3Q1-5in1

ニューラルネット機械翻訳における自動コーパス生成適用

Automatic corpora generation applied to neural machine translation

今出 昌宏 *1 藤原 菜々美 *1 山内 真樹 *1, *2
Masahiro Imade Nanami Fujiwara Masaki Yamauchi

*1 パナソニック株式会社 先端研究本部 *2 情報通信研究機構
Advanced Research Division, Panasonic Corporation NICT

Recently, Neural Machine Translation (NMT) systems have been become widely used. NMT is a system whose model is trained on large quantities of parallel corpora. We have been developing a unique method which automatically generates large-size parallel corpora from a small number. It has been confirmed that generated corpora by the method is effective for improving the performance of Statistical Machine Translation (SMT), though has not for NMT yet. In this work, we have applied the method to NMT for the first time. NMT performance using the corpora generated by our system is improved more than 5.9 points on BLEU compared to the original small amount of corpora, and it is equal to or better than those manually generated by human. While this method will be able to greatly improve translation quality and reduce the cost for corpora preparation, it is also suggested that excessive corpora expansion may lead to degradation of NMT performance.

1. はじめに

訪日外国人観光客の急増や 2020 年オリンピック・パラリンピック東京大会の開催を背景に，交通・道路・観光・サービスの各分野で多言語対応の取り組みが進められている．各所で言葉の壁をなくす多言語翻訳システムの社会実装が推進されており，その要となるのが機械翻訳技術である．とくに統計的機械翻訳(SMT: Statistical Machine Translation) [KOEHN 03]とニューラルネット機械翻訳(NMT: Neural Machine Translation) [Forcada 97]の 2 つが自動翻訳には欠かせない技術であり，SMT については「旅行会話」などの領域で実証実験段階となっている [松田 13]．また，NMT についてはここ数年で急速な進化をとげており [Cho 14] [Sutskever 14] [Bahdanau 15] [Luong 15]，NMT ベースの翻訳サービスも公開されつつある[Wu 16]．

SMT と NMT，どちらも大量の対訳コーパス(原言語と目的言語の対訳文集)から，翻訳に必要なモデルを学習し，そのモデルを通じて翻訳を行う機械翻訳システムである．実用的なモデル構築には約 100,000 文～1,000,000 文単位での対訳コーパスが必要とされているが，対訳コーパスの収集は一般に困難である．特に初期段階で準備できる対訳コーパス量は，ドメインに依らずおおよそ 1,000～10,000 文程度である．少量の対訳コーパスでは学習に十分な情報が得られず，機械翻訳性能は著しく低下する．

これに対し我々は，少量の対訳コーパスから，機械翻訳エンジン構築に適用可能な品質でかつ十分量の対訳コーパスを自動的に獲得すべく，自動対訳コーパス生成手法(ACG: Automatic Corpora Generation)を開発している[山内 16] [藤原 16a,16b,17] ．これは，種となる少量の対訳コーパス(原文)に対して単語や文節(フレーズ)の言い換え処理を行い，原文とは別表現で同じ意味内容（同意異表現）であることが期待される類似候補文を生成し，さらにその中からより好ましい文を識別出力する手法である．原文は少量でよく，低コストで対訳コーパスを拡充することができる．

これまでに ACG で拡充した対訳コーパスが SMT の翻訳性能向上に寄与することを確認している．しかしながら，SMT で有効なそれらの手法が，NMT においても有効であるかは，これまで確認されていない．そこで，NMT への ACG 技術適用を目的に以下の評価を実施した．

i. NMT 学習への同意異表現コーパス拡張手法適用による翻訳性能変化により，NMT に対する同手法の効果を評価
ii. ACG および人手による同意異表現コーパス拡張による翻訳性能比較により，NMT における ACG 実用性を評価

本稿では上記評価内容について， NMT における自動コーパス生成適用の第一報として報告する．

{ imade.masahiro, fujiwara.nanami, yamauchi.masaki } at jp.panasonic.com
{ yamauchi } at nict.go.jp

2. 自動対訳コーパス生成

ACG の構成概要図を Fig. 1 に示す．ACG は，種となる対訳コーパス(原文)を入力とし，「類似候補文生成」器と「候補識別」器により多量の対訳コーパス(識別結果文)を生成する．これに類する先行研究もいくつかあるが[Madnani 13] [Yuval 13]，本構成に類似の枠組みについては，これまでのところ見出だせていない．

2.1 類似候補文生成

「類似候補文生成」器では，言い換え表現のデータベース(換言 DB)を言語資源(WordNet[Word 09], ALAGIN 言語資源[ALAGIN 10] , PPDB[Mizukami 14], 内容語換言辞書[山形 14]等)，及び手作業から構築し，入力文に対して適用することで類似候補文を得る．

類似候補文の生成模式図を Fig. 2 に示す．原文(ここでは日本語文) 1 文に含まれる語句・文節に対して同時に 1 箇所の置換を行い，類似候補文を生成する．ただし，必ずしも正しい文

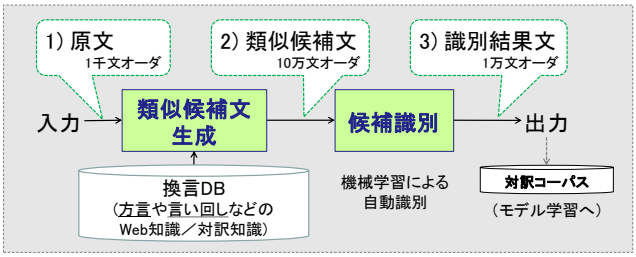

Fig. 1 Illustration of automatic corpora generation

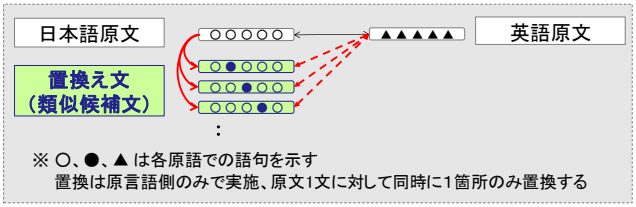

Fig. 2 Paraphrasing illustration

章のみが生成されるのではなく，不自然な，意味的・文法的に破綻した文も生成される可能性がある．これは，原文の対訳コーパスの想定ドメインが換言 DB のエントリと必ずしも合致しないことや，エントリ自身のノイズ等に起因する．

次段の「候補識別」器では，このような破綻文を除外し，対訳コーパスに適切な文を識別する．

2.2 候補識別

「候補識別」器では，類似候補文に対して識別器を適用し，“良い文”の集合として識別結果文を得る．類似候補文から人手で良質な対訳コーパスを抽出することによって，翻訳性能が向上することが確認されているため[藤原 16a]，如何に品質の良い対訳コーパスを自動で識別できるかが鍵となる．

識別タスクにおける素性には，N-gram [矢田 10] を用いる．語句の置換えが発生した箇所を含む素性から，“良い文”“悪い文”の識別を行う「候補識別」器を構築する．「候補識別」器で識別された文は，識別結果文として出力され，SMT あるいは NMT の訓練文(対訳コーパス)として使用する．

識別器の具体的な処理として，ある類似候補文 1 文において，語句置換えが行われた語を最低 1 語含む N-gram を k 個取得する．k 個のそれぞれについて N-gram の出現確率の対数尤度を求め，その平均値に対して閾値判定を行う．語句置換えが行われた箇所の周辺のフレーズに対し，それらがある一定以上の出現確率を持つ場合に，“良い文”として選択された識別結果文となる[藤原 16b]．

N-gram の出現確率は，単語 $\omega_1\omega_2\cdots\omega_n$ の出現確率を $P(\omega_1\omega_2\cdots\omega_n)$ と表すと，以下の式で求めることができる．

$$P(\omega_1\omega_2\cdots\omega_n)\cong\prod_{i=1}^{n}P(\omega_i|\omega_{i-N+1}\cdots\omega_{i-1})$$

出現頻度から N-gram 確率を推定する場合，N-gram の学習モデル中に出現しない単語も多く存在するため，学習モデルに含まれない単語の確率値は 0 となる場合がある．その回避のため，加算スムージングを行い，出現確率を求める際に，N-gram の出現回数に一定値を加える．

$$P(\omega_i|\omega_{i-N+1}\cdots\omega_{i-1})=\frac{C(\omega_{n-N+1}^{n})+\delta}{C(\omega_{n-N+1}^{n-1})+\delta V}$$

ここで $C(\omega_m^n)$ は単語列 $\omega_m\omega_{m+1}\cdots\omega_{n-1}\omega_n$ が N-gram の学習モデル中に出現する回数，V は単語列の異なり総数，δ はスムージング定数 (=0.5)である．

3. 実験評価

本章では，第1章で触れた以下の評価内容の詳細と評価結果について述べる

(i)NMT における同意異表現コーパス拡張手法の効果確認
(ii)ACG および人手拡張コーパスの NMT における性能比較

評価は日英翻訳タスクの性能を指標として実施した．比較評価に必要な複数の訓練コーパスセットと評価文セットで NMT 学習および翻訳タスクを実施し，それぞれの翻訳結果に対し BLEU (BiLingual Evaluation Understudy) [Papineni 02]による客観評価，および主観評価を行うことで，性能比較を実施した．

3.1 評価条件

比較評価に用いる訓練コーパスセットを以下に示す．

(A) 原文コーパス：道案内における行動指示などで使われる言い回しを含んだ対訳コーパス（独自収集・作成）
(B) 人手拡張コーパス：原文の日本語側表現を人手で同義言い換えして拡張したコーパス（原文コーパス自体も含む）
(C) 自動生成コーパス 1：原文コーパスを ACG で拡張したコーパス（原文コーパス自体も含む）
(D) 自動生成コーパス 2：人手拡張コーパスを ACG で拡張したコーパス（人手拡張コーパス自体も含む）

(A)原文コーパスでの翻訳性能は評価内容(i)における評価基準となる．(A)に対し，(B)～(D)で翻訳性能が良化すれば，同意異表現コーパス拡張手法が，NMT においても効果有りといえる．

(B)と(C)の比較は評価内容(ii)に相当する．加えて(D)は(C)の拡張規模を増大したものに相当し，(C)と(D)の比較で拡張規模の影響を評価できる．各コーパスセットに含まれる対訳コーパス数を Table.1 に示す．(B)と(C)は(A)の原文数に対し 4 倍量程度であるのに対し，(D)は約 16 倍量になっているのがわかる．

評価文セットは以下の 2 種を用意した．

(a) クローズ評価文：(A)原文コーパスからランダムに抽出した 50 文 (訓練コーパス(A)～(D)の全てに含まれる)
(b) オープン評価文：(A)原文コーパスと同ドメインで収集・作成した 50 文 (訓練コーパス(A)～(D)のいずれにも含まれない)

ドメイン内の汎用的な性能評価としての意味を持つのは(b)オープン評価文であるが，(a)クローズ評価文もあわせて評価することで，適正な学習が行われているかを判断するのに役立つ．もし(a)の翻訳結果の精度が非常に良いにもかかわらず(b)の精度が極端に悪い場合には過学習が疑われる．また，もし(a)(b)と

Table. 1 Number of training corpora

訓練コーパスセット	コーパス数
(A) 原文コーパス	21,621
(B) 人手拡張コーパス	82,472
(C) 自動生成コーパス 1	76,155
(D) 自動生成コーパス 2	335,368

もに精度が悪ければ学習不足と考えられる.

NMT の学習および翻訳には OpenNMT ツールキット[Klein 17]を用いた. 学習設定は, 2 層 LSTM による RNN エンコーダー/デコーダーで, 隠れ層と単語埋め込みは 1380 次元とした. パラメータの最適化手法には学習率 0.001 で Adam を使用した. またドロップアウト確率は 0.5 とした. 評価文翻訳時の設定は OpenNMT のデフォルト設定であるビーム幅 5 のビーム探索とした. 各訓練コーパスセットの学習は 20 epoch まで実施し, 評価文の翻訳タスクは全 epoch のモデルについて実施した.

翻訳結果は BLEU により客観評価し, 結果を比較した. また最終的な epoch 20 の翻訳結果については, 人手による主観評価を実施し, (A)～(D)の翻訳品質をそれぞれ比較した. 主観評価基準は以下の 4 段階で設定した.

SA-rank : 流暢な表現で意味も正しい
B-rank : 非ネイティブな表現だが意味は伝わる
C-rank : 表現には誤りを含むがある程度の意味は推測可能
D-rank : 翻訳として破綻

3.2 客観評価

Fig.3 に, (A)～(D)の各訓練コーパスセットで学習したモデルで epoch 毎に評価した BLEU 値の推移を示す. 縦軸は BLUE 値, 横軸は学習の epoch 数である. また, Table. 2 に epoch 20 における(A)～(D)の各モデルで翻訳した評価文(a)(b)の BLEU 値を示す.

Fig.3 を見ると, (a)クローズ評価文と(b)オープン評価文のいずれにおいても, 学習の進行に伴って BLUE 値が向上しているのが確認できる. また, (a)クローズ評価文の BLUE 値は約 80～90 の非常に高い値に収束し, かつ(b)オープン評価文の BLEU 値もそれなりの値に達している. これらの結果から, NMT の学習は, 過学習や学習不足に陥ることなく, 適切に進んだものと考えられる.

原文のみの(A)に比べ, 人手または ACG で同意異表現コーパス拡張した(B)～(D)は, より高い BLUE 値へと収束している. Table.2 に, 最終的に epoch 20 において到達する BLEU 値をまとめた. ドメイン内汎用翻訳性能を確認するための評価文(b)の BLUE 値は, それぞれ(A)に対して(B) +5.1, (C) +5.9, (D) +9.3 といずれも良化していることがわかる. これらの結果より, 評価内容(i)に関しては, SMT と同様, NMT においても同意異表現コーパス拡張が翻訳性能向上に効果のあることが確認された. さらには, (B)～(D)は, (A)より少ない学習度で(早い段階で)一定レベルの BLUE 値に到達しており, この結果より, 同意異表現によるコーパス拡張は, ある翻訳レベルに達するまでの学習回数を少なくできる効果もあるといえる. 今回の結果の中では, コーパス数最多の(D)において特にその傾向が顕著である.

Fig.3 および Table.2 において(B)と(C)を比較すると評価文(a)(b)のいずれにおいても, (C)は(B)と同等以上の良化を示している. これらの結果より, 評価内容(ii)に関しても, SMT 同様 NMT においても, 翻訳性能向上に対する ACG の効果は人手拡張と同等以上であることが確認された.

以上より, ACG は NMT においても低コストでの翻訳品質改善に有用であることが示された.

3.3 主観評価

(b)オープン評価文に対する(A)～(D)の epoch 20 における翻訳結果についての主観評価結果を Fig.4 に示す. 評価文数

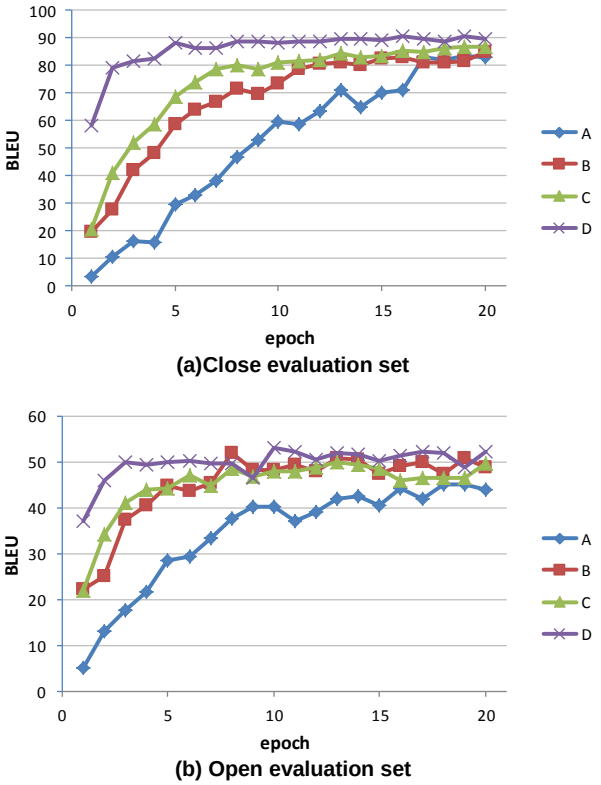

Fig. 3 BLEU plots for (a)close evaluation set and (b)open evalution set.

Table. 2 BLEU at epoch 20

訓練コーパスセット	BLEU 評価文(a)	BLEU 評価文(b)
(A) 原文コーパス	82.8	43.9
(B) 人手拡張コーパス	84.7	49.0
(C) 自動生成コーパス 1	87.0	49.8
(D) 自動生成コーパス 2	89.9	53.2

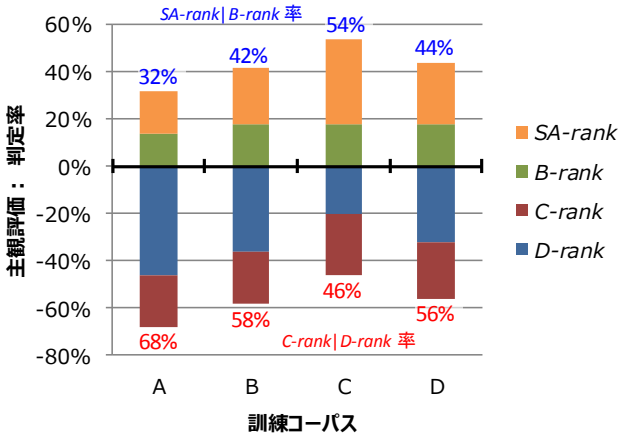

Fig. 4 Subjective evaluation results for open evaluation set

に占める *SA-rank* | *B-rank* | *C-rank* | *D-rank* の各判定率を棒グラフで表し，翻訳された目的言語(英語)において原言語(日本語)の意味が通じると判断された *SA-rank* | *B-rank* 判定率 は正値，正確な意味の通じ難い *C-rank* | *D-rank* 判定率は負値でプロットしている．

主観評価においても，客観評価と同様に(B)～(D)では(A)に対する良化(*SA-rank*|*B-rank* 率で+10%～+22%)が確認できる．また(B)と(C)の比較においては(C)が *SA-rank*|*B-rank* 率で+12%勝っており，評価内容(i)(ii)についての結果は，主観評価においても客観評価で確認された傾向と基本的には同じといえる．

しかしながら，(C)と(D)を比較した場合の傾向については客観評価とは異なり，(C)が最良の結果となった．この結果は，同意異表現のコーパスを過剰に追加すると，実質的な翻訳性能の向上への寄与は小さく，むしろ逆効果，すなわち翻訳の品質が向上する事例よりも，翻訳の品質の下がる事例が増えてしまう場合があることを示唆している．ただし，本結果は限定された 50 文のオープン評価文に対する結果であるので，今後，クロスバリデーション等でより汎化した性能確認が必要である．また拡張コーパスで良化/悪化する評価文の分類等による詳細な分析も必要と考える．

4. まとめ

少量対訳コーパスからの実用的な機械翻訳エンジンの構築を狙いとして，対訳コーパスを同意異表現で自動拡張する手法開発を行っている．

本手法で生成した拡張コーパスを NMT 学習へ適用して得た翻訳結果に対する客観評価および主観評価により，(i)同意異表現コーパス拡張が NMT においても翻訳性能向上に効果があること，(ii)ACG による拡張コーパスは一般的な人手拡張コーパスと同等以上の翻訳性能の向上効果（原文のみに対し BLUE 値+5.9～+9.3 ポイント向上）があることを確認した．これらの結果より，ACG は NMT においても低コストでの翻訳品質改善に有用であることが示された．

ただし，ACG のコーパス拡張規模と翻訳性能との相関性に関しては，客観評価と主観評価とでは異なる傾向が見られ，過剰なコーパス拡張は実質的な翻訳性能向上に対し逆効果になる可能性が示唆された．今後，より詳細な性能確認と分析実施の必要がある．

参考文献

[ALAGIN 10] ALAGIN 言語資源: https://alaginrc.nict.go.jp

[Bahdanau 15] Bahdanau, D., et al.: Neural machine translation by jointly learning to align and translate, in *Proc ICLR*(2015)

[Cho 14] Cho, K. et al: Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation, in *Proc. of EMNLP* 2014, pp. 1724–1734 (2014)

[Forcada 97] Forcada, M., et al: Recursive hetero-associative memories for translation, *Neuroscience to Technology*, vol. 1240, pp. 453-462 (1997)

[Klein 17] Klein, G., et al.: OpenNMT: Open-Source Toolkit for Neural Machine Translation, *ArXiv e-prints arXiv:1701.0281* (2017)

[KOEHN 03] KOEHN, P.: Statistical Phrase-Based Translation, *Proc. Joint Conference on Human Language Technologies and the Annual Meeting of the North American Chapter of the Association of Computational Linguistics* (HLT-NAACL-03), (2003)

[Luong 15] Luong, M., et al.: Effective Approaches to Attention-based Neural Machine Translation, in *Proc of EMNLP* (2015)

[Madnani 13] Madnani, N., et al.: Generating targeted paraphrases for improved translation, *ACM Trans. Intell. Syst. Technol.*4, 3, Article 40 (2013)

[Mizukami 14] Mizukami, M., et al.: Building a Free, General-Domain Paraphrase Database for Japanese: The 17th Oriental COCOSDA Conference (2014)

[Papineni 02] Papineni, K., et al.: BLEU: a method for automatic evaluation of machine translation, *40th Annual meeting of the Association for Computational Linguistics*, p.311-318 (2002).

[Sutskever 14] Sutskever, I., et al.: Sequence to Sequence Learning with Neural Networks, in *Proc. of NIPS 2014*, pp3104–3112 (2014)

[Word 09] Japanese Wordnet (v1.1): http://compling.hss.ntu.edu.sg/wnja/

[Wu 16] Wu Y., et al.: Google's Neural Machine Translation System: Bridging the Gap between Human and Machine Translation, *arXiv reprint arXiv:1609.08144* (2016)

[Yuval 13] Yuval, M., et al.: Distributional Phrasal Paraphrase Generation for Statistical Machine Translation, *ACM Trans. Intell. Syst. Technol*.4, 3, Article 39 (2013)

[藤原 16a] 藤原他: 自動コーパス生成による少量対訳コーパスからの統計的機械翻訳， 言語処理学会第 22 回大会 (2016)

[藤原 16b] 藤原他: コーパスの自動生成・識別による少量コーパスからの統計的機械翻訳， 第 15 回情報科学技術フォーラム (2016)

[藤原 17] 藤原他: 自動コーパス生成とユーザフィードバックによる機械翻訳, 言語処理学会第 23 回大会 (2017)

[松田 13] 松田他: 多言語音声翻訳システム"VoiceTra"の構築と実運用による大規模実証実験, 信学 D, No.10, pp.2549-2561(2013)

[矢田 10] http://s-yata.jp/corpus/nwc2010/ngrams/

[山内 16] 山内他: 自動コーパス生成とフィードバックによる少量コーパスからの統計機械翻訳, 人工知能学会第 30 回大会 (2016)

[山形 14] 山形他: 普通名詞換言辞書の構築: 言語処理学会第 20 回年次大会, pp.7-10 (2014)