

ニューラルネットワーク日英機械翻訳における品詞情報の利用

増田 嵩志 鶴岡 慶雅

東京大学 工学部電子情報工学科

{masuda, tsuruoka}@logos.t.u-tokyo.ac.jp

1 はじめに

対訳コーパスから翻訳ルールを統計的に学習する統計的機械翻訳には数多くの手法が存在する。フレーズベース機械翻訳や木構造ベース機械翻訳といった従来の手法では、確率モデルを翻訳モデルと言語モデルに分けて学習を行っていた [1]。しかし近年ニューラルネットワークのみを用いて入力文から出力文を生成する確率を直接学習するモデルが考案され注目されている [2]。このニューラルネットワーク機械翻訳は英語・フランス語間といった西洋言語間の翻訳で従来の手法を上回る精度を達成している [3]。基本的な Encoder-Decoder モデルに加え、アライメントの存在を仮定し翻訳に関係する部分に着目しながら翻訳を行う注意型ニューラルネットワーク機械翻訳モデル [4] などの応用モデルも考案されている。

しかしニューラルネットワーク機械翻訳におけるエンコード側もしくはデコード側での特徴量の考慮は、注意型ニューラルネットワーク機械翻訳モデルでのアライメントの考慮以外には特に行っていないのが現状である。それに対し、統計的機械翻訳だけでなく他のタスクにおいて品詞情報を特徴量として用いることで翻訳精度が向上したとの報告がある [5]。品詞情報を用いて品詞曖昧性を解消することで、多義語や、同じ綴りであるが品詞や役割の違いを持つ単語の区別化が可能になる。以上から品詞情報の利用はニューラルネットワーク機械翻訳の精度向上に有効に働くと考えられる。

そこで本稿では、品詞ラベルを付加した対訳コーパスを用いて学習を行うことで、ニューラルネットワーク機械翻訳の精度向上を目指す。結果として品詞情報を付加した対訳コーパスを用いることで翻訳精度が下がってしまったが、同時に今後の課題や方針が見つかった。

2 ニューラルネットワーク機械翻訳

統計的機械翻訳は、翻訳元言語文 x が与えられた時に、条件付き確率 $P(y|x)$ が最大となる翻訳先言語文 y を求めるタスクとしてされる。このため対訳コーパスを用いて適切な文のペアで条件付き確率が最大になるようにパラメータを調整する必要がある。従来では、言語モデルと翻訳モデルに分けて条件付き確率のモデルを学習する手法が行われていたが、近年ではニューラルネットワークを用いることで、条件付き確率全体のモデルを直接学習する手法が提案された [2]。このニューラルネットワーク機械翻訳では、基本的に2つの Recurrent Neural Network (RNN) から構成されるモデルを用いて翻訳を行う。

2.1 Encoder-Decoder モデル

Cho らや Sutskever らが提案したモデルは2つの RNN から構成されるモデルで翻訳を行っている [6, 3]。これらのモデルでは、一方の RNN で可変長の翻訳元言語文を入力として固定長ベクトルにエンコードし、もう一方の RNN でベクトルから可変長の翻訳先言語文を出力している (Encoder-Decoder モデル)。Cho らが考案したモデル [6] の概略を図 1 に示す。以下ではこのモデルの基本的な仕組みを説明する。

エンコード

翻訳元言語の単語列 x_1, x_2, \dots, x_m に対応する固定長のベクトル $s_1, s_2, \dots, s_m (\in \mathbb{R}^{d_e \times 1})$ を単語ベクトルとする。単語ベクトルの系列を順番に RNN に入力することで、文の情報をエンコードした固定長の文ベクトルとして、隠れ層ベクトル $h_k (\in \mathbb{R}^{d_h \times 1})$ が生成される。 h_k は重み行列 $W_1 (\in \mathbb{R}^{d_h \times d_e}), W_2 (\in \mathbb{R}^{d_h \times d_e})$ を用いて式 (1) より算出される。なお f は非線形活性化関数であり、シグモイド関数や双曲線正接関数などが用いられるが、長期依存の学習効率を向上させるため

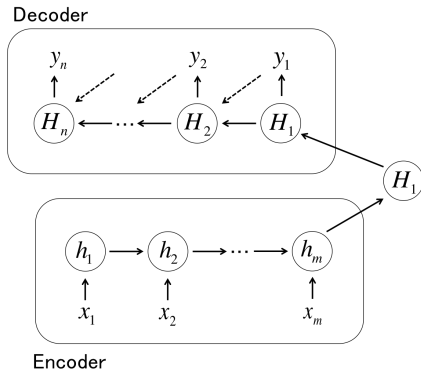


図 1: Encoder-Decoder モデルの概略

Sutskever らのように f に Long Short-Term Memory (LSTM) [7] を用いる例もある [4]。

$$\mathbf{h}_k = f(W_1 \mathbf{h}_{k-1} + W_2 \mathbf{s}_k) \quad (1)$$

そうして得られた翻訳元言語の最終的な文ベクトル \mathbf{h}_m に重み行列 $W_3 (\in \mathbb{R}^{d_h \times d_h})$ を掛け合わせることで、翻訳先言語の文ベクトル $\mathbf{H}_1 = W_3 \mathbf{h}_m (\in \mathbb{R}^{d_h \times 1})$ を生成する。

デコード

翻訳先言語の各単語 y_i に対して単語ベクトル $\mathbf{t}_i (\in \mathbb{R}^{d_e \times 1})$ 及び重みベクトル $\mathbf{g}_i (\in \mathbb{R}^{d_h \times 1})$ を与えておく。翻訳先言語の文ベクトル \mathbf{H}_1 から、式 (2) よりソフトマックス関数によって出力単語を生成する。式 (2) 中の \mathbf{H}_n は式 (3) によって、重み行列 $W_4 (\in \mathbb{R}^{d_h \times d_h})$, $W_5 (\in \mathbb{R}^{d_h \times d_e})$ を用いることで単語を出力する度に算出される。

$$y_l = \arg \max_i \frac{\exp(\mathbf{g}_i \cdot \mathbf{H}_l)}{\sum_j \exp(\mathbf{g}_j \cdot \mathbf{H}_l)} \quad (2)$$

$$\mathbf{H}_l = f(W_4 \mathbf{H}_{l-1} + W_5 \mathbf{t}_{l-1}) \quad (3)$$

2.2 注意型ニューラルネットワーク機械翻訳

基本的な Encoder-Decoder モデルで翻訳を行うことを考えた場合、翻訳元言語文の全情報を固定次元のベクトルで表現する必要があることになる。そのため文の長さが長くなるにつれ翻訳の精度は急速に落ちてしまう。この欠点を克服するため、デコーダーによる出力単語の各生成時に、翻訳元言語文の中でどの部分が最も関連する箇所なのか (アライメント) を探索す

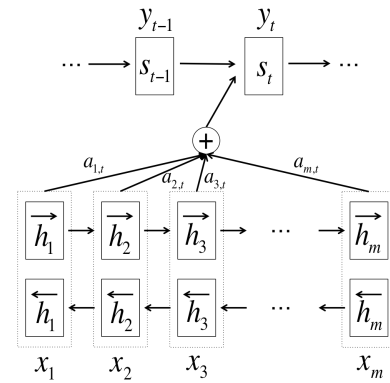


図 2: 注意型ニューラルネットワーク機械翻訳モデル

る、注意型ニューラルネットワーク機械翻訳モデルが考案された [4]。図 2 にモデルの概略を示す。

前述の Encoder-Decoder モデルでは、翻訳元言語文を前から一語ずつ入力することで再帰的に隠れ層を生成するが、このモデルではそれに加え、逆順に一語ずつ入力することで隠れ層をもう一つ生成する両方向の RNN (\vec{h} , \overleftarrow{h}) で形成されている。また基本的な Encoder-Decoder モデルでは文の全情報を一つのベクトルに集約するため、単語の系列からただ一つの隠れ層を生成する。しかしこのモデルでは複数の隠れ層を利用する。両方向の RNN で生成された隠れ層の系列に対し重みベクトル \mathbf{a} で重み付けを行うことで、これからデコードを行う単語 y_t と関連する翻訳元言語文の単語 x_1, x_2, \dots, x_m の考慮を行う。

3 ニューラルネットワーク機械翻訳における品詞情報の利用

前述のニューラルネットワーク機械翻訳では単語の系列を順番に RNN に入力することで隠れ層を生成するが、エンコード側もしくはデコード側での特徴量の考慮は、注意型ニューラルネットワーク機械翻訳モデルでのアライメントの考慮以外には特に行っていない。

一方で、従来の統計的機械翻訳のみならず他のタスクにおいても、品詞情報を利用することで精度が向上したとの報告がある [5]。このように統計的機械翻訳の特徴量として品詞情報を利用することは有効であり、ニューラルネットワーク機械翻訳の精度向上にも貢献すると考えられる。

そこで本稿では、翻訳元言語のみ、または翻訳先言語のみ、またはその両方で品詞情報を利用することで機械翻訳の精度を向上することを考える。単語の系列

```

how do you spell `` pretty " ?
why do n't we go home ?
↓
how/WRB do/VBP you/PRP spell/VB ``/'` pretty/RB "/" ?/.
why/WRB do/VBP n't/RB we/PRP go/VB home/NN ?/.

```

図 3: 品詞ラベルが付加された対訳コーパス

を順に入力するニューラルネットワーク機械翻訳で品詞情報を最も簡単に利用する方法として、対訳コーパス中の単語それぞれに品詞ラベルを付け、それを一つの単語とみなす手法が考えられる。図 3 にその例を示す。

ここで図 3 で登場する “spell” という単語について考える。図 3 中では「単語を綴る」という意味で動詞の役割をもつが、この単語は文脈によっては「魔法、魔力」という全く違う意味の名詞の役割を持つことがある。同様に “pretty” という単語には「かわいい」という意味の形容詞の役割もあれば、「非常に」という意味の副詞の役割もある。このような単語は、前述のニューラルネットワーク機械翻訳では単語の表層しか見ないために同じ単語として扱われてしまうが、意味や品詞を考えれば分かるように全く異なるものとして扱うべきである。

そこで品詞ラベルを付加し品詞曖昧性を解消することで、前述した多義語や、同じ綴りであるが過去形・過去分詞形といった役割の違いを持つ単語の区別化が可能になる。これにより翻訳精度の向上が期待できる。

4 実験

本稿では、品詞曖昧性の解消によるニューラルネットワーク機械翻訳の影響を調査するため、その問題を多く持つ英語で品詞情報を利用することを考えた。また品詞情報が暗に持つ語順等の単語の並び法則を考慮し、言語構造の異なる日本語・英語間で翻訳を行った。

4.1 実験設定

4.1.1 コーパス

今回の実験では対訳コーパスとして田中コーパス [8] を用い、日本語から英語の翻訳を行った。このコーパス中の対訳文は学生が作成しているため英文の品質が低く、口語的な表現も多いためノイズが大きい。その

ため、文の単語数が一定以上の例、対訳間で単語数が極端に異なる例、コーパス中に一定回未満しか出現しない単語を含む例を除外したのから 45,000 文を学習データ、2,560 文を開発データ、1,000 文を評価データとした。

英語の単語分割は Stanford Tokenizer ^{*1} で行い、品詞タグ付けには Lookahead Part-Of-Speech Tagger ^{*2} を用いた。日本語の単語分割には KyTea ^{*3} を用いた。英語側では品詞タグ付け済みのコーパスとそうでない通常のコーパスを準備して実験を行なった。

単語辞書に関しては、同じスペリングなのに品詞が異なるために、品詞ラベル付きで見たときに異なるものになってしまう語の登場頻度が分散してしまうため、今回は個数制限や閾値制限を設けず、学習データ中に登場する全単語を対象とした。文開始タグ・文末タグ・未知語タグに加えると、日本語側は 8,388 語、英語側は品詞ラベルの無い通常のコーパスが 6,379 語、品詞ラベル付きのものが 8,561 語となった。なお品詞ラベル付きのものに関して、未知語タグは品詞の数だけ用意した。

4.1.2 モデル

今回、ニューラルネットワーク機械翻訳のモデルとして、一層の LSTM を組み込んだ、前述の Encoder-Decoder モデル (EncDec) を用いた。このモデルは、github で公開されているソースコード ^{*4} をもとに実装を行った。

単語埋め込みは 256 次元、隠れ層の大きさは 1,024 次元とした。学習には学習率 0.5 の確率的勾配降下法を用い、バッチサイズ 128 のミニバッチ学習を行った。また評価データの翻訳時には、貪欲法に基づく出力に加え、ビーム幅 10 のビーム探索も行った。

なお入力に関して、文の逆順で行なうことで文頭のあたりの長期依存が解消されるため学習効率が向上するとの報告がある [3] ので、これに従い入力文の単語系列を逆順に入力した。

4.2 結果と考察

品詞ラベルが付加されたコーパスとそうでない通常コーパスの 2 種類のコーパスで学習させたところ、学

^{*1}<http://nlp.stanford.edu/software/tokenizer.shtml>

^{*2}<http://www.logos.ic.i.u-tokyo.ac.jp/~tsuruoka/lapos/>

^{*3}<http://www.phontron.com/kytea/index-ja.html>

^{*4}https://github.com/odashi/chainer_examples

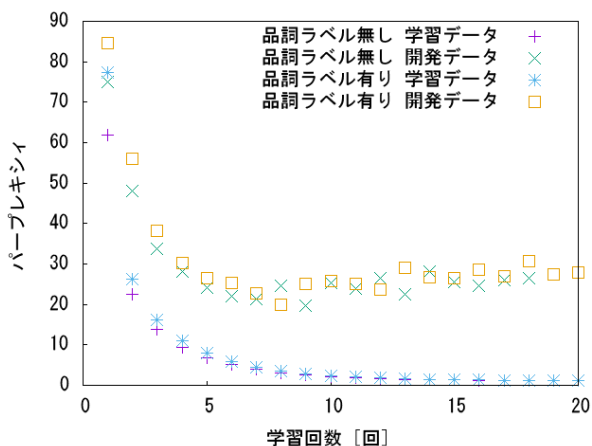


図 4: 学習回数毎のパープレキシィの変化

品詞タグの有無	ビーム幅	BLEU	RIBES
×	1	9.48	0.664
	10	8.68	0.660
×	10	13.57	0.697
	10	11.90	0.679

表 1: モデル毎のスコア

習データ・開発データのパープレキシィは図 4 のようになった。学習回数に対してパープレキシィが低下しているため学習はうまく進んで入ることが分かる。品詞ラベルの付いた単語を一つの単語とみなして辞書を作成した場合でも学習は可能であると言える。

評価データに対する翻訳結果を、機械翻訳の評価尺度 BLEU[9] と RIBES [10] で評価した結果を表 1 に示す。この結果から分かるように、品詞情報を利用していない場合と比較して、提案手法である品詞情報を利用した場合の方が翻訳の精度が低いという結果になった。

今回用いた田中コーパスは口語調のものが多く、誤った品詞タグ付けが一定数存在すると考えられる。その結果品詞情報の利用が悪影響を及ぼしてしまった箇所があると考えられる。

また誤った品詞タグ付けと関連するが、辞書サイズの増加が精度低下の原因として挙げられる。本実験では文のノイズの軽減のため低頻度の語句が登場する文を削除したため、対訳コーパスの文数から考慮しても辞書サイズはかなり小さいものになっている。その小さな辞書が、品詞ラベルを付加したことによってサイズが約 1.35 倍になった。これは単純に、デコード時に出力される単語の候補が 1.35 倍になったことになり、増えてしまった語彙がノイズとして働いてしまってい

ると考えられる。

5 おわりに

本稿では品詞情報を利用したニューラルネットワーク機械翻訳を提案した。実験では単語に関する品詞の曖昧性が高い英語に品詞情報を付加し、翻訳先言語として日本語・英語間の翻訳を行なったが、品詞情報を付加したことによって翻訳の精度が低下した。しかし今後はその原因をもとに改善することを試みる。今後の課題として、語彙の増加に左右されないようにより大規模なコーパスでの実験や、注意型ニューラルネットワーク機械翻訳での実験、より効率的に品詞情報を付加したデコーダー処理の実装を試みる。

謝辞

本研究は、JST、CREST の支援を受けたものになります。また対訳コーパスのデータ整形の際には NAIST の小田さんにご協力をいただきました。心より感謝いたします。

参考文献

- [1] 奥村学, 渡辺太郎, 今村賢治, 賀沢秀人, Graham Neubig, and 中澤敏明. 機械翻訳. コロナ社, 2014.
- [2] Michael Auli, Michel Galley, Chris Quirk, and Geoffrey Zweig. Joint language and translation modeling with recurrent neural networks. In *EMNLP*, 2013.
- [3] Ilya Sutskever, Oriol Vinyals, and Quoc V. V Le. Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems 27*. Curran Associates, Inc., 2014.
- [4] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *CoRR*, 2014.
- [5] Andreas Zollmann and Stephan Vogel. A word-class approach to labeling pscfg rules for machine translation. In *ACL*, 2011.
- [6] 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 *EMNLP*, 2014.
- [7] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 1997.
- [8] 田中康仁. 日英・パラレルコーパスの作成. In 言語処理学会第 8 回 年次大会, 2003.
- [9] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. Bleu: a method for automatic evaluation of machine translation. In *ACL*, 2002.
- [10] Hideki Isozaki, Tsutomu Hirao, Kevin Duh, Katsuhito Sudoh, and Hajime Tsukada. Automatic evaluation of translation quality for distant language pairs. In *EMNLP*, 2010.