双方向ニューラル機械翻訳の反復的な教師なし適応の検討

森田 知熙　　秋葉 友良　　塚田　元
豊橋技術科学大学
morita@nlp.cs.tut.ac.jp akiba@cs.tut.ac.jp tsukada@brain.tut.ac.jp

1　はじめに

近年，機械翻訳ではニューラルネットワークを用いた系列変換モデルによるニューラル機械翻訳 (NMT) モデルが広く使われている．NMT は，従来の統計的機械翻訳モデルよりも高い翻訳性能を持ち，人手による辞書やルールの作成の必要がないという利点を持つが，十分な翻訳性能を獲得するためにはこれまで以上に大量の対訳コーパスを必要とする．対訳コーパスの構築はコストが高く，大規模な対訳コーパスの構築は極めて困難である．その上，翻訳対象と異なるドメイン (ドメイン外) の対訳コーパスから学習したモデルでは，翻訳対象ドメイン (ドメイン内) で十分な翻訳性能を発揮することはできない深刻な問題がある．

そのため，ドメイン内の学習データが少ない場合にそれを使って，ドメイン外の豊富な学習データから学習したモデルのドメイン内での翻訳性能を改善させる教師ありドメイン適応の研究が行われている [2, 1]．しかし，教師ありドメイン適応手法では，少量とはいえドメイン内の対訳コーパスを必要とするため，対訳コーパスが存在しない場合は適応を行うことができない．

本稿では，このようなケースにおいても使用できる対訳でない原言語側と目的言語側の 2 つの単言語コーパスを用いた教師なしドメイン適応手法を提案する．本手法は，従来の NMT モデルに変更を加える必要はない．ドメイン内単言語コーパスから擬似的に作成したドメイン内対訳コーパスをドメイン外対訳コーパスに追加し再学習することを繰り返すことで実現できる．双方向の翻訳モデルを相互に性能を改善するように学習を繰り返すため，同時に両方向の翻訳モデルの性能を改善できる．

英日及び日英翻訳タスクにおいて，Asian Scientific Paper Excerpt Corpus (ASPEC) から NTCIR-8 PATMT コーパスへドメイン適応の実験を行った結果，ドメイン適応を行わなかったモデルと比べ，日英翻訳で+14.5，英日翻訳で+20.1 ポイント BLEU が向上した．また，Sennrich らの先行研究と比べても，日英翻訳で+12.6，英日翻訳で+12.1 ポイント BLEU が向上したことを確認した．

2　関連研究

2.1　ニューラル機械翻訳の半教師あり学習

翻訳モデルの学習に対訳コーパスと併用して入手が容易な単言語コーパスを活用する半教師あり学習手法が提案されている．Gülçehre ら [3] は目的言語側の単言語コーパスから学習した RNN 言語モデルをニューラル機械翻訳モデルと統合する方法を考案している．Zhang ら [9] は目的言語側と原言語側の単言語コーパスを利用し，両方向の翻訳モデルを連結して得られるオートエンコーダとマルチタスク学習を行う方法を考案している．Xia ら [4] は対訳コーパスから双方向の翻訳モデルが学習できることを利用し，両モデルの関係をノイズ混じりの通信経路に見立て，強化学習により同時に学習する方法を提案した．また，Zhang ら [10] は，原言語側と目的言語側の単言語コーパスを逆翻訳して，同時に学習を行う手法を EM アルゴリズムで定式化している．本研究は Zhang ら [10] のアプローチを単純化した手法を提案し，ドメイン適応に適用するものである．

2.2　ニューラル機械翻訳のドメイン適応

ドメイン適応手法として，ドメイン内の学習データに対訳コーパスを用いるもの (教師あり)[2, 1] と単言語コーパスを用いる方法 (教師なし)[8] がある．Freitag ら [2] は転移学習した翻訳モデルと学習前の翻訳モデルをアンサンブルすることによりドメイン内コーパスへのオーバーフィットを防ぎ，翻訳性能を改善した．Chu ら [1] はドメイン外コーパスとドメイン内コーパスにそれぞれ”out-of-domain”,”in-domain”タグを付加し，

転移学習をすることで翻訳性能が向上したことを報告している．

Sennrich ら [8] は，目的言語側の単言語コーパスを逆翻訳し，学習データに加えることで翻訳性能が大きく向上することを報告した．この手法は，単言語コーパスによる半教師あり学習手法であるが，ドメイン適応においても有効であることを示している．提案法は，Sennrich らの手法を双方向かつ反復的に拡張したものになっている．

3 提案手法

提案手法は，ドメイン外の対訳コーパスの他に，翻訳対となる 2 つの言語のドメイン内単言語コーパスを用いてドメイン適応を行う．2 つの単言語コーパスは対応付けられたコンパラブルコーパスである必要はない．提案法は両方向の翻訳システムを同時に学習するので，原言語と目的言語の区別は重要ではなく，2 つの言語は対等である．以降，2 つの言語をそれぞれ X,Y と記し，言語 X から Y への翻訳を X-Y，Y から X への翻訳を Y-X と記す．

提案法の手順は以下の通りである (**図 1**)．

1 ドメイン外の対訳コーパス D_X^{out}, D_Y^{out} から X-Y，Y-X の両方向の NMT モデルを学習する．以降，これをモデル 0 と呼び，モデル番号 i を 0 に初期化する．

2 X-Y NMT モデルを以下の手順で再学習する．

2.1 Y のドメイン内単言語コーパス D_Y^{in} から Y-X NMT モデル i を用いて翻訳結果 $D_X^{in\prime}$ を得る．

2.2 $D_X^{in\prime}$ と D_Y^{in} の組を疑似対訳コーパスとして D_X^{out}，D_Y^{out} と混合し，X-Y NMT モデルを学習し，X-Y NMT モデル i+1 とする．

3 Y-X NMT モデルを以下の手順で再学習する．

3.1 X のドメイン内単言語コーパス D_X^{in} から X-Y NMT モデル i を用いて翻訳結果 $D_Y^{in\prime}$ を得る．

3.2 $D_Y^{in\prime}$ と D_X^{in} の組を疑似対訳コーパスとして D_Y^{out}，D_X^{out} と混合し，Y-X NMT モデルを学習し，Y-X NMT モデル i+1 とする．

4 i ← i+1 としてステップ 2 に戻る．

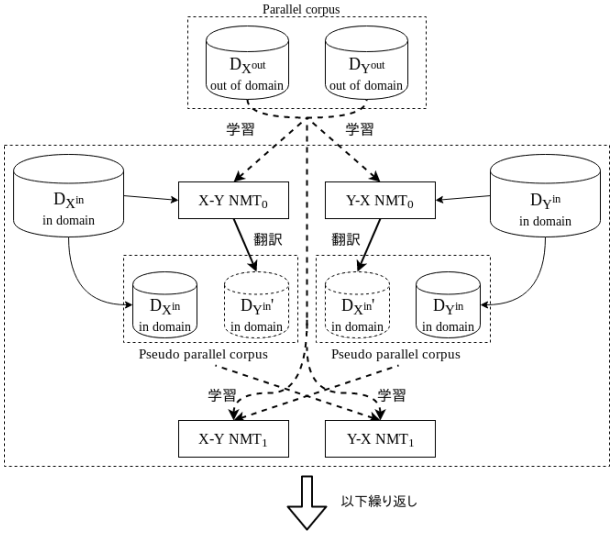

図 1: 翻訳対となる 2 つの言語の単言語コーパスを用いた反復学習のフロー

上記の 2〜4 の手順を繰り返し，疑似対訳コーパスの品質を向上させることで，NMT モデルの性能改善を図る．

4 実験

本実験では，提案手法のドメイン適応での効果を調べるため，単言語コーパスを用いて提案手法で適応を行い，ドメイン内テストデータに対する翻訳精度の調査を行う．翻訳精度の評価には BLEU[7] を使用する．ドメイン適応を行わないモデル (モデル 0, ベースライン)，Sennrich ら [8] の手法 (モデル 1 に相当)，ドメイン内対訳コーパスから教師あり学習を行ったモデルと性能を比較する．

4.1 データセット

ドメイン外の教師ありデータには Asian Scientific Paper Excerpt Corpus (ASPEC)[6] の英日対訳コーパスを利用した．逆翻訳システムの学習には対訳文の全文 (1,000,000 対) を用いた．

ドメイン内の教師なし単言語コーパスには NTCIR-8 PATMT の英日対訳コーパスを利用した．また，2 つの言語の単言語コーパスに対訳となるような文ペアが含まれないことを確実にするために，このコーパスの先頭 10 万行を除いた 3,086,284 対を 2 分割し，前半側からは英語のみ，後半側からは日本語のみを抽

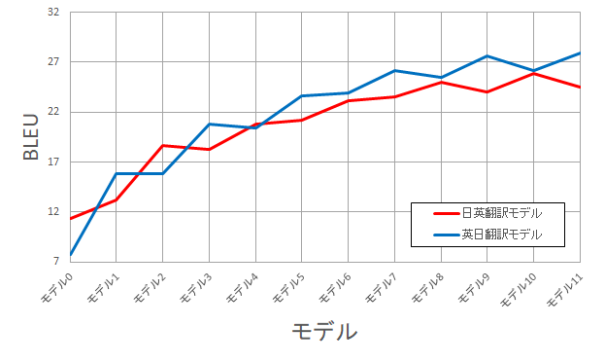

図 2: ドメイン適応の反復学習の各学習段階における BLEU 推移

表 1: テストデータに対する未知語率 (%)

翻訳モデル	英日翻訳	日英翻訳
モデル 0	9.3	21.0
モデル 1	13.0	26.5
モデル 2	4.2	1.4
モデル 5	2.1	1.3
モデル 9	1.9	1.2

出し，単言語コーパスとして用いた．テストデータは NTCIR-8 PATMT の学習データ以外から 899 対，開発データは 2000 対を用いた．両方のコーパスとも，日本語文は MeCab[1]により分かち書きを行った．

4.2 実験条件

ニューラル機械翻訳システムには OpenNMT[5] を用いた．英日翻訳システム，日英翻訳システム共にエンコーダは 1 レイヤーの BRNN(500 次元)，デコーダは 1 レイヤーの RNN(500 次元) とし，学習アルゴリズムは Adam，学習率は 0.001 で 10 エポック学習した．ボキャブラリサイズは 30K とした．

ASPEC コーパスの全文から英日，日英の双方向の初期モデル (モデル 0) を学習した．モデル 0 のボキャブラリは，ASPEC の対訳コーパスで出現頻度の高い 30K を選択した．モデル 1 以降のボキャブラリは，モデル i の再学習の度に NTCIR8-PATMT の単言語コーパスから作成する疑似対訳コーパスだけを用いて出現頻度の高い 30K を選択した．

[1] http://taku910.github.io/mecab/

4.3 実験結果

図 2 は，提案手法の反復学習の各 NMT モデル i の BLEU 推移を表している．**表 2** に，各条件での BLEU を示す．**図 2** から英日翻訳，日英翻訳共にモデルを再学習するごとに BLEU が向上しており反復学習が有効なことがわかる．モデル 0(ドメイン適応を行わずドメイン外対訳コーパスのみで翻訳モデルを学習するベースライン) に対し英日翻訳で BLEU が+20.1(モデル 11)，日英翻訳で+14.5(モデル 10) ポイント改善された．また，Sennrich ら [8] の手法に相当するモデル 1 と比べても，英日翻訳で BLEU が+12.1，日英翻訳で＋ 12.6 ポイント改善されている．これらにより提案手法の反復学習が翻訳精度の大幅な改善に効果的であることがわかる．

図 2 の英日翻訳モデルと日英翻訳モデルの BLEU を比べると，モデル 2 では日英翻訳の精度向上が大きく，次のモデル 3 では英日翻訳，モデル 4 では再び日英翻訳，というように改善される翻訳方向が交互に入れ替わっていることがわかる．精度向上は学習に用いる擬似対訳コーパスの質，すなわち一つ前の逆翻訳モデルの精度，に依存するため，大きく改善した方向の翻訳モデルは次のモデルでは反対方向の翻訳モデルの改善に大きく貢献する．本実験結果では結局，モデル 0 で精度の高い方から派生するモデル (日英で i が偶数，英日で i が奇数) の方が，もう一方の精度の低い方から派生するモデル (日英で i が奇数，英日で i が偶数) より一貫して精度が良い．モデル 11 まで学習した場合は，日英モデル 10，英日モデル 11 がベストなモデルとなっている．

また，ドメイン内の対訳コーパスを用いて学習した場合 (**表 2** の Oracle) と比較すると，提案法 (英日 27.91，日英 25.83) は 10 万対の対訳コーパスで学習したモデル (英日 25.30，日英 25.21) と同等の性能を達成していることがわかる．このことは，対訳コーパスの 15 倍のサイズ (文タイプで数えると 30 倍) の単言語コーパスを収集すれば，本手法により教師あり学習と同等の翻訳性能を達成できることを示している．

表 1 に，各翻訳モデルのテストデータに対する原言語側の未知語率を示す．モデル 0(ベースライン)，モデル 1(Sennrich らの手法) の未知語率が高いのに対し，モデル 2 で急激に低下し，以降少しずつ未知語が減少している．未知語の観点からも，Sennrich らの手法を 2 回以上適用する提案法がドメイン適応において特に有効であることが分かる．

表 2: 反復学習における教師なしドメイン適応と従来法の比較

ドメイン外コーパス (ASPEC)	ドメイン内コーパス (NTCIR PATMT)	翻訳モデル (適応手法)	英日翻訳 (BLEU)	日英翻訳 (BLEU)
対訳 100 万ペア	-	モデル 0 (適応なしベースライン)	7.78	11.31
対訳 100 万ペア	単言語 150 万+150 万	モデル 1 ([Sennrich et al.])	15.79	13.17
対訳 100 万ペア	単言語 150 万+150 万	モデル 10 (提案法)	26.21	**25.83**
対訳 100 万ペア	単言語 150 万+150 万	モデル 11 (提案法)	**27.91**	24.53
-	対訳 10 万ペア	Oracle	25.30	25.21
-	対訳 150 万ペア	Oracle	39.96	36.02

翻訳モデルの原言語側ボキャブラリは，擬似コーパスを作成するのに用いられる直前の反対方向の翻訳モデルのボキャブラリに制限される．例えば，日英翻訳のモデル 1 の日本語ボキャブラリは，英語の単言語コーパスを英日翻訳のモデル 0 で翻訳した結果から構築されるため，モデル 0 の目的言語側ボキャブラリの範囲に制限されるが，モデル 0 はドメイン外対訳コーパスで学習されているためドメイン内の語彙を十分にカバーできない．さらに，モデル 0 がドメイン外対訳コーパスから直接ボキャブラリを構築しているのに対し，モデル 1 はこの対訳コーパスのボキャブラリの範囲で単言語コーパスを翻訳した結果から構築するため，モデル 0 のボキャブラリのサブセットとなる．これがモデル 0 よりもモデル 1 の未知語率が高くなる理由である．一方，日英翻訳のモデル 2 の日本語ボキャブラリは，英語の単言語コーパスを英日翻訳のモデル 1 で翻訳した結果から構築される．このモデル 1 の目的言語側ボキャブラリは日本語単言語コーパスから直接構築されるため，ドメイン内の語彙をカバーすることができる．モデル 3 以降も同様である．これがモデル 2 以降で未知語率が改善されることの理由である．

5 おわりに

本研究では対になる双方向の翻訳モデルを作成して単言語コーパスの逆翻訳と学習を繰り返すことで 2 つのモデルを相互に改善する方法を提案し，ドメイン適応での有効性を検証した．実験の結果，単言語コーパスによる学習の繰り返しは特にドメイン適応において効果が大きく，英日翻訳で+12.1，日英翻訳で+12.6 ポイントと Sennrich らの手法を大きく上回る BLEU を達成した．また，対訳コーパスの 15 倍のサイズ (文タイプ数では 30 倍) の単言語コーパスにより同等の性能を持つ翻訳モデルが構築可能であることを確認した．

謝辞

本研究は JSPS 科研費 18H01062 および 16K00153 の助成を受けた．

参考文献

[1] C. Chu, R. Dabre, and S. Kurohashi. An empirical comparison of simple domain adaptation methods for neural machine translation. In *ACL*, 2017.

[2] M. Freitag and Y. Al-Onaizan. Fast domain adaptation for neural machine translation. *CoRR abs/1612.06897*, 2016.

[3] Ç. Gülçehre, O. Firat, K. Xu, K. Cho, L. Barrault, H. Lin, F. Bougares, H. Schwenk, and Y. Bengio. On using monolingual corpora in neural machine translation. *CoRR abs/1503.03535*, 2015.

[4] D. He, Y. Xia, T. Qin, L. Wang, N. Yu, T.-Y. Liu, and W.-Y. Ma. Dual learning for machine translation. In *Advances in Neural Information Processing Systems 29*, pages 820–828. 2016.

[5] G. Klein, Y. Kim, Y. Deng, J. Senellart, and A. M. Rush. OpenNMT: Open-source toolkit for neural machine translation. *CoRR abs/1701.02810*, 2017.

[6] T. Nakazawa, M. Yaguchi, K. Uchimoto, M. Utiyama, E. Sumita, S. Kurohashi, and H. Isahara. ASPEC: Asian scientific paper excerpt corpus. In *LREC*, 2016.

[7] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu. BLEU: A method for automatic evaluation of machine translation. In *ACL*, pages 311–318, 2002.

[8] R. Sennrich, B. Haddow, and A. Birch. Improving neural machine translation models with monolingual data. In *ACL*, 2016.

[9] J. Zhang and C. Zong. Exploiting source-side monolingual data in neural machine translation. In *EMNLP*, pages 1535–1545, 2016.

[10] Z. Zhang, S. Liu, M. Li, M. Zhou, and E. Chen. Joint training for neural machine translation models with monolingual data. In *AAAI*, 2018.