

人工知能学事典 再校のお願い

次ページ以降の内容につきまして、著者校正をお願いいたします。

- 締切は **2017 年 2 月 21 日** です。
- 修正できるのは修正必須の誤りだけです。文章の推敲や追加はできません。
- 校正が終わりましたら、修正がない場合は「修正なし」、修正がある場合は修正内容をメールでお知らせください。
- メールでは伝わりにくい修正指示は、この PDF を印刷して赤字を入れ、スキャンまたは写真撮影の上、メールに添付してください。
- 共著の項目は、代表者がまとめてご回答ください。

ご連絡事項

- 再校では、aidic ウェブシステム上での初校の結果を反映した原稿データを、最終的な紙版の状態に近づけた組版状態でご覧いただきます（初校の状態を確認するには、再校通知メール（この PDF を開いたメール）の下部に記載したリンクを開いてください）。
- 再校通知メールに個別の連絡を記載している場合があります。再校通知メールの **■本項目固有のご連絡■** を必ずお読みください。
- 組版は、旧版の器に原稿データを流し込んで現段階でできる調整・補正を施した状態です。
- 後の工程で、ページデザインやフォントサイズを変更して総ページ数を圧縮するため、版面はこのあと変化します。配置・空きの不具合や、長い数式の飛び出しなど、**組版上の見た目の問題はその後修正しますので、個々のご指摘は不要です。**
- 内容的には、全項目の再校が済んだ後の全体処理（術語の統一など）を経て変化します。
- **索引** は出るべきものが出ているかを確認してください（**英語の大文字/小文字による索引重複は無視**）。aidic ウェブシステムと紙版とでポリシーが異なるため、**並び順や対訳・略語の表示方法などは後にまとめて調整・変更**します（稀に先行して直しているところもあります）。
- 項目タイトル部の項目番号は、現状どの項目も 1（例えば 15 章なら「15-1」）（コラムの場合、章番号も出ず「0-a」）と出ていますが問題ありません。一方、他項目参照の項目番号は、再校 PDF を生成した時点での aidic ウェブシステム上の項目ラベルになっています。以下に一覧があります。

http://www.gravel.co.jp/pdfs/aidic_itemlist.htm

- aidic ウェブシステムによる紙版向けの TeX データで、稀に数式や索引語句などの再現における事故が発生しています。別途共立出版にて不整合のチェックを行いますが、お気づきの点がありましたらお知らせください。

校正終了の通知やお問合せは、再校通知メール（この PDF を開いたメール）への返信を使い、下記宛てをお願いいたします。

株式会社グラベルロード

学術コンテンツ編集：伊藤裕之、山田ひとみ

aidic@gravel.co.jp

7-1

リカレントネットワークによる 文法学習

Learning Grammar by Recurrent Neural Network

はじめに

リカレントニューラルネットワークで文法を扱う研究は、エルマン [1][2][3] 以来の伝統がある。論文のタイトルである「小さく始めることが重要」とは、広義にカリキュラム学習 (curriculum learning) [4] の先駆けと見なせる。2010 年、Mikolov ら [5] は音声認識にリカレントニューラルネットワークを適用し、さらに自然言語処理へ適用可能であることを示した。これをリカレントニューラルネットワーク言語モデル (recurrent neural network language model; RNNLM) という。リカレントニューラルネットワーク言語モデルにおいては単語や品詞の選択という離散的意思決定問題になるので、隠れマルコフモデル (hidden Markov model; HMM) の最短パス探索や、条件付き確率場 (conditional random field; CRF) における最適解の探索に比肩しうる。単語配列の生起確率を最大化する条件付き確率場と観察単語系列の荷重和とを求め、損失を最小化する意思決定がなされる。この手続きをグラフ上の各ノードに対して繰り返し適用することで、連続値の場合の確率的勾配降下法 (stochastic gradient descent; SGD) に相当する。クヌース [6] が拡張文脈自由文法 (augmented context free grammar; ACFG) に関してコンパイラ設計で言及したコンピュータ言語に関するものでも、自然言語処理に関するものでも、リカレントニューラルネットワークを扱う手法が用いられている。加えて、リカレントニューラルネットワークを用いて数学の演算子を文法として扱う応用も提案されている [7]。ここでは、リカレントニューラルネットワークによる文法学習として、(1) 系列予測課題、(2) リカレントニューラルネットワーク、(3) 機械翻訳、(4) 構文解析を紹介する。

系列予測と系列生成

エルマンは系列予測学習課題を用いて単純再帰型ニューラルネットワークに逐次単語を入力し、その都度次の単語を予測させる訓練を行った [3]。系列予測

課題とは、時刻 $t-1$ におけるシステムの内部状態を $s(t-1)$ とし、入力を $x(t)$ とすれば条件付き確率

$$P(x(t+1) | x(t), s(t-1)) \quad (1)$$

となり、入出力 $(x_t, y_t)_{t \in T}$ を音素とすれば音韻予測、単語とすれば単語予測課題となる。

一方、入力を BOW (bag-of-word) などに固定し、文脈層の遷移によって出力系列を生成する課題を系列生成課題と呼ぶ。リカレントニューラルネットワークは系列予測および生成の能力を持つ。

単語予測課題：次の例文には主格となりうる名詞が複数存在する。

The girls who the teacher has picked for the play which will be produced next month practice every afternoon.

主語 “girls” と動詞 “practice” との間に中央埋め込み文 (center embedded sentence) が存在する。動詞との間で数の一致が必要であるため、リカレントニューラルネットワークはこの例文の主語 “girls” が複数形であることを保持し、三人称単数現在の s を付けないことを学習する。このとき、単純再帰型ニューラルネットワークは単文をまず学習し、続いて関係代名詞節を含む複雑な文章を学習することで増分学習 (incremental learning) が用いられる。すなわち、リカレントニューラルネットワークにおける文法学習とは、与えられた文章の構文解析木を返すという意味での文法を表すのではなく、時制、数、性の一致などの系列予測課題によって統語規則を正しく学習することを指していた。系列予測課題は、次の単語を予測することから、明示的に教師信号を用意しなくてもよいので、チョムスキーの指摘した刺激の貧困 (poverty of stimulus)、拡大して考えればプラトン問題 (Plato's problem) へのニューラルネットワークによる一つの解答だとも見なせる。

ニューラルネットワークを単語予測課題による言語モデルとして考えた場合、フィードフォワード型の Bengio [8] とリカレントニューラルネットワーク型の Mikolov ら [5] との両モデルを従来の比較参照モデルとして扱うことが多い。

リカレントニューラルネットワーク言語モデル

Mikolov ら [5] は、音声認識にエルマンネットワーク (単純再帰型ニューラルネットワーク) [1] を用い、

リカレントニューラルネットワークの自然言語処理への応用への道を開いた. x, h, y をそれぞれ中間, 文脈, 出力層の出力とする. 時刻 t における入力語彙を $v(t)$ とすれば,

$$x(t) = f(v(t) + h(t)) \quad (2)$$

$$h_i(t) = f\left(\sum_{j \in V, H} w_{ij} x_j(t-1)\right) \quad (3)$$

$$y_k(t) = g\left(\sum_{j \in H} w_{kj} x_j(t)\right) \quad (4)$$

ここで, f はロジスティック関数

$$f(x) = \frac{1}{1 + \exp(-x)} \quad (5)$$

であり, g はソフトマックス関数 (softmax function)

$$g(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (6)$$

である. また, w_{qp} は p から q への結合係数である.

Mikolov らは, ニューヨークタイムズ紙の 370 万語中 64 万語 (30 万文) を用いてニューラルネットワークを訓練した. 学習時間は要するが, ニーザーネイ (Kneser-Ney) の平滑化などの従来手法より訓練語彙数が少ないにもかかわらず, 単語誤認識率 (word error ratio) で比較した場合の性能が向上したことが報告された. Mikolov らは低頻度語を以下のように扱った.

$$p(v_i(t+1) | w(t), s(t-1)) = \begin{cases} \frac{y_{\text{rare}}(t)}{C_{\text{rare}}}, & w_i(t+1) \text{ が低頻度語の場合} \\ y_i(t), & \text{それ以外} \end{cases} \quad (7)$$

ここで, C_{rare} は閾値以下しか発生しない単語数である.

以降, リカレントニューラルネットワークに LSTM (long short-term memory) やゲート付き再帰ユニット (gated recurrent unit) を用いた研究が増加した.

ニューラルネットワーク機械翻訳

機械翻訳は Weaver [9] や Brown [10] 以来の歴史がある. ニューラルネットワーク機械翻訳 (neural network machine translation; NMT) と直接対比される概念に, SMT (statistical machine translation; 統計的機械翻訳) がある [10]. SMT で整備されてきた枠組みをリカレントニューラルネットワークに適用して, 機械翻訳を実装する試みがなされている. 画像処理と自然

言語処理とを同一の枠組みで記述する試み [11][12][13] もある. SMT では, 入力文 (ソース言語) x に対応する翻訳文 (ターゲット言語) y を条件付き確率 $p(y|x)$ と見なし, y の条件付き確率をベイズ則から次式のよう

$$p(y|x) \propto p(x|y)p(y) \quad (8)$$

上式 (8) を対数変換して

$$\log(y|x) = \log p(x|y) + \log p(y) + C \quad (9)$$

とおけば, 右辺第 1 項は翻訳モデル, 第 2 項は言語モデルとなる [14][15]. 式 (9) は BLEU [16] など为目标関数として学習される.

ニューラルネットワークを用いた機械翻訳 (NMT) [17] では, リカレントニューラルネットワークを用いてソース言語の文を目標言語の文の確率分布へと変換する. ソース言語と文脈層とを用いた

$$h_t = \sigma(W^{hx}x_t + W^{hh}h_{t-1}) \quad (10)$$

$$y_t = W^{yh}h_t \quad (11)$$

などが一般形であろう [18]. 機械翻訳では, ソース言語とターゲット言語における単語が一对一に対応するとは仮定できないため, 系列長も対応がとれるとは限らず, T と T' とを可変として $p(y_1, \dots, y_{T'} | x_1, \dots, x_T)$ を考える. LSTM を用いた言語モデル (LSTM-LM) (図 1) で考えれば,

$$p(y_1, \dots, y_{T'} | x_1, \dots, x_T) = \prod_{t \in T'} p(y_t | v, y_1, \dots, y_{t-1}) \quad (12)$$

となる. 上式 (12) 右辺 $p(y_t | v, y_1, \dots, y_{t-1})$ はソフトマックスを用いる. 文頭を <SOS> (start of sentence), 文末を <EOS> (end of sentence) という単語として扱う場合が多い.

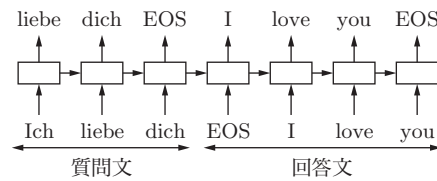


図 1 LSTM 言語モデル

この時点で最高性能のモデルの一つである Sutskever らの文献 [18] では, ソース言語とターゲット言語, すなわち入出力系列を異なる LSTM によってモデル化している. 加えて, ソース言語の語順を逆転させる (ターゲット言語は逆順にしない) と成績が向上したと報告している. なぜソース言語を逆順にして入力する

とよいかの理由については言及されていない。また、ニューラルネットワーク機械翻訳 (NMT) では、双方向リカレントニューラルネットワーク (bidirectional recurrent neural network; BRNN) [19] を用いる場合がある (図 2)。系列 ($\tau = 0, \dots, T$) における隠れ層の状態 h_τ を順方向 \vec{h}_τ と逆方向 \overleftarrow{h}_τ の二つの方向に情報が流れる二つの独立した中間層を用意する。両中間層を合併した $h_\tau = [\vec{h}_\tau, \overleftarrow{h}_\tau]$ を用いて、時刻 t における中間層の出力 a_t は、以下のように表記される。

$$a_t = f(s_t, h_t, y_{t-1}) \quad (13)$$

ここで、 s_t は入力、 y_{t-1} は 1 時刻前の出力である。

$$p(y_i | y_1, \dots, y_{i-1}, x, \theta) = g(y_{i-1}, s_i, c_i, \theta) \quad (14)$$

を考える。リカレントニューラルネットワークの内部状態 s は

$$s_i = f(s_{i-1}, y_{i-1}, c_i) \quad (15)$$

である。文脈ベクトル c_i は

$$c_i = \sum_{j \in T_s} \alpha_{ij} h_j \quad (16)$$

であり、 α_{ij} と BRNN の順行および逆行中間層の状態である h_j との積の総和である。各 BRNN ユニットの出力 $\alpha_{i \in T}$ は、ソフトマックスである。

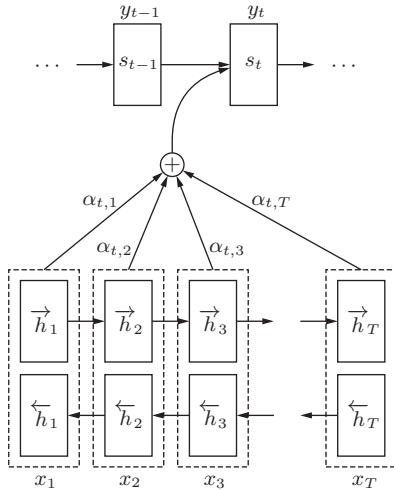


図 2 BRNN による NMT の概念図 (文献 [20], Figure 1 を改変)

ニューラルチューリングマシン (neural Turing machine) [21] の成果に基づいたニューラル翻訳機械 [22] と呼ばれるアプローチもある。これは、通常の深層リ

カレントニューラルネットワークが LSTM ユニットの多層に重ねるのに対して、入力層 (入力メモリ) → ヘッドから読み出し → コントローラ → ヘッドから書き出し → 次の入力層、というチューリングマシンの 1 ステップを多層化したモデルである。

Graves ら [21] は、LSTM、ニューラルチューリングマシン + LSTM、ニューラルチューリングマシン + フィードフォワード制御のモデルを比較し、ニューラルチューリングマシンに LSTM の制御をつけたモデルの学習が速いことを報告している [23]。

構文解析

リカレントニューラルネットワークによって構文解析木を出力するモデルがある [23]。日本語では “recurrent” と “recursive” をともに “再帰” と訳す。ニューラルネットワークの文脈では、“recursive” は同じ演算子やパラメータセットをスケールの異なる種々の状態、成分、因子、構造の計算に繰り返し適用する操作モデルを指す。一方、“recurrent” はフィードバック結合を有するニューラルネットワークモデルを指す場合に用いる。ここでは、recursive なモデルを RecNN, recurrent なモデルをリカレントニューラルネットワークと表記する。Socher ら [11] は、RecNN による構文解析を提案した。RecNN は、語や句、節など文の要素を入力として受け取り、それらの関係を出力とする。RecNN は有向非環グラフ (directed acyclic graph; DAG) であるから、構文解析木との相性は高い。図 3 は RecNN の例である。RecNN は同一の係数行列をすべての端点ノードに適用する。葉ノードは単語を表す n 次元ベクトル表現である。

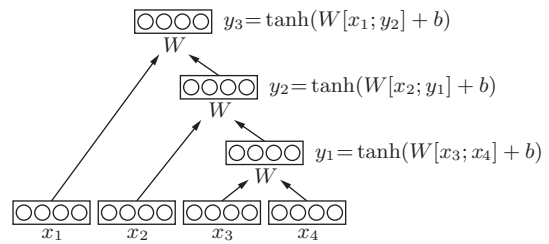


図 3 リカレントニューラルネットワーク

$$p(B|A) = \prod_{t \in T_B} p(y_{B_t} | A_1, \dots, A_{T_A}, B_1, \dots, B_{t-1}) \quad (17)$$

$$\equiv \prod_{t \in T_B} z(W_0 h_{T_{A+t}})' \delta_{B_t} \quad (18)$$

ここで, $x = (A_1, \dots, A_{T_A}, B_1, \dots, B_{T_B})$ であり, h_t は LSTM の h 空間上で t 番目の要素である. z はソフトマックスを表す. W_0 は各記号の表象行列であり, δ はクロネッカーのデルタである. 図 4 に RecNN による解析木の例を示す.

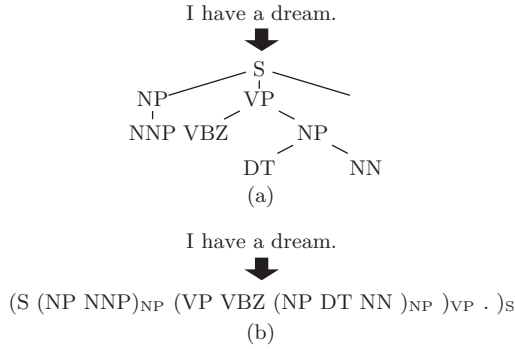


図 4 RecNN による構文解析の例. (a) 構文解析木, (b) 対応する文字列による出力 (文献 [23], Figure 2 を改変)

注意の導入

画像処理や自然言語処理において注意を導入する場合がある. Vinyals ら [23] も, Bahdanau ら [20] と同じく注意機構を導入した. 注意は次式のように表記できる.

$$u_i^t = v^T \phi(W_1' h_i + W_2' d_t) \quad (19)$$

$$a_i^t = z(u_i^t) \quad (20)$$

$$d_t' = \sum_{i \in T_A} a_i^t h_i \quad (21)$$

v , W_1' , W_2' は学習可能なパラメータである. u_i^t は長さ T のベクトルであり, これをソフトマックス z にかけて中間層の出力となる. ソフトマックスをかけることを注意と呼んでいる.

参考文献

- [1] Elman, J. L. Finding structure in time. *Cognitive Science*, Vol. 14, No. 2, pp. 179–211, 1990.
- [2] Elman, J. L. Incremental learning, or The importance of starting small. Technical report, University of California, San Diego, 1991.
- [3] Elman, J. L. Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning*, Vol. 7, pp. 195–225, 1991.
- [4] Bengio, Y., Louradour, J., Collobert, R., and Weston, J. Curriculum Learning. In *Proc. the 26th Annual International Conference on Machine Learning (ICML '09)*, pp. 41–48, 2009.

- [5] Mikolov, T., Karafiát, M., Burget, L., Černocký, J., and Khudanpur, S. Recurrent Neural Network Based Language Model. In *Proc. INTER-SPEECH2010*, pp. 1045–1048, 2010.
- [6] Knuth, D. E. Semantics of Context-Free Languages. *Mathematical Systems Theory*, Vol. 2, No. 2, pp. 127–145, 1968.
- [7] Zaremba, W., Kurach, K., and Fergus, R. Learning to Discover Efficient Mathematical Identities. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., and Weinberger, K., editors, *Advances in Neural Information Processing Systems*, pp. 1278–1286, 2014.
- [8] Bengio, Y., Ducharme, R., Vincent, P., and Jauvin, C. A Neural Probabilistic Language Model. *Journal of Machine Learning Research*, Vol. 3, pp. 1137–1155, 2003.
- [9] Weaver, W. Translation. In *Machine Translation of Languages*, pp. 15–23. MIT Press, Cambridge, Massachusetts, USA, 1949.
- [10] Brown, P. E., Pietra, S. A. D., Pietra, V. J. D., and Mercer, R. L. The Mathematics of Statistical Machine Translation: Parameter Estimation. *Computational Linguistics*, Vol. 19, No. 2, pp. 263–311, 1993.
- [11] Socher, R., Manning, C., and Ng, A. Learning Continuous Phrase Representations and Syntactic Parsing with Recursive Neural Networks. In *Proc. Advances in Neural Information Processing Systems*, 2010.
- [12] Socher, R., Lin, C. C., Ng, A. Y., and Manning, C. D. Parsing Natural Scenes and Natural Language with Recursive Neural Networks. In Getoor, L. and Scheffer, T., editors, *Proc. the 28th Annual International Conference on Machine Learning (ICML)*, 2011.
- [13] Socher, R., Bauer, J., Manning, C. D., and Ng, A. Y. Parsing With Compositional Vector Grammars. In *Proc. the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2013.
- [14] Cho, K., van Merriënboer, B., Gulcehre, C., Bougares, F., Schwenk, H., and Bengio, Y. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. In *Proc. the Empirical Methods in Natural Language Processing (EMNLP)*, 2014.
- [15] Koehn, P. Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proc. the Tenth Machine Translation Summit*, pp. 79–86, 2005.
- [16] Papineni, K., Roukos, S., Ward, T., and Zhu, W. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proc. the 40th Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 311–318, 2002.

- [17] Kalchbrenner, N. and Blunsom, P. Recurrent Continuous Translation Models. In *Proc. the 2013 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2013.
- [18] Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to Sequence Learning with Neural Networks. In *Proc. Advances in Neural Information Processing Systems (NIPS)*, pp. 3104–3112, 2014.
- [19] Schuster, M. and Paliwal, K. K. Bidirectional Recurrent Neural Networks. *IEEE Transactions on Signal Processing*, Vol. 45, No. 11, pp. 2673–2681, 1997.
- [20] Bahdanau, D., Cho, K., and Bengio, Y. Neural Machine Translation by Jointly Learning to Align and Translate. In *Proc. the International Conference on Learning Representations (ICLR)*, 2015.
- [21] Graves, A., Wayne, G., and Danihelka, I. Neural Turing Machines. *arXiv:1410.5401*, Vol. 1410.5401, 2014.
- [22] Meng, F., Lu, Z., Tu, Z., Li, H., and Liu, Q. Neural Transformation Machine: A New Architecture for Sequence-to-Sequence Learning. *arXiv:1506.06442*, 2015.
- [23] Vinyals, O., Kaiser, L., Koo, T., Petrov, S., Sutskever, I., and Hinton, G. Grammar as a Foreign Language. In *Proc. the International Conference on Learning Representations (ICLR)*, 2015.

執筆者：浅川伸一

和英索引

【か】

拡張文脈自由文法 (augmented context free grammar; ACFG) 2

隠れマルコフモデル (hidden Markov model; HMM) 2

カリキュラム学習 (curriculum learning) 2

【し】

刺激の貧困 (poverty of stimulus) 2

条件付き確率場 (conditional random field; CRF) 2

【そ】

双方向リカレントニューラルネットワーク (bidirectional recurrent neural network; BRNN) 4

ソフトマックス関数 (softmax function) 3

【ち】

中央埋め込み文 (center embedded sentence) 2

【に】

ニューラルチューリングマシン (neural Turing machine) 4

【ふ】

プラトン問題 (Plato's problem) 2

【り】

リカレントニューラルネットワーク言語モデル (recurrent neural network language model; RNNLM) 2

英和索引

【A】

augmented context free grammar; ACFG (拡張文
脈自由文法) 2

【B】

bidirectional recurrent neural network; BRNN (双
方向リカレントニューラルネットワーク)
4

【C】

center embedded sentence (中央埋め込み文) 2
conditional random field; CRF (条件付き確率場)
2
curriculum learning (カリキュラム学習) 2

【H】

hidden Markov model; HMM (隠れマルコフモデル)
2

【N】

neural Turing machine (ニューラルチューリング
マシン) 4

【P】

Plato's problem (プラトン問題) 2
poverty of stimulus (刺激の貧困) 2

【R】

recurrent neural network language model;
RNNLM (リカレントニューラルネッ
トワーク言語モデル) 2

【S】

softmax function (ソフトマックス関数) 3