

# 自然言語処理における ディープラーニングの発展

Yuta Tsuboi

IBM Research Tokyo

yutat@jp.ibm.com

2015-03-16

# 出版予定のサーベイ論文の内容を元にお話します

- 坪井 祐太, 自然言語処理におけるディープラーニングの発展, オペレーションズ・リサーチ, Vol.60, No.4 (In press)

# 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-

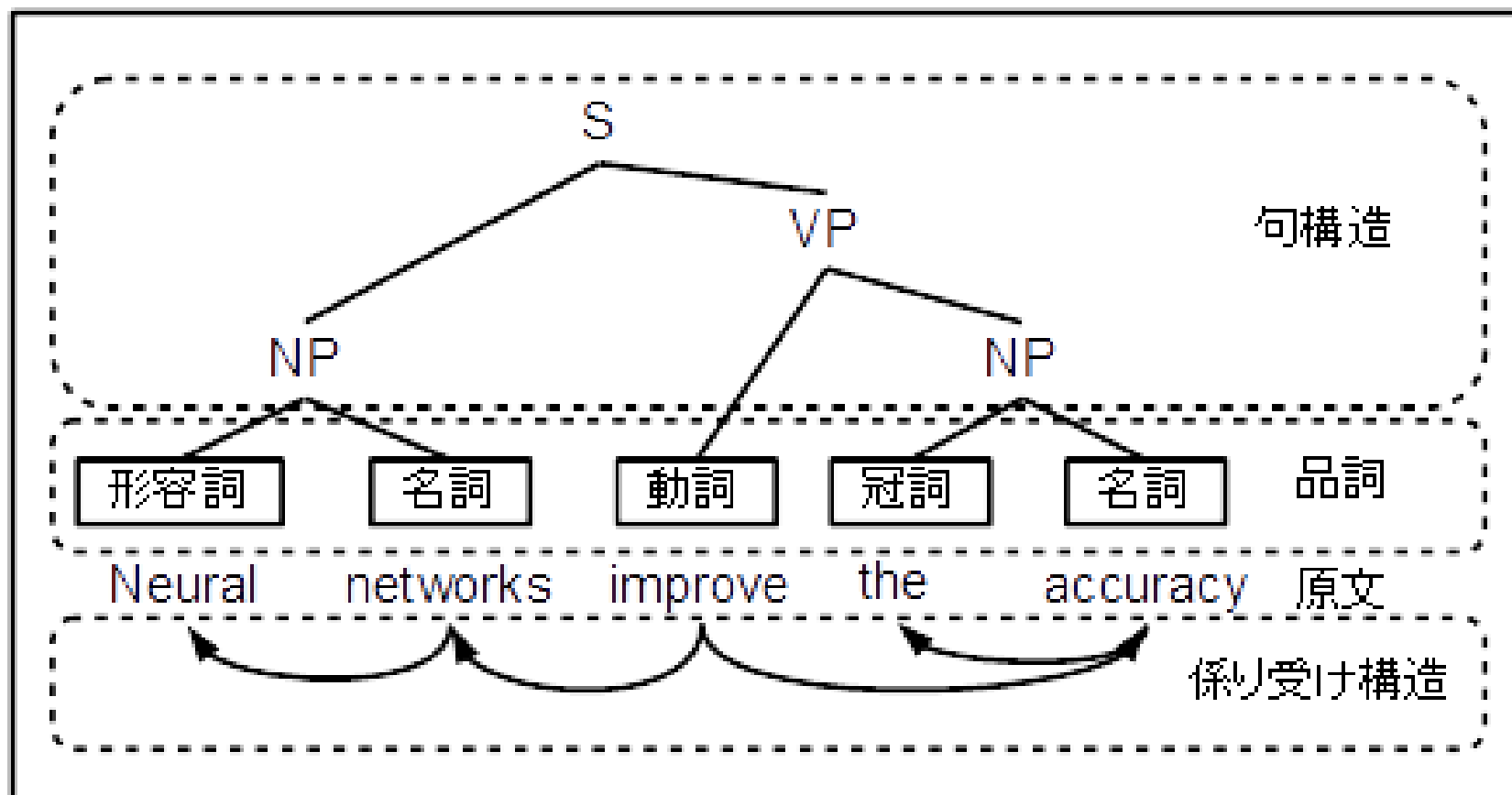
## • 離散入力

- 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

# 自然言語処理の例



↓ 翻訳

ニューラルネットワークが精度を改善する

# 自然言語処理における ディープラーニングの現状

- ニューラルネットワークを用いた手法の台頭
  - 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
- うまく行けば自然言語処理以外の似たような特徴を持つタスクでも有効な手法になる可能性
  - 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)

# 従来の自然言語処理での機械学習手法

- 線形分類器が中心
  - カーネル法が2000年前後に流行
  - 予測時の遅さのために近年ではあまり使われていない
    - ランダム射影によるカーネル法の高速化は？
- 構造予測
  - 列構造・木構造などの出力変数間の依存関係
  - 構造全体を当てる目的関数(条件付確率場, 構造化SVM, etc.)
  - 効率的な構造の推定・列挙(動的計画法, 整数計画, etc.)

# 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)

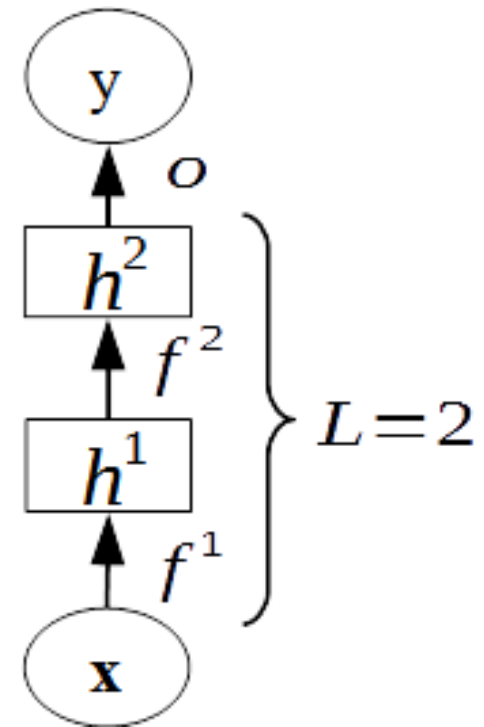
# フィードフォワードニューラルネットワーク (Feed-forward Neural Networks)

- 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)

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

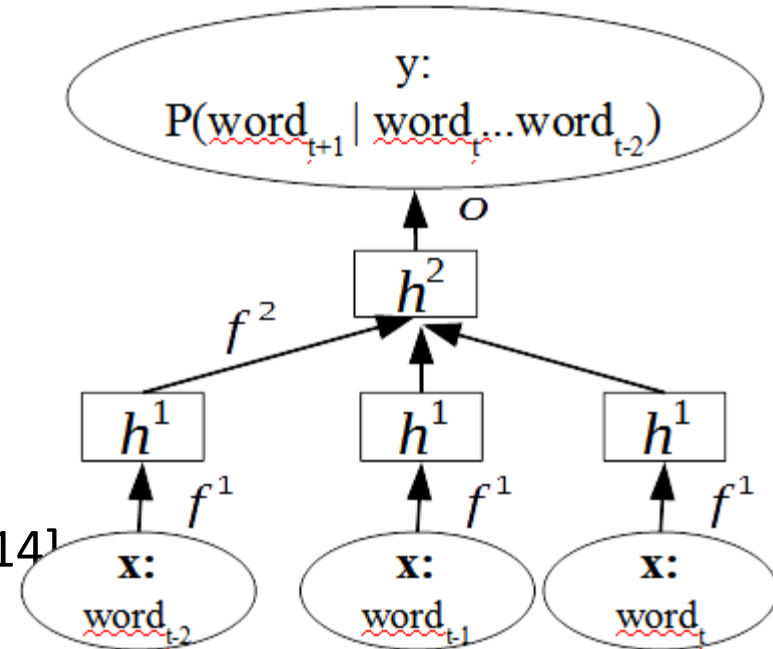
$$\hat{\mathbf{y}} = \mathbf{o}(\mathbf{W}^0 \mathbf{h}^L)$$





# フィードフォワードニューラルネットワークの応用例

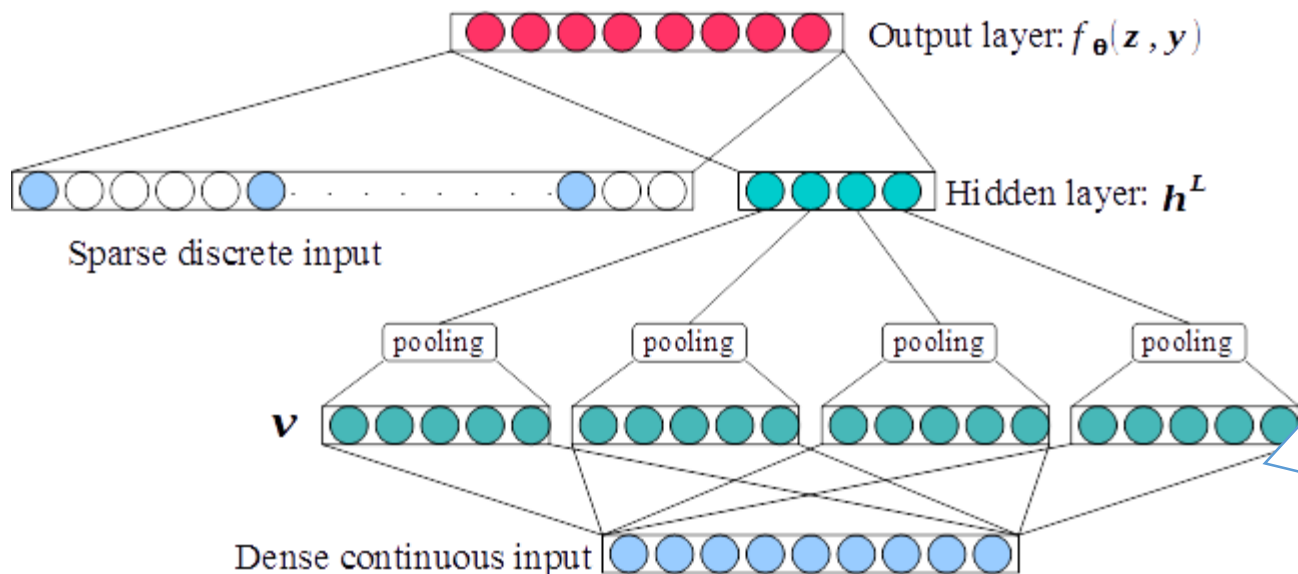
- 言語モデル [Bengio et al., 2003]
  - 次の単語を予測するモデル
  - 文の生成などに利用される
  - 1層目で窓幅の単語を個別に非線形変換
- 機械翻訳 [Devlin et al., 2014]
  - 原言語と対象言語両方を入力
- 品詞タグ付 [Ma et al., 2014] [Tsuboi, 2014]
- 構文解析 [Chen and Manning, 2014]



- 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

# フィードフォワードニューラルネットワークによる英語品詞タグ付 [Tsuboi, 2014] -手前味噌-

- 現時点での最高性能を実現
- 線形分類器とニューラルネットワークを統合したモデル
  - 線形分類器: 従来の離散疎ベクトル
  - ニューラルネットワーク: 単語分布など連続密ベクトル

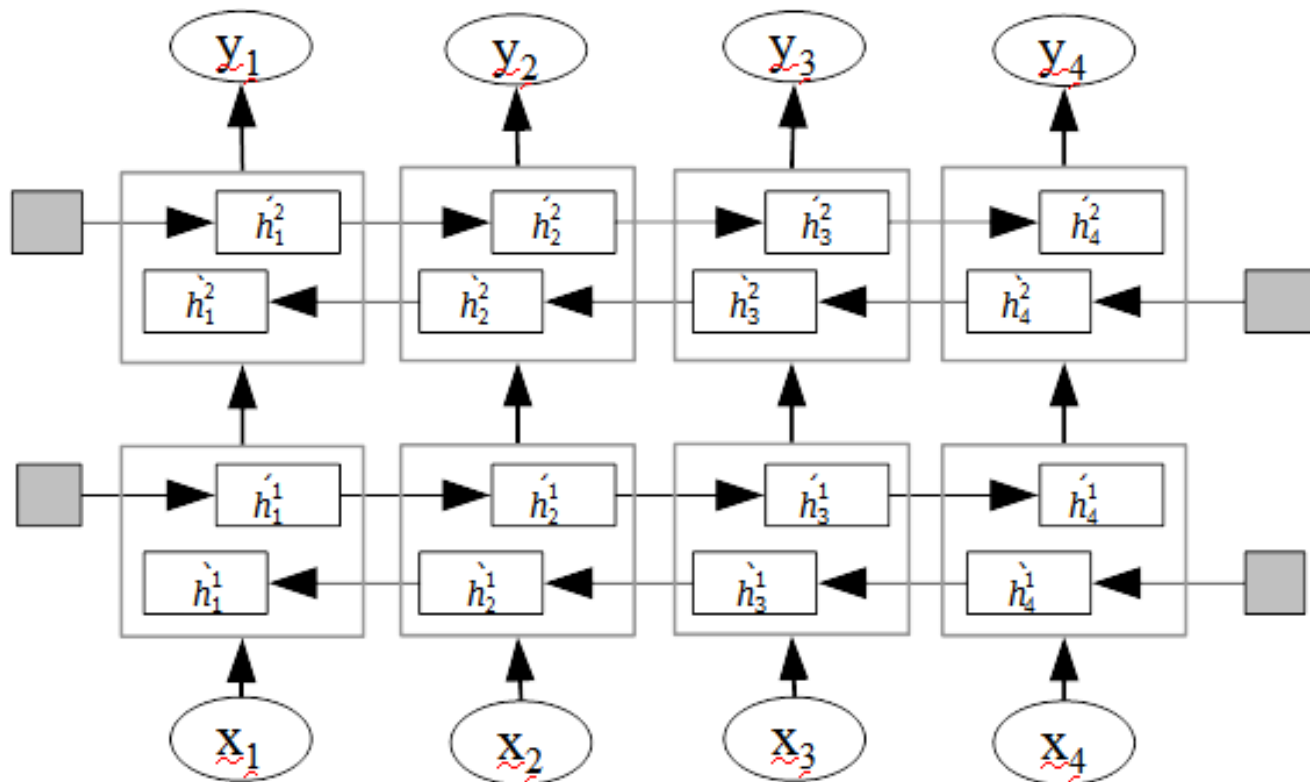


Lp-pooling (Gulcehre et al., 2014) やMaxout networks (Goodfellow et al., 2013) などプーリングタイプの活性化関数を使用

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

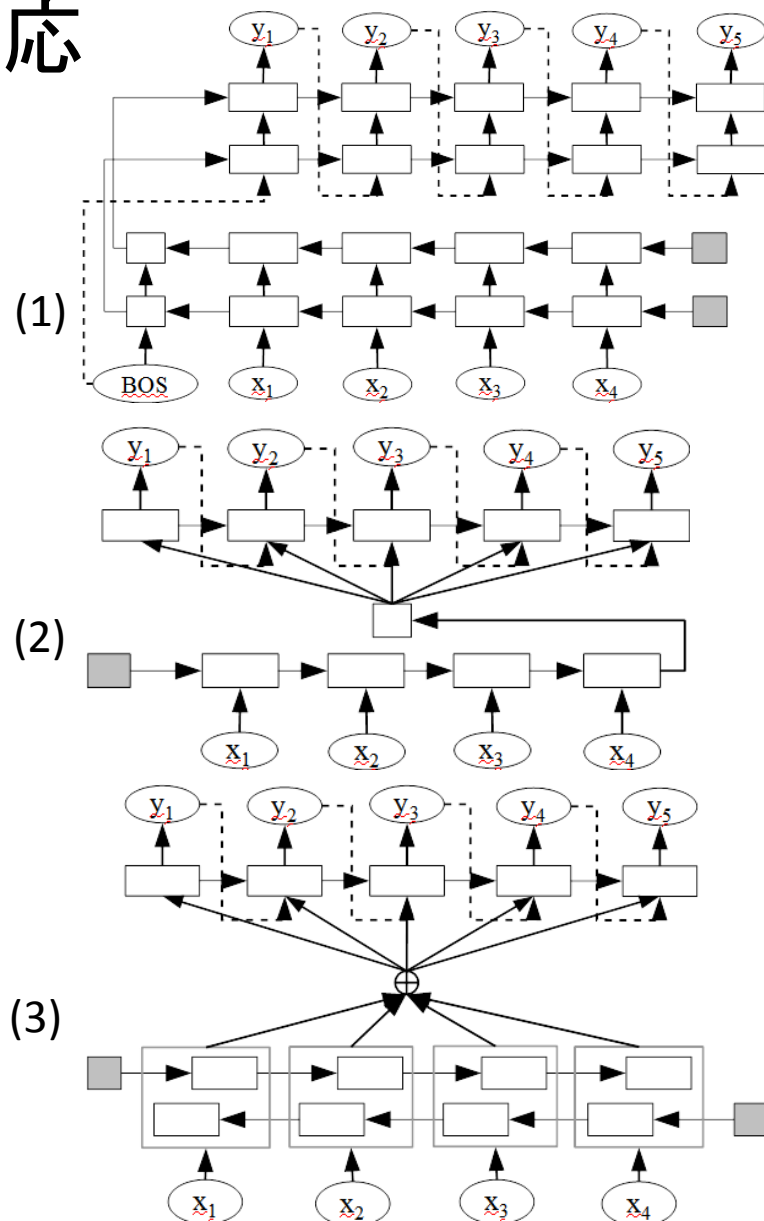
- 双方向RNNの例  $\hat{h}_t^\ell = f^\ell(W[\hat{h}_{t-1}^\ell; \hat{h}_t^{\ell-1}; \hat{h}_t^{\ell-1}]) \leftarrow$  後向き走査  
 $\hat{h}_t^\ell = f^\ell(W[\hat{h}_{t+1}^\ell; \hat{h}_t^{\ell-1}; \hat{h}_t^{\ell-1}]) \leftarrow$  前向き走査

- 課題: 入力と出力の長さが同じであることが必要



# 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

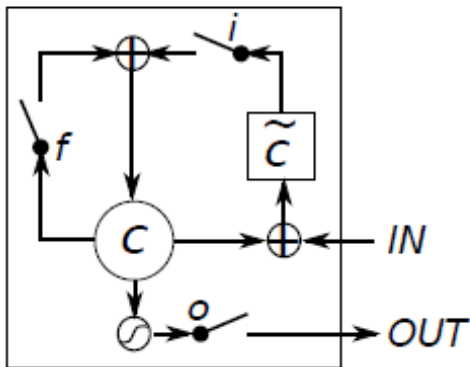
1. エンコーダの末尾状態をデコーダの先頭に接続  
[Sutskever et al., 2014]
2. エンコーダの末尾状態をデコーダのすべての点で参照  
[Cho et al., 2014]
3. エンコーダの各点の状態を重みつき線形和したベクトルをデコーダのすべての点で参照  
(a.k.a. ソフトアテンションモデル)  
[Bahdanau et al., 2014]



# 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

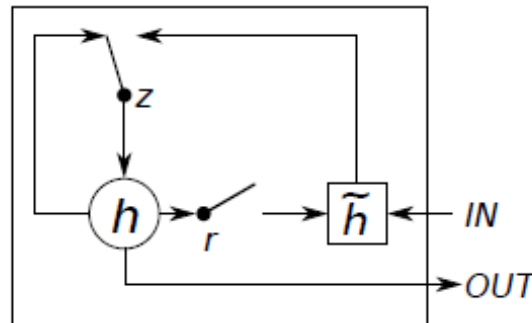
- Multiplicative Gate Units
  - Figures from (Chung et al., 2014)

- Figures from (Mikolov et al., 2015)



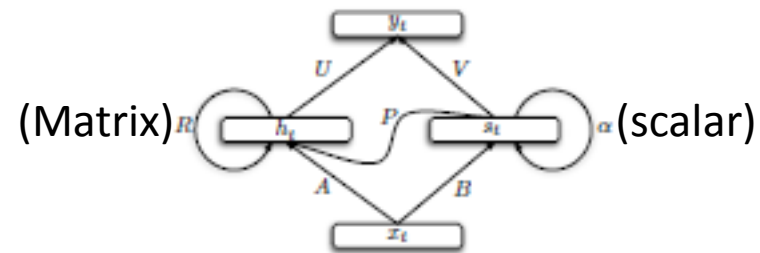
(a) Long Short-Term Memory

(Hochreiter and Schmidhuber, 1997)



(b) Gated Recurrent Unit

(Cho et al., 2014)



$$\begin{aligned}
 s_t &= (1 - \alpha) B x_t + \alpha s_{t-1}, \\
 h_t &= \sigma (P s_t + A x_t + R h_{t-1}), \\
 y_t &= f (U h_t + V s_t)
 \end{aligned}$$

structurally constrained recurrent nets (SCRN)

# 画像エンコーダ (CNN) とテキストデコーダ (RNN)

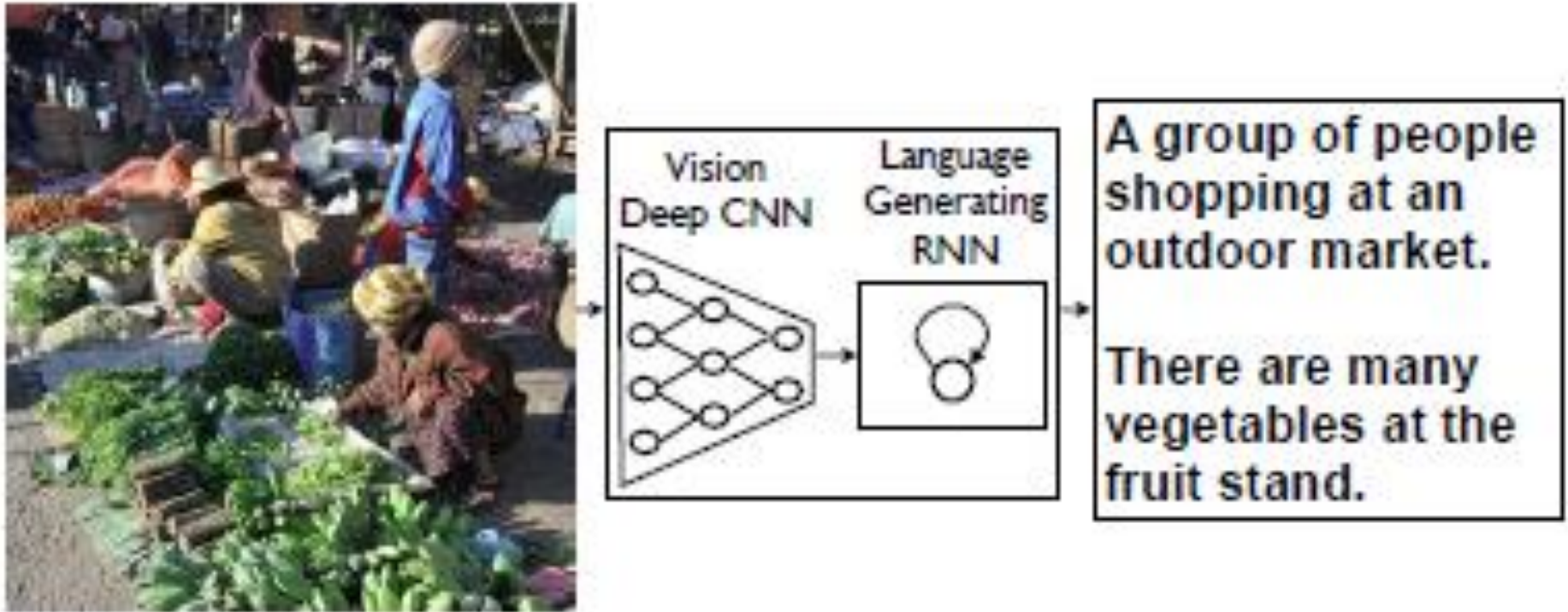


Figure from the original paper [Show and Tell: A Neural Image Caption Generator (Vinyals et. al., 2014)]

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

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

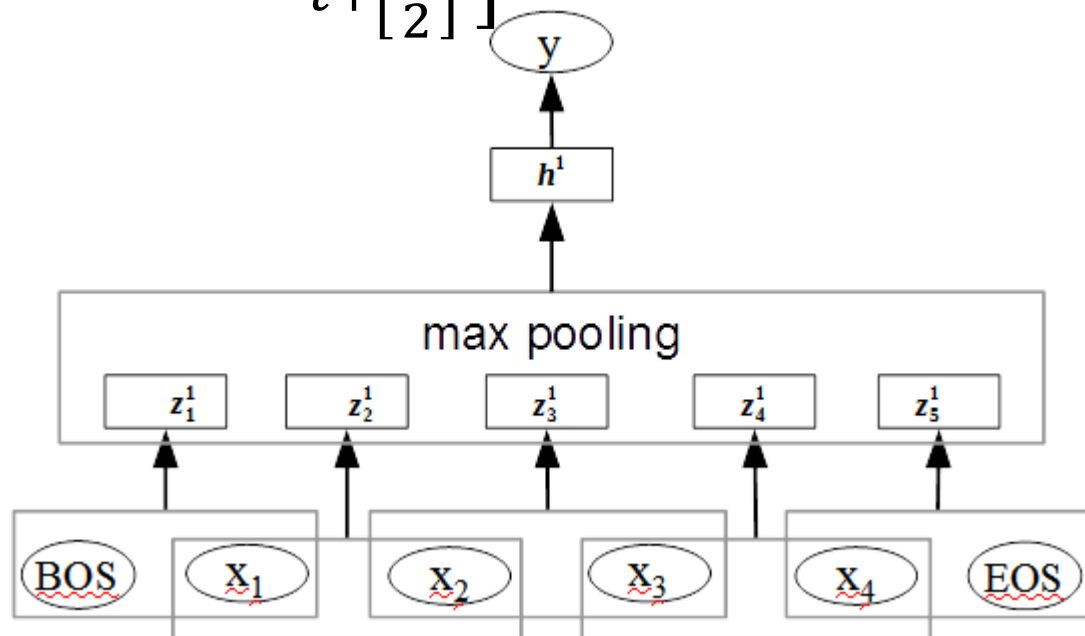
- 1次元畳み込み(窓幅 $w$ )
- 最大値プーリングにより可変長入力を固定長に変換

Feature map

$$\mathbf{z}_t^\ell = \mathbf{W}^\ell \left[ \mathbf{h}_{t-\lfloor \frac{w}{2} \rfloor}^{\ell-1}; \dots; \mathbf{h}_t^{\ell-1} \dots; \mathbf{h}_{t+\lfloor \frac{w}{2} \rfloor}^{\ell-1} \right]$$

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

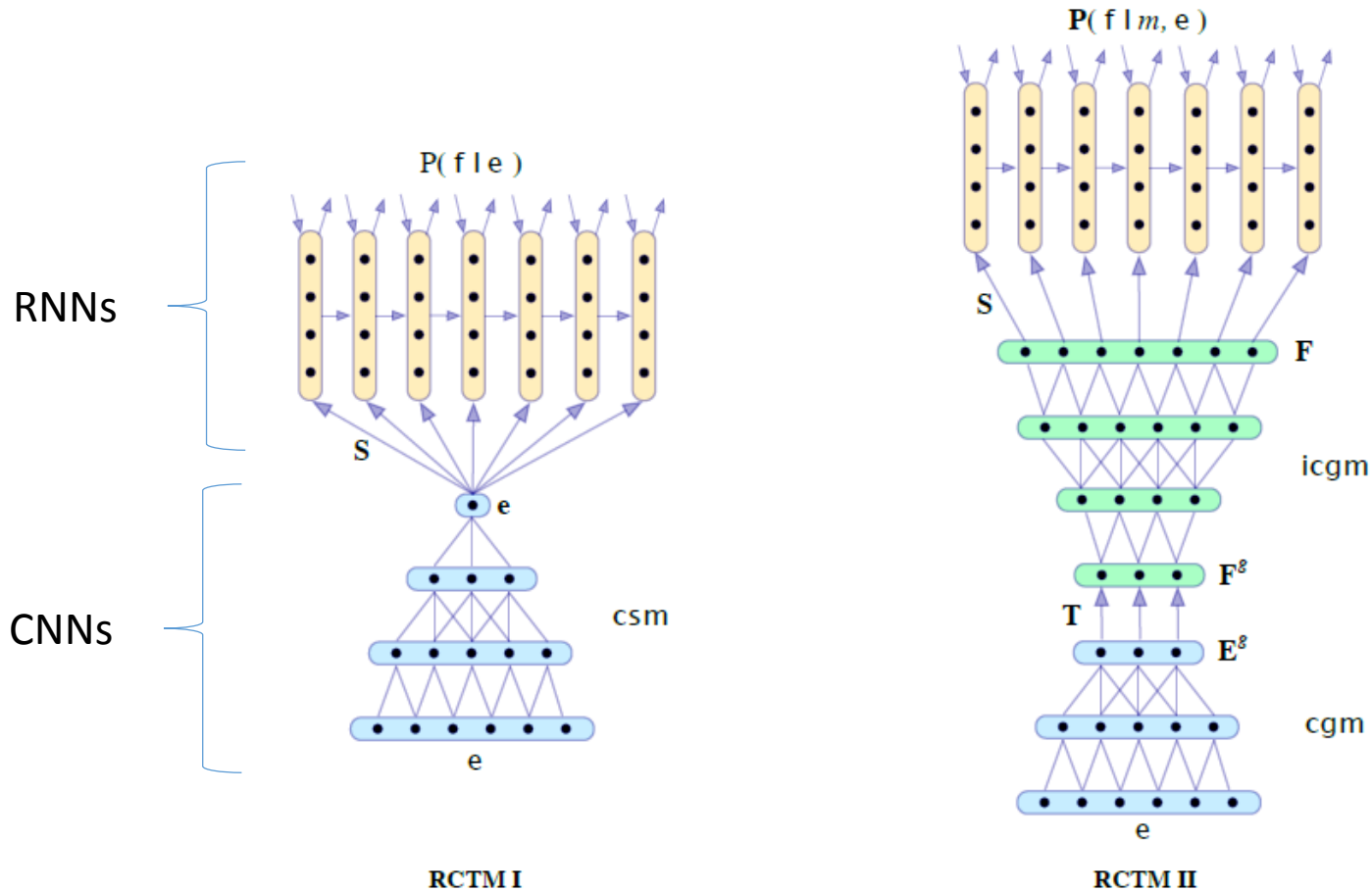
- 基盤処理タスクをマルチタスク学習(品詞タグ付け, 句構造チャンキング, 固有表現抽出, 意味ラベル付与タスク) [Collobert et al., 2011]
  - 当時の最先端の性能に肉薄
- 文字単位でのCNN: 未知語に対応可能
  - 単語&文字CNN: 活用形が多い言語の処理やテキストに頑健 [Santos and Zadrozny, 2014] [Santos and Gatti, 2014]
  - Bag of 文字N-gram: 部分文字列でハッシング [Gao et al., 2014]
  - 文字CNN: 9層の深いネットワークを実現 [Zhang and LeCun, 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]

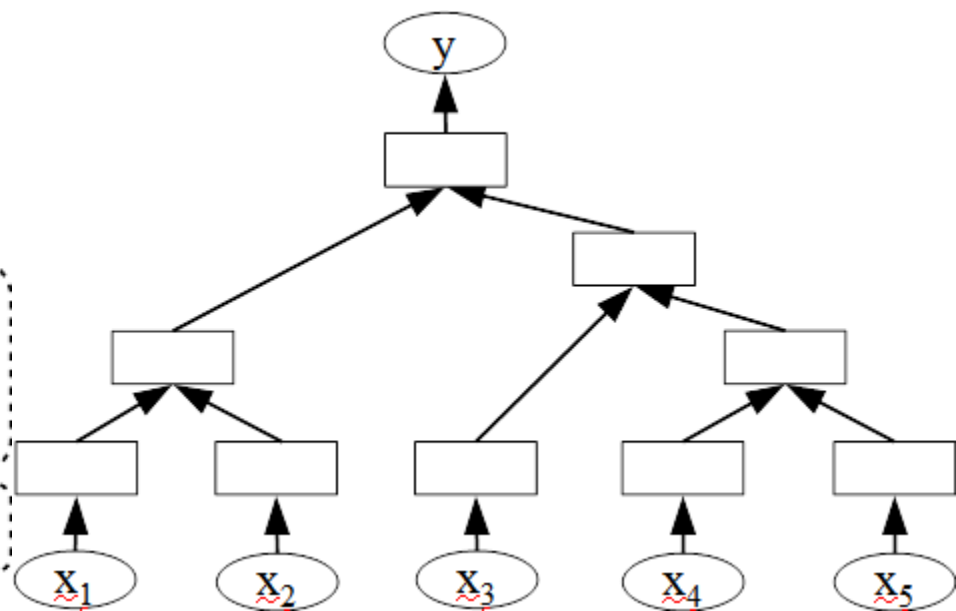
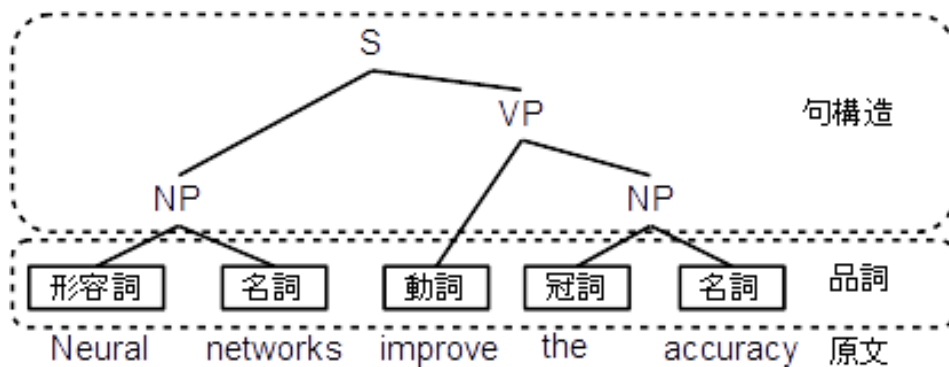
- Figures from the original paper



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

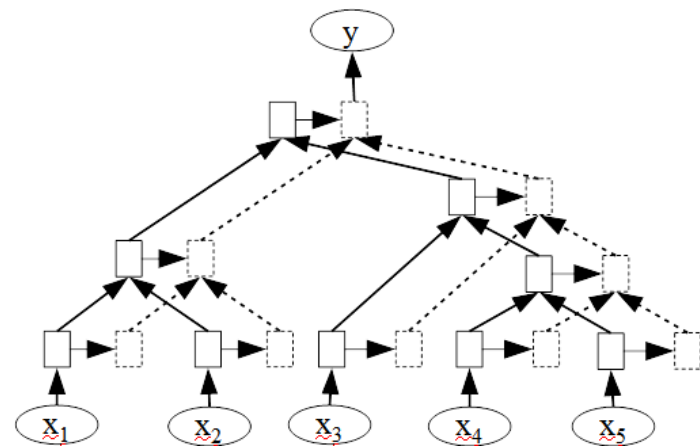
- RNNの一般化(Sequence  $\rightarrow$  DAG)
- 自然言語処理では構文解析結果の木構造を使い、文や句のベクトル表現を得る [Socher, 2014]
- 2分木を仮定すると:

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



# 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]



# 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。流行中
  - 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
  - さまざまな前処理(品詞タグ付・構文解析等)が不要になる?
- お話していないが重要なこと学習アルゴリズム/最適化
  - 自然言語処理特有のアルゴリズムはないようだがAdaGrad利用率が高い?

# 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, Lukáš 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.

# 日本語の参考文献

- Danushka Bollegala. 自然言語処理のための深層学習. 人工知能, Vol. 29, No. 2, pp. 195–201, 2014.
- 渡辺太郎. 深層学習による機械翻訳. 情報・システムソサイエティ, Vol. 19, No. 3, pp. 8–9, 2014.