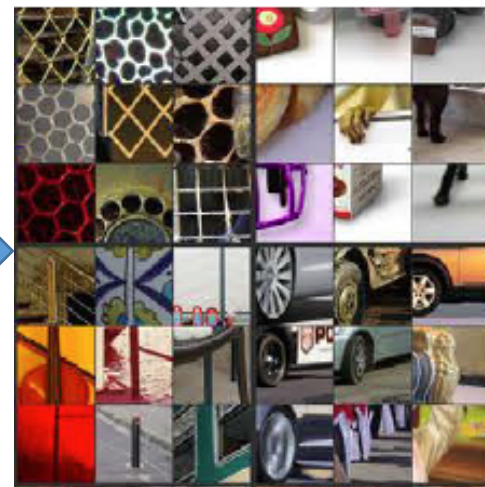
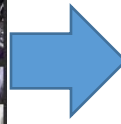
入力画像



隠れ層1
（エッジ）



隠れ層2
（コーナー、等高線）



隠れ層3
（パーツ）



出力（画像ラベル）

なぜ今ディープラーニングなのか

- 大規模データの出現



- 高速計算モジュールのコモディティ化 (GPU, FPGA)



大規模書き起し音声データ

表の数値は以下の論文より引用

<http://arxiv.org/pdf/1412.5567v2.pdf>

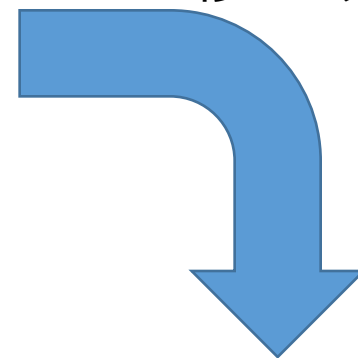
<http://static.googleusercontent.com/media/research.google.com/ja//pubs/archive/38130.pdf>

データセット	タイプ	時間	話者
WSJ	読み上げ	80	280
Switboard	会話	300	4000
Fisher	会話	2000	23000
Baidu	読み上げ	5000	9600
Google	音声検索	5780	N/A
Google	Youtube	1400	N/A

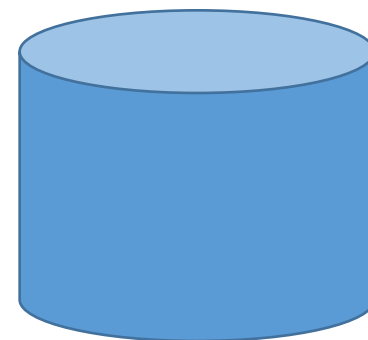
コンシューマサービスを通して生成される膨大なデータが訓練データとしても利用可能に



クリックログ,
修正ログ, etc.

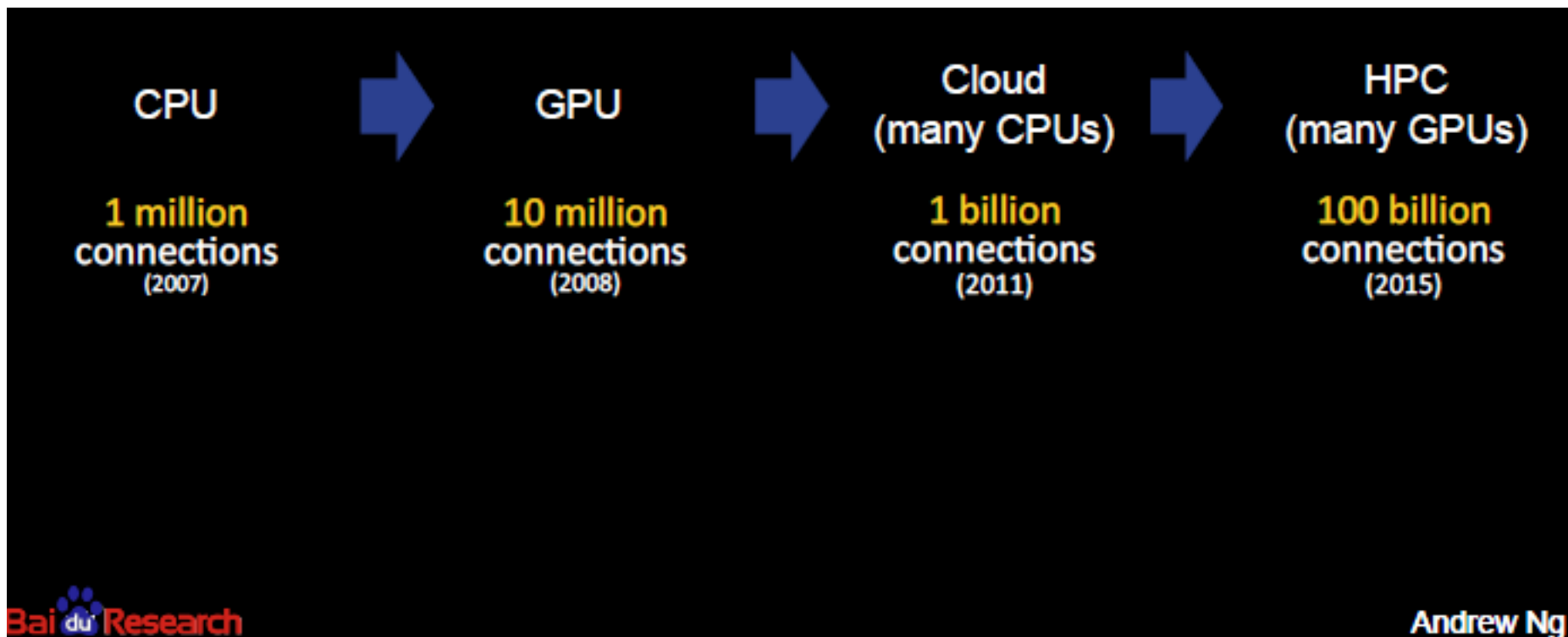


訓練データ



高速計算モジュールのコモディティ化により大規模なニューラルネットワークの学習が可能に

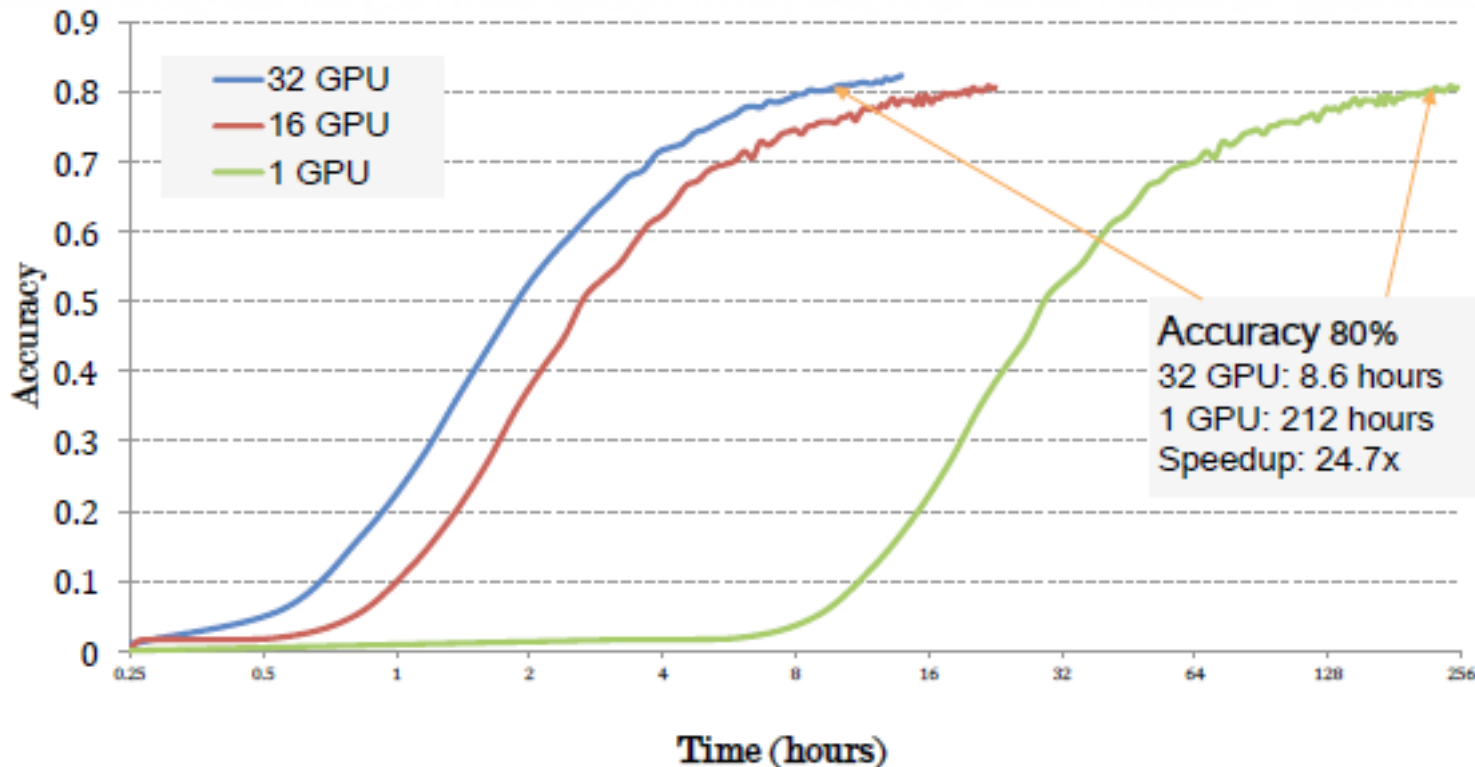
- 図は <http://on-demand.gputechconf.com/gtc/2015/presentation/S5818-Keynote-Andrew-Ng.pdf> より引用



ディープラーニングの学習はHPCへ

図は <http://on-demand.gputechconf.com/gtc/2015/presentation/S5280-Ren-Wu.pdf>より引用

Speedup (wall time for convergence)



Validation set accuracy for different numbers of GPUs

データセンターでの運用に向けて低消費電力にも焦点

Microsoft ResearchによるFPGA/GPUの分類速度・消費電力の比較

表はAccelerating Deep Convolutional Neural Networks Using Specialized Hardware <http://research.microsoft.com/pubs/240715/CNN%20Whitepaper.pdf> より引用

	CIFAR-10 [4]	ImageNet 1K [1]	ImageNet 22K [2]	Max Device Power
Catapult Server + Stratix V D5 [3]	2318 images/s	134 images/sec	91 images/sec	25W
Catapult Server + Arria 10 GX1150 [8]	-	~233 images/sec (projected)	~158 images/sec (projected)	~25W (projected)
Best prior CNN on Virtex 7 485T [5]	-	46 images/sec ³	-	-
Caffe+cuDNN on Tesla K20 [6]	-	376 images/sec	-	235W
Caffe+cuDNN on Tesla K40 [6]	-	500-824 images/sec ⁴	-	235W

Table 1: Comparison of Image Classification Throughput and Power.

ディープラーニング専用集積回路 (ASIC)

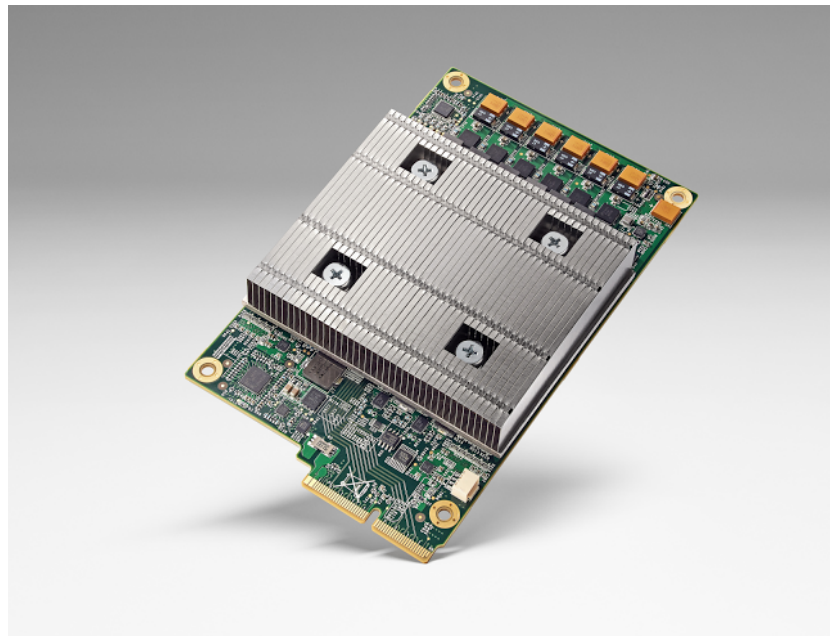
- Google Tensor Processing Unit (TPU)

- 図は

- <https://cloudplatform.googleblog.com/2016/05/Google-supercharges-machine-learning-tasks-with-custom-chip.html> より

- データセンターで運用中

- Web Search
 - Street View
 - AlphaGo

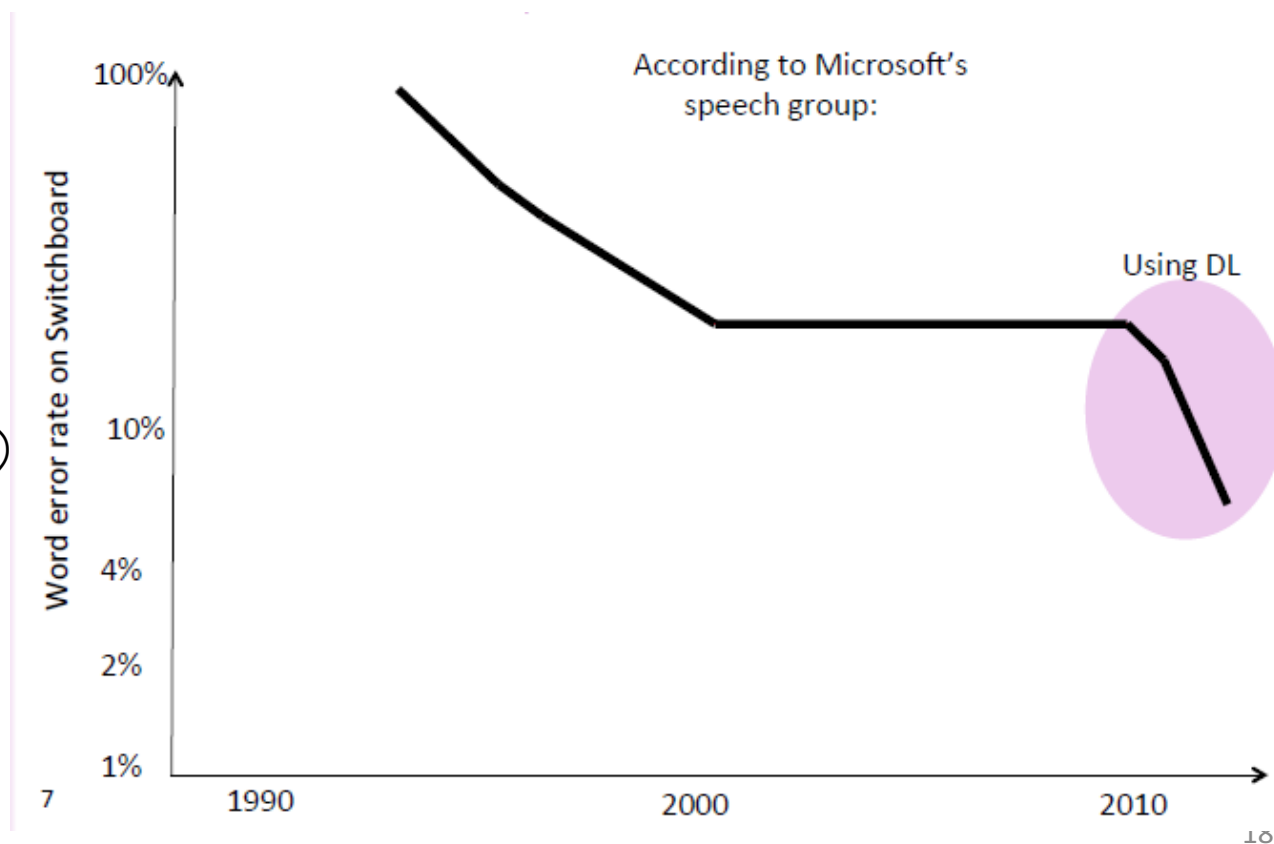


ディープラーニングの発展

音声認識における発展

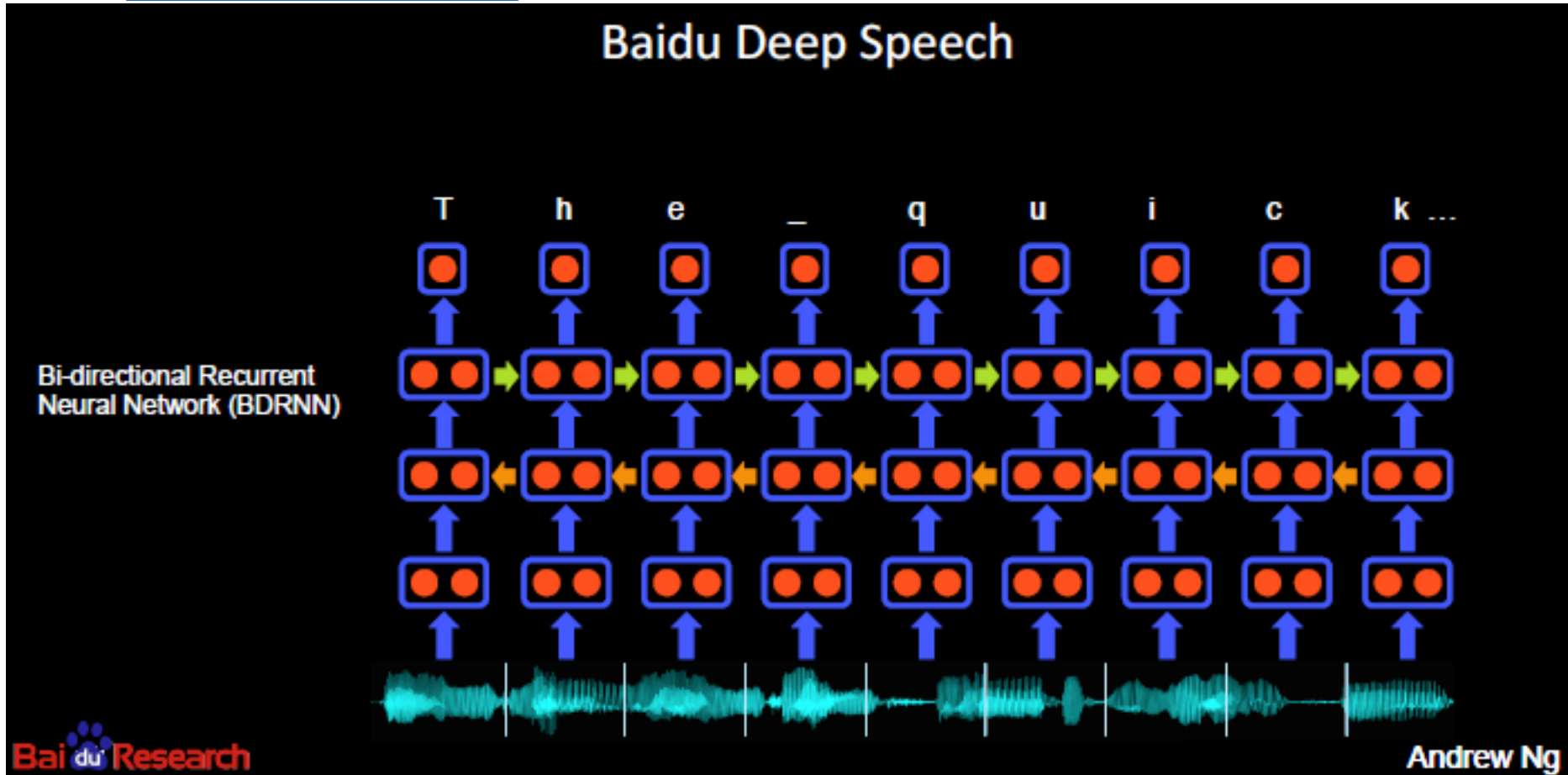
- 図は<http://www.iro.umontreal.ca/~bengioy/talks/KDD2014-tutorial.pdf>より引用

単語誤認識率
(小さい方がよい)



ディープラーニングによる音声認識 (Baidu)

- 図は <http://on-demand.gputechconf.com/gtc/2015/presentation/S5818-Keynote-Andrew-Ng.pdf> より引用



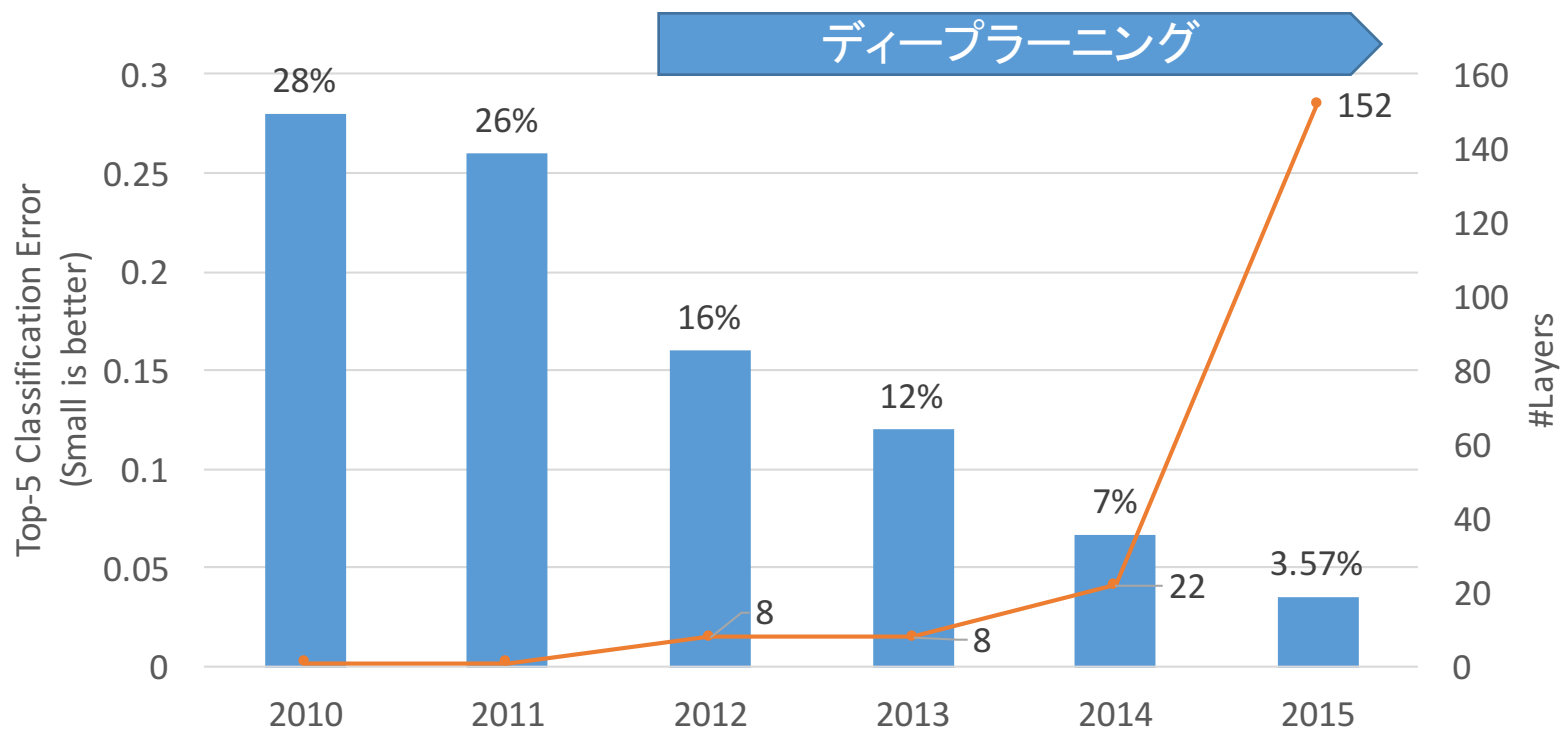
IBMのディープラーニング音声認識システム

- 表はSoan et al., "The IBM 2015 English Conversational Telephone Speech Recognition System", 2015より引用

System	AM training data	SWB	CH
Vesely et al. [8]	SWB	12.6	24.1
Seide et al. [9]	SWB+Fisher+other	13.1	–
Hannun et al. [10]	SWB+Fisher	12.6	19.3
Zhou et al. [11]	SWB	14.2	–
Maas et al. [12]	SWB	14.3	26.0
Maas et al. [12]	SWB+Fisher	15.0	23.0
Soltau et al. [13]	SWB	10.4	19.1*
This system	SWB+Fisher+CH	8.0	14.1

画像認識では150層以上

- ImageNet Large Scale Visual Recognition Challengeの1000クラス分類タスクのエラー率の推移
 - 数値はhttp://image-net.org/challenges/talks/ILSVRC2015_12_17_15_clsloc.pdf より



ImageNet Large Scale Visual Recognition Challenge は人でも難しい課題

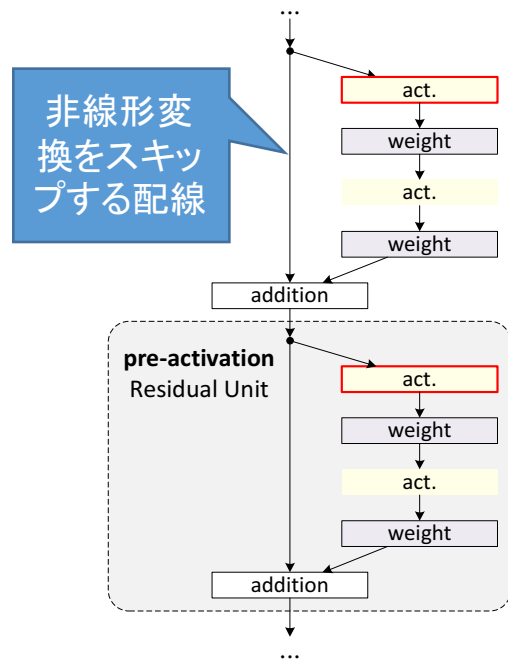
- 図は Russakovsky et al., ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015. より



...画像認識では1000層以上

- 図・表はHe et al. Identity Mappings in Deep Residual Networks, 2016より
- CIFAR-10/100分類エラー率
 - 100層台から1000層に増やすことでエラーが低下

dataset	network	baseline unit	pre-activation unit
CIFAR-10	ResNet-110(1layer)	9.90	<u>8.91</u>
	ResNet-110	6.61	<u>6.37</u>
	ResNet-164	5.93	<u>5.46</u>
	ResNet-1001	7.61	<u>4.92</u>
CIFAR-100	ResNet-164	25.16	<u>24.33</u>
	ResNet-1001	27.82	<u>22.71</u>



音声認識でも約50層

- 表はGhahremani et al. "LINEARLY AUGMENTED DEEP NEURAL NETWORK", 2016より
- 深くすることでエラー低下とともにパラメータ数を減らせた

Model	Network Size			Training		Validation		
	Num.of.H Layers	Layer.Size	Params	CE	% Frame Err	CE	% Frame Err	% PER
DNN+Sig	2	2048X2048	10.9M	0.66	21.39	1.23	37.67	23.63
LA-DNN+Sig	4	1024X512	5.9M	0.61	20.5	1.18	35.83	22.39
LA-DNN+Sig	6	1024X512	8.0M	0.60	20.3	1.18	35.7	22.28
LA-DNN+Sig	8	1024X256	5.6M	0.58	19.7	1.19	35.7	22.08
LA-DNN+Sig	12	512X512	7.2M	0.65	22.08	1.16	35.74	22.08
LA-DNN+ReLU	3	1024X256	2.9M	0.61	20.7	1.20	35.77	22.39
LA-DNN+ReLU	6	1024X256	4.5M	0.54	18.6	1.22	35.5	22.08
LA-DNN+ReLU	12	512X256	3.8M	0.55	19.2	1.21	35.5	21.8
LA-DNN+ReLU	24	256X256	3.5M	0.55	19.31	1.21	35.3	22.06
LA-DNN+ReLU	48	256X128	3.4M	0.56	19.5	1.21	35.4	21.7

Table 1. Results on TIMIT using different models.

企業が大物教授を次々と迎える

- 写真は<http://www.kdnuggets.com/wp-content/uploads/photo.jpg>より引用

Geoffrey E.
Hinton (Google
& Toronto大学)

Yoshua Bengio (Montreal大学)
→ IBM Research/Watson Groupと協業
<http://asmarterplanet.com/blog/2015/07/promise-deep-learning.html>

Yann LeCun
(Facebook AI
Research &
NYU)

Andrew Ng
(Baidu / 元
Stanford大学)



自然言語処理における ディープラーニングの現状

- 機械学習コミュニティから自然言語処理が注目されている
 - Christopher D. Manning. 2015. Last Words. Computational Linguistics より引用
 - Yann LeCun: ディープラーニングの次のターゲットは自然言語理解である。単語だけでなくパラグラフや文を理解することを目指す。
 - Geoff Hinton: ディープラーニングで次の5年で最もエキサイティングな領域はテキストとビデオの理解だ。5年後にYoutubeビデオを見て説明する仕組みが出てきていなかったとしたら残念だ。
 - Yoshua Bengio: 機械翻訳システムなど言語処理に向けたディープラーニング研究を促進している
 - Michael Jordan: もし10億ドルあったら何をしますか、と尋ねられたら自然言語処理に特化したNASA規模のプログラムを作ると答えるだろう
 - ニューラルネットワークを用いた手法が多く発表されている
 - →画像認識や音声認識ほど大成功を収めているとは言えない
 - 多くのタスクで既存手法と同等または若干上回る程度

自然言語処理(Natural Language Processing; NLP)の特徴 -画像認識・音声認識との比較-

• 離散入力

- テキストは記号列（「あ」と「い」の距離は定義されない vs 画像のRGB）
- 典型的には疎な離散値ベクトルとして入力を表現
- 離散であるため組み合わせ特徴量の構成が直感的で比較的構成しやすい
 - E.g. “New” \wedge “York” \rightarrow “New York”
- 課題: 組み合わせ特徴量は指数的に増大

• 可変長入出力

- テキストは可変長
- 課題1: 機械学習アルゴリズムで扱うためには、入力テキストを固定長の特徴ベクトルで表現する必要がある
- 翻訳・要約・質問応答などの応用では、出力もテキスト(入力と出力の長さは普通異なる)
- 課題2: 可変長記号列を出力する必要性

ネットワーク構造に基づく自然言語 処理におけるディープラーニング適 用の分類

- フィードフォワードニューラルネットワーク (Feed-forward Neural Networks)
- リカレントニューラルネットワーク (Recurrent Neural Networks; RNNs)
- 畳み込みニューラルネットワーク (Convolutional Neural Networks; CNNs)
- 再帰ニューラルネットワーク (Recursive Neural Networks)

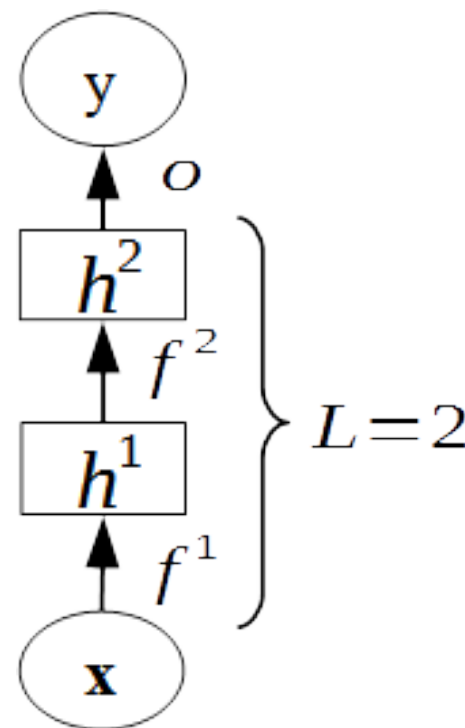
フィードフォワードニューラルネットワーク (Feed-forward Neural Networks)

• フィードフォワードニューラルネットワークの例

- x : 入力ベクトル
- y : 予測
- l : 階層インデックス
- h : 隠れ変数ベクトル
- W : 重み行列
- f : 活性化関数 (シグモイドなど)

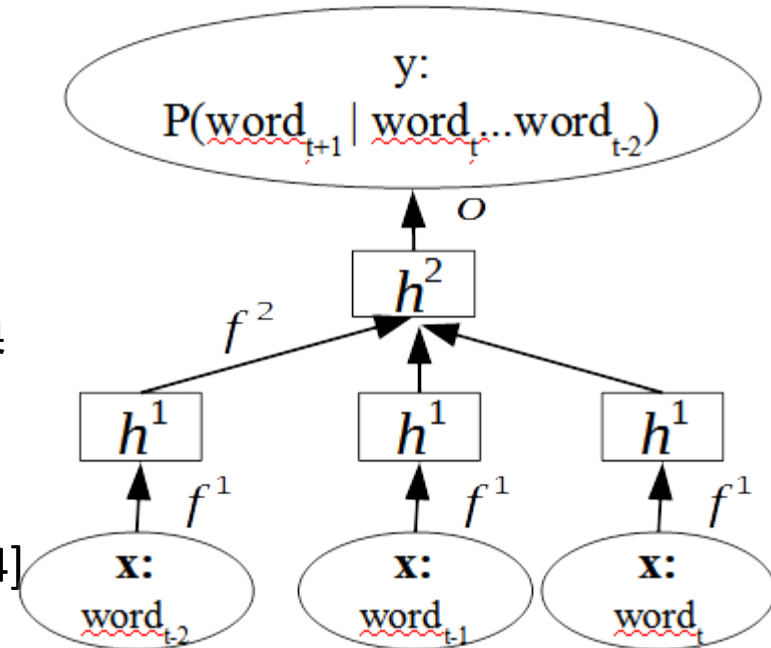
$$h^{\ell} = f^{\ell}(W^{\ell}h^{\ell-1})$$

$$\hat{y} = o(W^o h^L)$$



フィードフォワードニューラルネットワークの応用例

- 言語モデル [Bengio et al., 2003]
 - 次の単語を予測するモデル
 - 文の生成などに利用される
 - 1層目で窓幅の単語を個別に非線形変換
- 機械翻訳 [Devlin et al., 2014]
 - 原言語と対象言語両方を入力
- 品詞タグ付 [Ma et al., 2014] [Tsuboi, 2014]
- 構文解析 [Chen and Manning, 2014]



- 利点: 明示的に組み合わせ特徴量を使うことなく、暗に特徴量の組み合わせを考慮できる
- 課題: 固定長の入力を得るために試行錯誤が必要

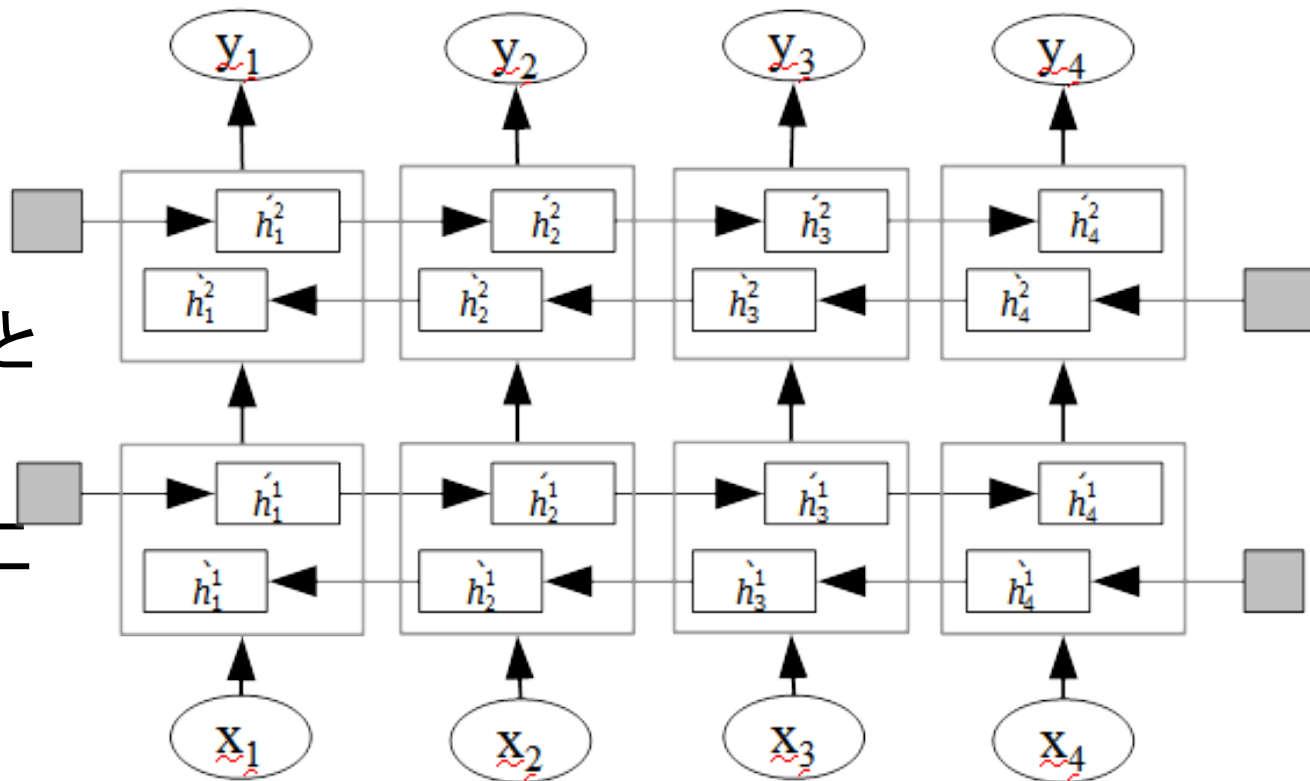
リカレントニューラルネットワーク (Recurrent Neural Networks; RNNs)

- 双方向RNNの例

$$\hat{h}_t^\ell = f^\ell(W[\hat{h}_{t-1}^\ell; \hat{h}_t^{\ell-1}; \acute{h}_t^{\ell-1}]) \quad \leftarrow \text{後向き走査}$$

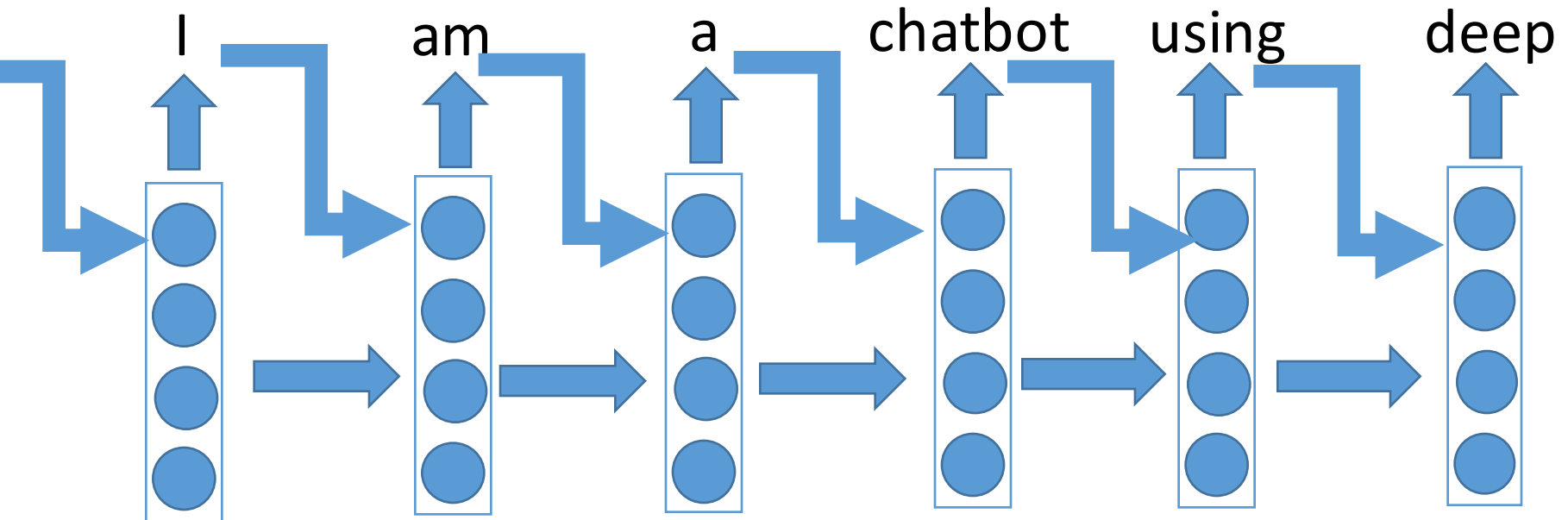
$$\acute{h}_t^\ell = f^\ell(W[\acute{h}_{t+1}^\ell; \acute{h}_t^{\ell-1}; \hat{h}_t^{\ell-1}]) \quad \leftarrow \text{前向き走査}$$

- 課題: 入力と出力の長さが同じであることが必要



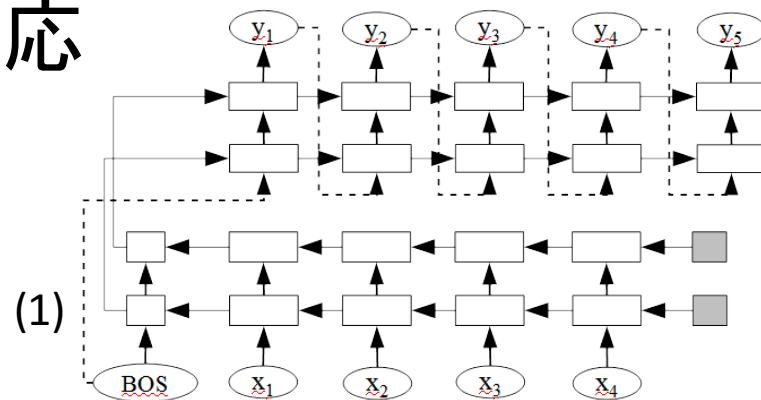
言葉を生成するためのRNN

- リカレントネットワークによる言語モデル[Mikolov et al., 2013]

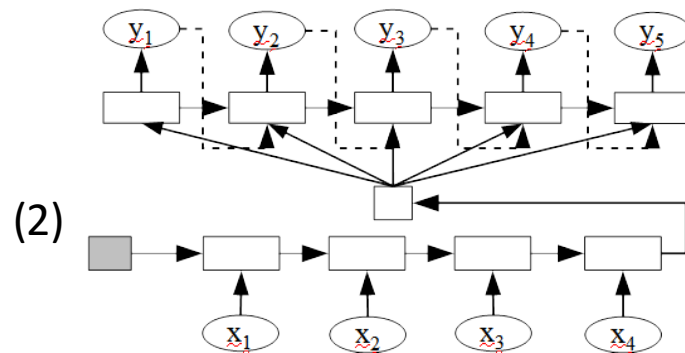


発展: 入力列のエンコーダと出力列のデコーダを接続し可変長出力に対応

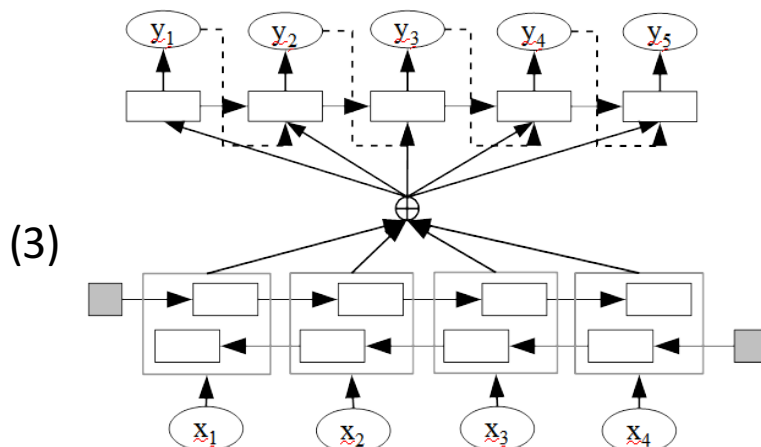
1. エンコーダの末尾状態をデコーダの先頭に接続
[Sutskever et al., 2014]



2. エンコーダの末尾状態をデコーダのすべての点で参照
[Cho et al., 2014]



3. エンコーダの各点の状態を重みつき線形和したベクトルをデコーダのすべての点で参照
(a.k.a. ソフトアテンションモデル)
[Bahdanau et al., 2014]

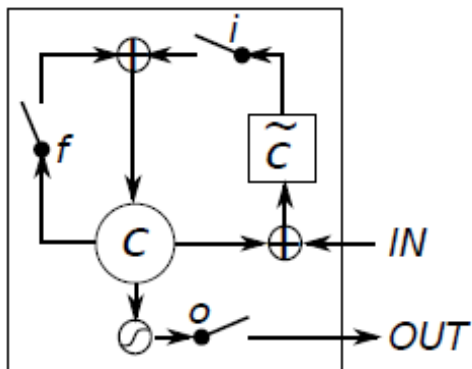


エンコーダには教師信号が直接入らないためRNNsに長期記憶を持たせるための手法が併用される

- Multiplicative Gate Units

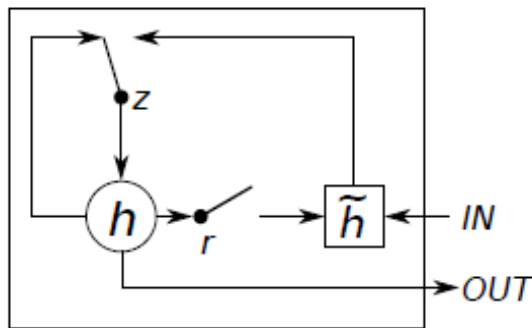
- Figures from (Chung et al., 2014)

- Figures from (Mikolov et al., 2015)



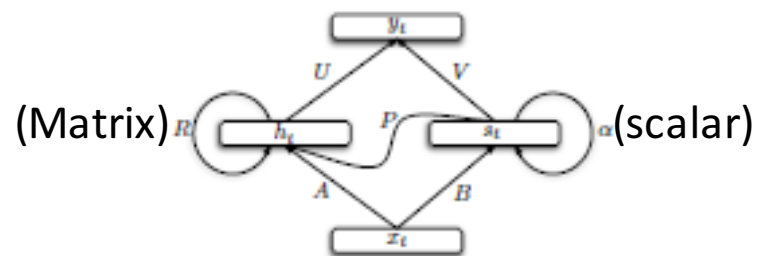
(a) Long Short-Term Memory

(Hochreiter and Schmidhuber, 1997)



(b) Gated Recurrent Unit

(Cho et al., 2014)



$$\begin{aligned}
 s_t &= (1 - \alpha)Bx_t + \alpha s_{t-1}, \\
 h_t &= \sigma(Ps_t + Ax_t + Rh_{t-1}), \\
 y_t &= f(Uh_t + Vs_t)
 \end{aligned}$$

structurally constrained recurrent nets (SCRN)

画像エンコーダ (CNN) とテキストデコーダ (RNN)

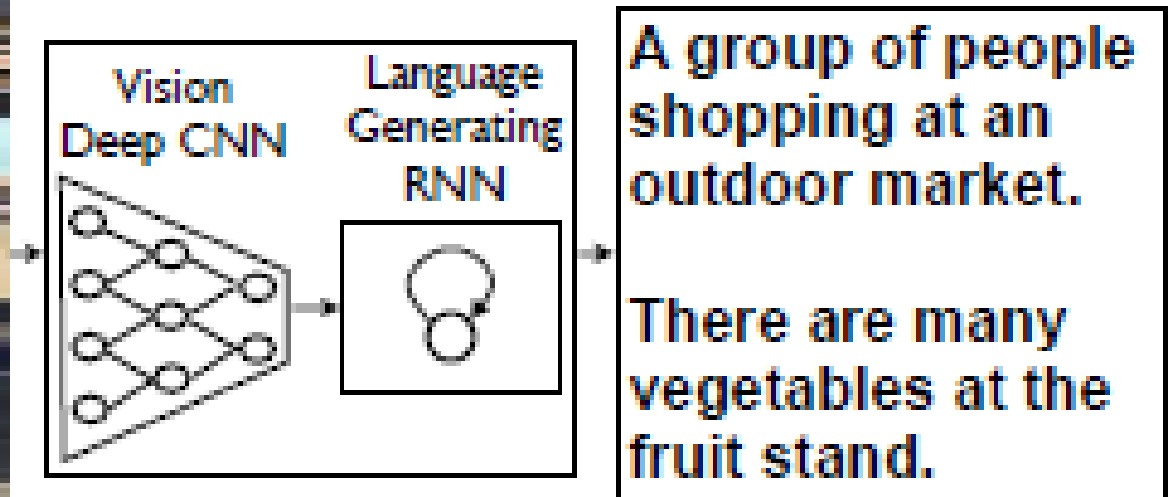


Figure from the original paper [Show and Tell: A Neural Image Caption Generator (Vinyals et. al., 2014)]

- [*New York Times Article: Researchers Announce Advance in Image-Recognition Software*](#)

動画の説明文生成

図はYao et al., Describing Videos by Exploiting Temporal Structure, 2015より引用



+Local+Global: A man and a woman are talking on the road

機械翻訳コンペティション (IWSLT2015) 英独翻訳で圧勝

TED : MT English-German (MT_{EnDe})

System	<i>case sensitive</i>		
	BLEU	NIST	TER
SU	30.85	6.9898	51.13
KIT	26.18	6.4640	55.52
UEDIN	26.02	6.4518	56.05
HDU	24.96	6.3170	56.94
PJAIT	22.51	6.0412	59.03
BASELINE	20.08	5.7613	61.37

System	BLEU
IWSLT'14 best entry [14]	26.2

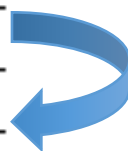
<i>Our systems</i>	
Single NMT (non-adapted)	25.6
Single NMT (adapted)	29.4 (+3.8)
Ensemble NMT (adapted)	31.4 (+2.0)

• 上図:
http://workshop2015.iwslt.org/downloads/IWSLT_Overview15.pdf より引用

• 2位以下を大きく離して圧勝

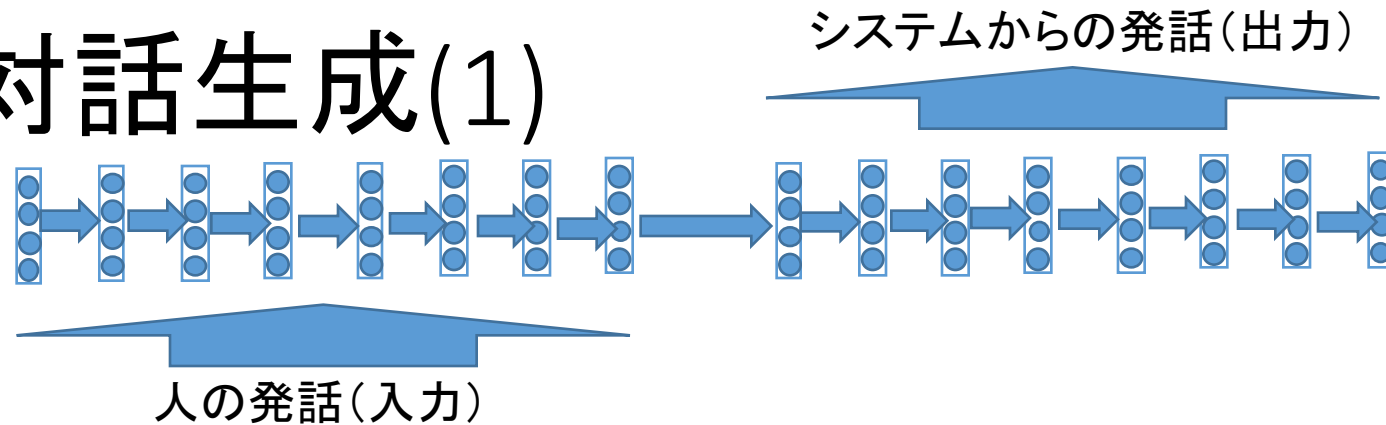
• 下図:
http://workshop2015.iwslt.org/downloads/IWSLT15_ThangLuoNg.pdfより引用

• 分野適応: WMTデータ4.5Mで学習したモデルをIWSLTデータ0.2M (TED talk) で再学習 (再学習は初期学習率=1.0から2epochごとに半減し12epoch)



分野適応

対話生成(1)



- IT Helpdeskデータで学習した対話生成例: 例は Vinyals and Le, “A Neural Conversational Model”, ICML Deep Learning Workshop 2015 より引用
 - **Describe your problem:** i am having issues accessing vpn
 - **Machine:** could you please let me know what are the operating systems you are using ?
 - **Human:** i am using linux
 - **Machine:** ok it was in the ssl vpn on which you stated it was broken
 - **Machine:** please check if you are logged in to corp network
 - **Machine:** what is the error that you are running please
 - **Human:** i am seeing an error related to vpn
 - **Machine:** what is the error message that you are getting when connecting to vpn using network connect ?
 - **Human:** connection refused or something like that
 - **Machine:** may i know the version of network connect you connect ?

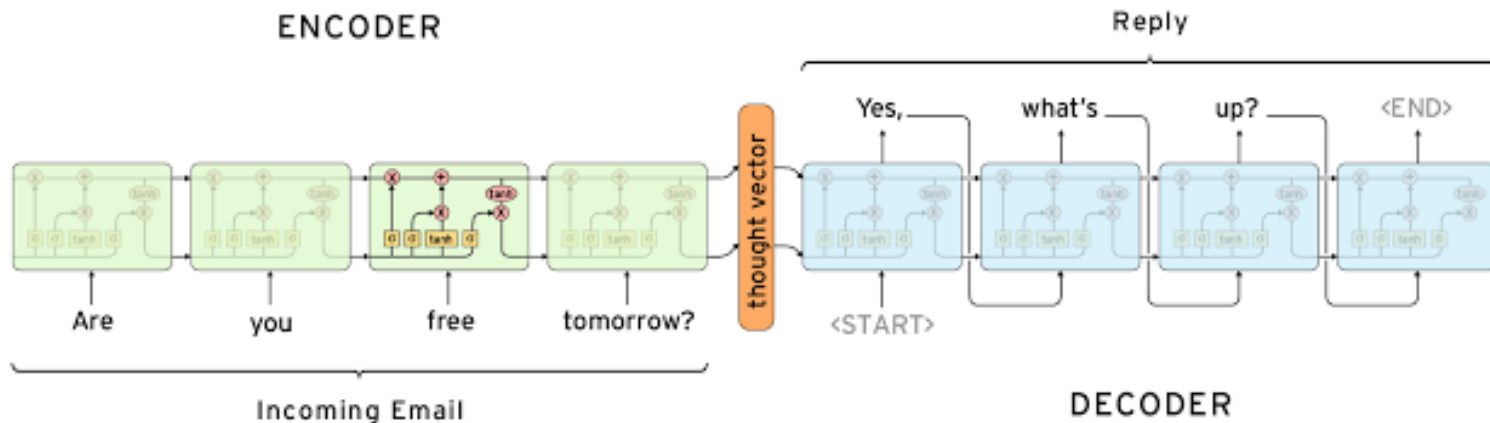
対話生成(2)

- 対話を真似るように学習しただけだが、タスク解決できているように見える
 -
 - **Machine:** *try this*
 - **Machine:** *goto <URL> page and enter your*
 - *name and password*
 - **Human:** *done*
 - **Machine:** *try logging in with <NAME> and*
 - *check if its working fine now*
 - **Human:** *yes , now it works !*
 - **Machine:** *great*
 - (<URL> は実際にVPNの情報を含んでいるURLを含んでいる)
- 模倣学習の利点: 数値的に達成度が図りにくいタスクに適している

メール自動返答生成

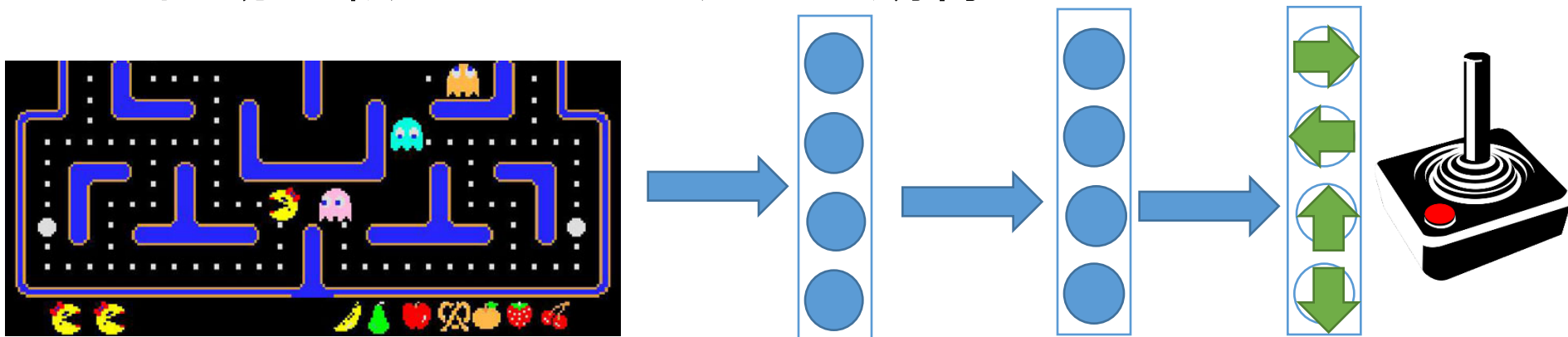
- エンコーダ・デコーダアプローチによりメールへの返答候補を自動生成
 - スマートフォンのメールアプリ(Inbox)で試用中
- 図は以下より引用

<http://googleresearch.blogspot.jp/2015/11/computer->



画像列からアクション列の生成

- V. Mnih et al., "Human-level control through deep reinforcement learning", Nature, 2015
- Silver et al., "Mastering the game of Go with deep neural networks and tree search"
- 自動運転などへの応用にも期待



畳み込みニューラルネットワーク (Convolutional Neural Networks; CNNs)

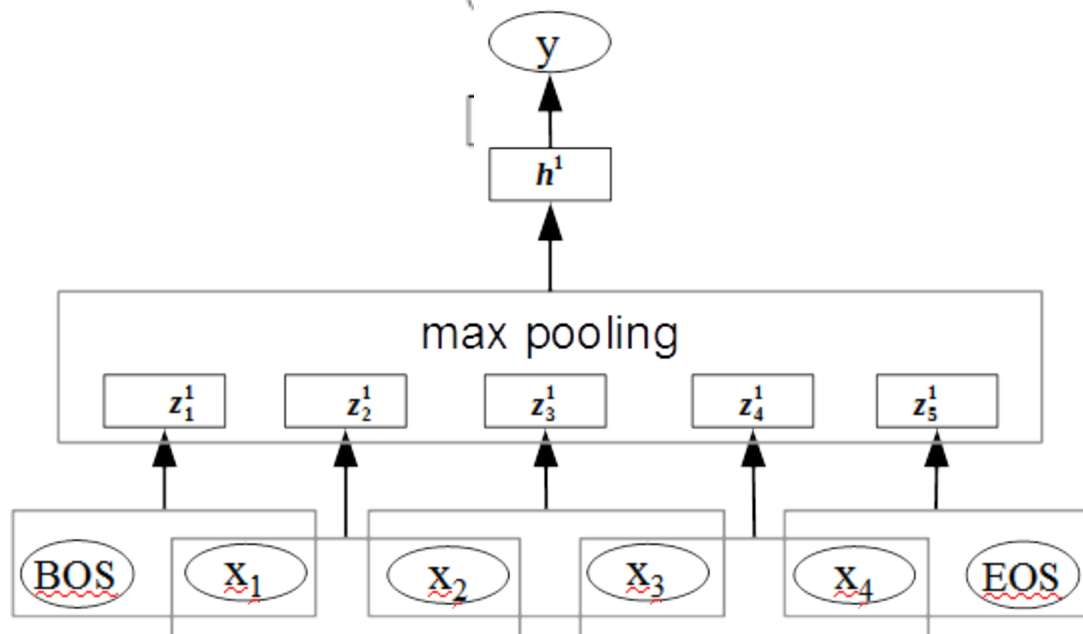
- 1次元畳み込み(窓幅 w)
- 最大値プーリングにより可変長入力を固定長に変換

Feature map

$$z_t^\ell = W^\ell \left[h_{t-\lfloor \frac{w}{2} \rfloor}^{\ell-1}; \dots; h_t^{\ell-1} \dots; h_{t+\lfloor \frac{w}{2} \rfloor}^{\ell-1} \right]$$

Kernel (t によらず共通)

$$h_i^\ell = \max_t f(z_{t,i}^\ell)$$



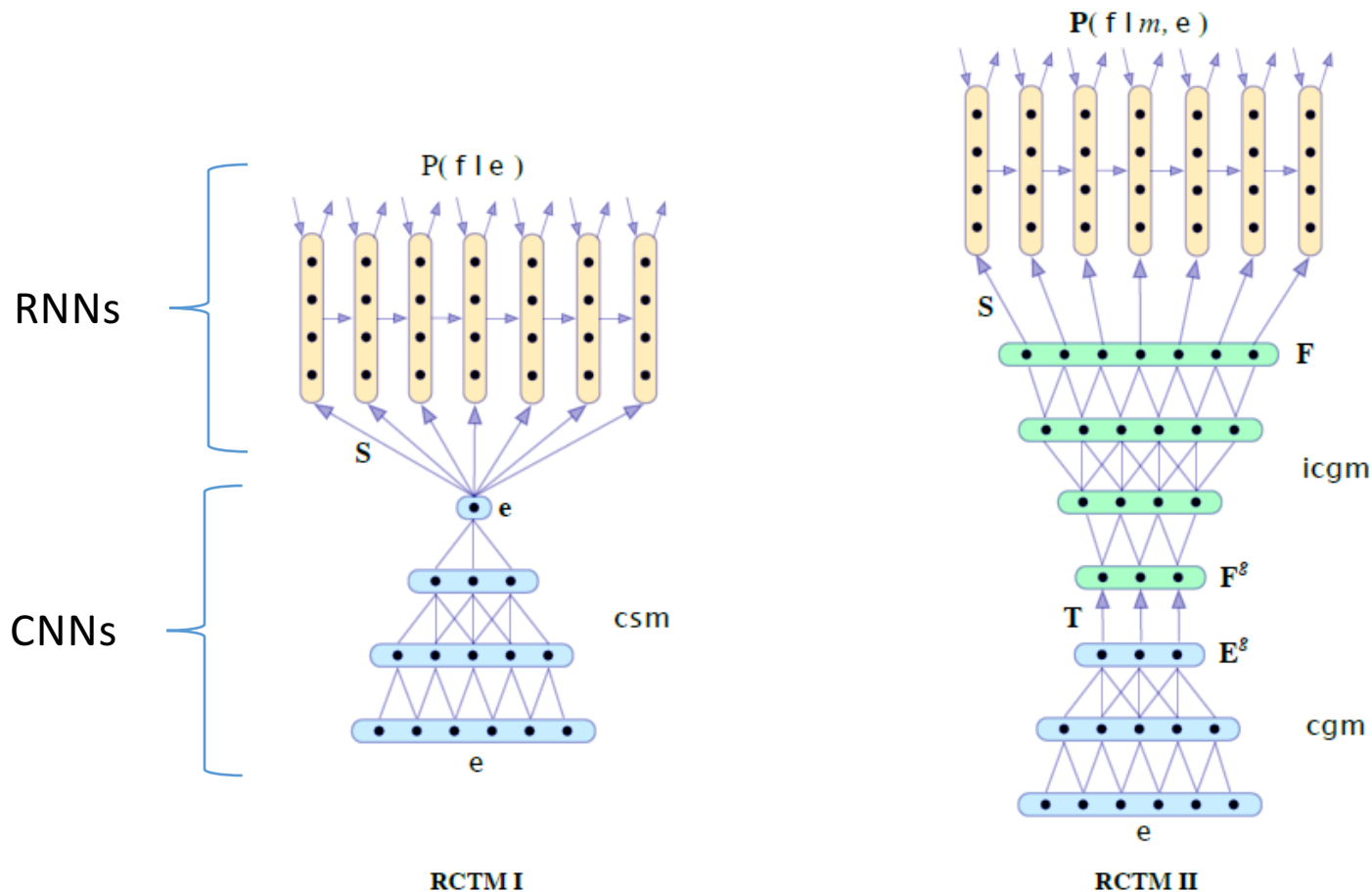
畳み込みニューラルネットワークの応用

- 基盤処理タスクをマルチタスク学習(品詞タグ付け, 句構造チャンキング, 固有表現抽出, 意味ラベル付与タスク) [Collobert et al., 2011]
 - 当時の最先端の性能に肉薄
- 文字単位でのCNN: 未知語に対応可能
 - 単語&文字CNN: 活用形が多い言語の処理やテキストに頑健 [Santos and Zadrozny, 2014] [Santos and Gatti, 2014]
 - Bag of 文字N-gram: 部分文字列でハッシング [Gao et al., 2014]
 - 文字CNN: 9層の深いネットワークを実現 [Zhang and LeCun, 2014]
- 動的k最大値プーリング [Kalchbrenner et al., 2014]
 - 上位k個のzを上位層に上げる。
 - Kは入力長Tに比例して決める
(仮定: 長い入力には情報量が多い)
$$k^\ell = \max \left(k^L, \left\lfloor \frac{L - \ell}{L} \right\rfloor T \right)$$
- 評判分析では最大値プーリングが、トピック分類では平均値プーリングが性能が高い [Johnson and Zhang, 2015]
 - タスクによって重要な部分が異なる: 一部 or 全体

CNNエンコーダとRNNエンコーダによる翻訳

[Nal and Blunsom, 2013]

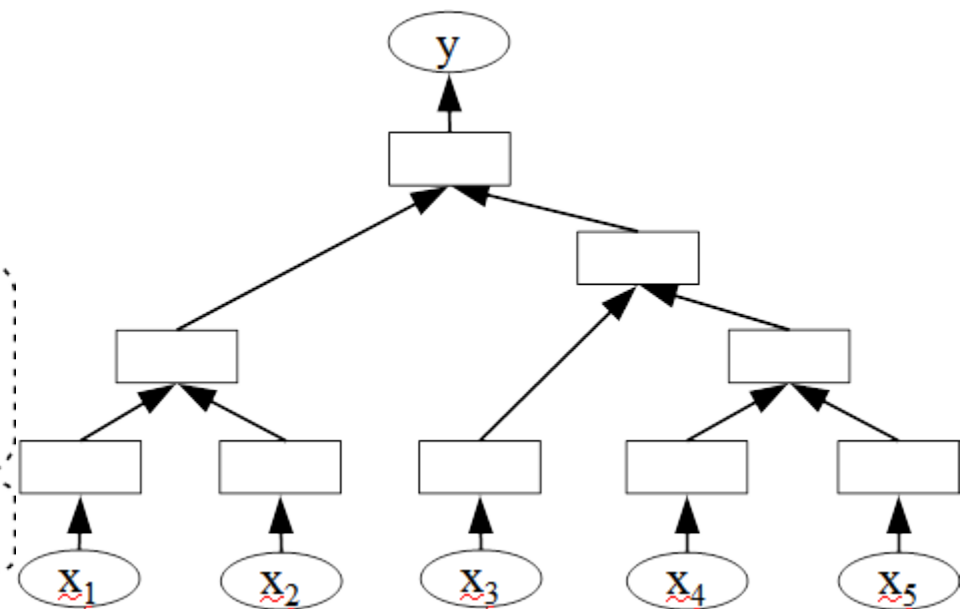
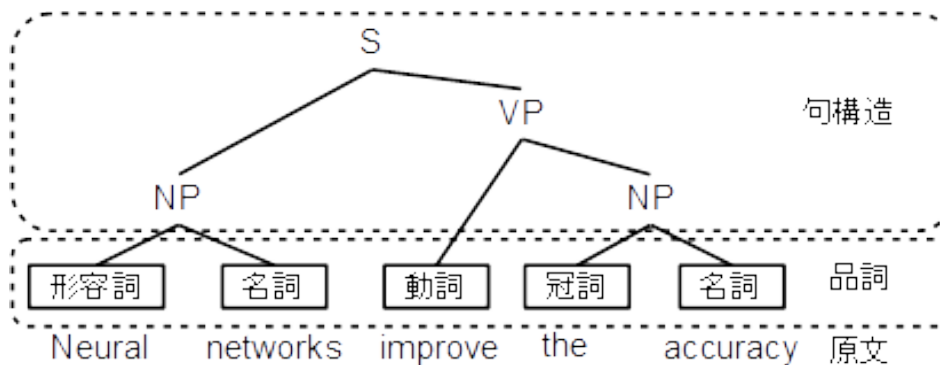
- Figures from the original paper



再帰ニューラルネットワーク (Recursive Neural Networks)

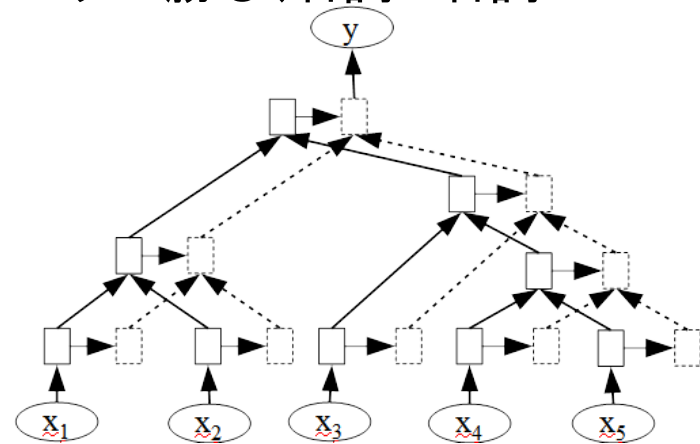
- RNNの一般化(Sequence \rightarrow DAG)
- 自然言語処理では構文解析結果の木構造を使い、文や句のベクトル表現を得る [Socher, 2014]
- 2分木を仮定すると:

$$h^\ell = f(W[h_L^{\ell-1}, h_R^{\ell-1}])$$



再帰ニューラルネットワークの応用

- 評判分析: 句のレベルで好評・不評を判定 [Socher et al., 2013]
- 質問応答: 質問文をベクトル表現し、該当する回答に分類 [Iyer et al., 2014]
- 長い依存関係が必要なタスクで有効 [Li et al., 2015]
 - 評判分析・質問応答・談話構造解析ではRNNsと差がない(または劣る)
 - 意味関係解析では再帰ニューラルネットワークが勝る(名詞と名詞の間の主語が重要なタスク)
- 空間的にも深い再帰ニューラルネットワーク [Irsoy and Cardie, 2014]



前半のまとめ

- 自然言語処理の特徴
 - 入力が離散
 - 入出力が可変
- ネットワーク構造による分類
 - フィードフォワードニューラルネットワーク: 線形モデルの置き換え
 - リカレントニューラルネットワーク: 可変長入出力が可能。
 - 畳み込みニューラルネットワーク: 文字単位の研究では先行
 - 再帰ニューラルネットワーク: 文法構造を活用できる
- 自然言語処理のパイプライン処理を置き換える可能性
 - End-to-endで学習できることが強み
 - さまざまな前処理(品詞タグ付・構文解析等)が不要になる?
- 自然言語処理以外の似た特徴を持つタスクにも適用可能になる可能性
 - 入力列が離散 (例: 商品の購買履歴)
 - 入出力長が可変 (例: アミノ酸配列)

参考文献

- Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate, 2014. arXiv:1409.0473.
- Jiwei Li, Dan Jurafsky and Eduard Hovy. When Are Tree Structures Necessary for Deep Learning of Representations, 2015. arXiv:1503.00185
- Jianfeng Gao, Patrick Pantel, Michael Gamon, Xiaodong He, Li Deng, Yelong Shen. Modeling Interestingness with Deep Neural Networks, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014.
- Rie Johnson and Tong Zhang. Effective Use of Word Order for Text Categorization with Convolutional Neural Networks, , In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics (NAACL)*, 2015.
- Yoshua Bengio, Réjean Ducharme, Pascal Vincent, and Christian Janvin. A neural probabilistic language model. *Journal of Machine Learning Research*, Vol. 3, No. 19, pp. 1137–1155, 2003.
- Danqi Chen and Christopher Manning. A fast and accurate dependency parser using neural networks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 740–750, 2014.

参考文献

- Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP)*, pp. 1724–1734, 2014.
- Ronan Collobert, Jason Weston, Léon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel P. Kuksa. Natural language processing (almost) from scratch. *Journal of Machine Learning Research*, Vol. 12, pp. 2493–2537, 2011.
- Jacob Devlin, Rabih Zbib, Zhongqiang Huang, Thomas Lamar, Richard Schwartz, and John Makhoul. Fast and robust neural network joint models for statistical machine translation. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1370–1380, 2014.
- Cicero Dos Santos and Bianca Zadrozny. Learning character-level representations for part-of-speech tagging. In *Proceedings of the International Conference on Machine Learning (ICML)*, pp. 1818–1826, 2014.
- Cicero Dos Santos and Maira Gatti. Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, pp. 69–78, 2014.
- Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.
- Ozan Irsoy and Claire Cardie. Deep recursive neural networks for compositionality in language. In *Advances in Neural Information Processing Systems (NIPS)*, pp. 2096–2104, 2014.

参考文献

- Mohit Iyyer, Jordan Boyd-Graber, Leonardo Claudino, Richard Socher, and Hal Daumé III. A neural network for factoid question answering over paragraphs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 633–644, 2014.
- Nal Kalchbrenner and Phil Blunsom. Recurrent continuous translation models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1700–1709, 2013.
- Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, 2014.
- Andrej Karpathy and Li Fei-Fei. Deep visual semantic alignments for generating image descriptions, 2014. arXiv:1412.2306.
- Yoon Kim. Convolutional neural networks for sentence classification. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1746–1751, 2014.
- Ji Ma, Yue Zhang, Tong Xiao, and Jingbo Zhu. Tagging the Web: Building a robust web tagger with neural network. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (ACL)*. The Association for Computer Linguistics, 2014.

参考文献

- Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernocký, and Sanjeev Khudanpur. Recurrent neural network based language model. In *Proceedings of the Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pp. 1045–1048, 2010.
- Tomas Mikolov, Wen tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pp. 746–751, 2013.
- Richard Socher. *Recursive Deep Learning for Natural Language Processing and Computer Vision*. PhD thesis, Stanford University, 2014.
- Richard Socher, Andrej Karpathy, Quoc V. Le, Christopher D. Manning, and Andrew Y. Ng. Grounded compositional semantics for finding and describing images with sentences. *Transactions of the Association for Computational Linguistics*, 2014.
- Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Chris Manning, Andrew Ng, and Chris Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the Conference on Empirical Methods on Natural Language Processing (EMNLP)*, pp. 1631–1642, 2013.
- Martin Sundermeyer, Tamer Alkhouli, Joern Wuebker, and Hermann Ney. Translation modeling with bidirectional recurrent neural networks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 14–25, 2014.

参考文献

- Ilya Sutskever, Oriol Vinyals, and Quoc V. V Le. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems (NIPS)*, pp.3104–3112. 2014.
- Wen tau Yih, Xiaodong He, and Christopher Meek. Semantic parsing for single-relation question answering. In *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*. Association for Computational Linguistics, 2014.
- Yuta Tsuboi. Neural networks leverage corpus-wide information for part-of-speech tagging. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 938–950, 2014.
- Oriol Vinyals, Lukasz Kaiser, Terry Koo, Slav Petrov, Ilya Sutskever, and Geoffrey Hinton. Grammar as a foreign language, 2014. arXiv:1412.7449
- Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator, 2014. arXiv:1411.4555.
- Daojian Zeng, Guangyou Zhou, and Jun Zhao. Relation classification via convolutional deep neural network. In *Proceedings of the International Conference on Computational Linguistics (COLING)*, 2014.

参考文献

- Vinyals and Le, A Neural Conversational Model. ICML Deep Learning Workshop. 2015.
- Alexander M. Rush, Sumit Chopra, Jason Weston. A Neural Attention Model for Sentence Summarization. In Proceedings of EMNLP. 2015
- Dzmitry Bahdanau, Kyunghyun Cho, Yoshua Bengio. Neural Machine Translation by Jointly Learning to Align and Translate. In Proceedings of ICLR. 2015.
- Li Yao, Atousa Torabi, Kyunghyun Cho, Nicolas Ballas, Christopher Pal, Hugo Larochelle, Aaron Courville. Describing Videos by Exploiting Temporal Structure. In Proceedings of ICCV. 2015.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Identity Mappings in Deep Residual Networks. 2016. arXiv:1603.05027
- Pegah Ghahremani, Jasha Droppo, and Michael L. Seltzer. Linearly Augmented Deep Neural Network. In Proceedings of ICASP 2016.

ICML 2015 ディープラーニングワークショップ パネルディスカッションより今後の展望

- <http://www.kyunghyuncho.me/home/blog/briefsummaryofthepaneldiscussionatdlworkshopicml2015> より抜粋
- 次に飛躍しそうなアプリケーション
 - Medicine/Healthcare (medical image analysis, drug discoveryなど)
 - Q&A, 対話, 連続意思決定, non-cognitive task (IoT?)
- 企業とアカデミア
 - 企業とアカデミアのギャップは大きい
 - GoogleやFacebookなどのデータドリブン企業以外はデータ不足に悩んでいる
- 如何に第2の冬の時代を避けるか
 - 査読やソフトウェアをオープンにし、課題点を明確にすることで過信を避ける
- AI脅威論
 - 気候変動や不平等拡散など他にもっと危急の課題がある
 - 人の知性ではなく、人の性格を真似ると脅威になる可能性がある