

目的言語文の固有表現タグ付与に基づく Transformer ニューラル機械翻訳

南端 尚樹,

田村 晃裕,
同志社大学 理工学部

加藤 恒夫

{cguc1041@mail4, aktamura@mail, tsukato@mail}.doshisha.ac.jp

概要

本研究では、目的言語文の固有表現 (NE) をタグの付与により考慮する Transformer に基づくニューラル機械翻訳 (NMT) モデルを提案する。目的言語文の NE を考慮する従来の NMT は、原言語と目的言語間で対応付いた NE 部分を、NE の種類を表す記号に置き換えて翻訳を行う。そのため、原言語側と対応付かない目的言語文の NE は考慮できない。また、NE の構成単語を原言語文の翻訳で考慮できない。そこで本研究では、目的言語文内の NE の前後に NE の種類と開始/終了情報を含むタグを追加して翻訳を行う、Transformer NMT モデルを提案する。WMT2014 の英語とドイツ語間のニュース翻訳タスクにおいて提案モデルを評価した結果、目的言語文の NE を考慮しないベースラインの NMT に比べて、英独翻訳では最大 0.09 ポイント、独英翻訳では最大 0.35 ポイント BLUE が向上した。

1 はじめに

NLP では古くから機械翻訳の研究が行われており、近年では、NMT の研究が盛んである。NMT の性能改善を目指す研究の流れの一つとして、原言語文や目的言語文中の単語の品詞や文構造といった言語学的素性を活用する試みが行われている。その中で、言語学的素性として、人名や地名、組織名といった特定の表現を表す NE に着目し、NE を活用する NMT の研究が行われている [2, 3, 4, 9]。

NE を活用する従来の NMT モデルのほとんどは、原言語文に対する NE の認識 (NER) 結果を用いる。NE には複合語が多く存在するため、原言語文の NE 情報を NMT に与えることで原言語文中の単語のチャンク情報を翻訳時に活用できる。また、多義語に対する NER は翻訳時の語義曖昧性を減らす効果があるという報告もある [9]。しかし、これらの従

来モデルは目的言語文の NE を活用していない。

目的言語文の NE を考慮する NMT モデルは、Li ら [2] により提案されている。この従来モデルは、原言語文と目的言語文の双方で NER を行い、原言語と目的言語間で対応付いた NE 部分を、その NE の種類を表す記号 (NE 種別記号)¹⁾で置き換えた文を NMT で翻訳する。そのため、原言語文と目的言語文の両方の NE タガーが必要である。そして、原言語と目的言語間で対応付いた NE しか活用できないため、原言語側と種類や粒度が異なる目的言語文の NE の情報は失われる。また、NE 部分は記号に置き換わってしまうため、NE の構成単語を考慮して原言語文を翻訳できない。

そこで本研究では、目的言語文中の NE の前後に、NE の種類と開始/終了情報を含むタグを付与して翻訳を行う、Transformer NMT モデルを提案する。提案モデルでは、原言語と目的言語間で NER 結果の対応付けを行わずに、目的言語の NER の結果にのみ基づいて目的言語文の NE を活用する。そのため、原言語と目的言語で NE の種類や粒度が異なっても、NER で認識された全ての NE を考慮できる。さらに、提案モデルでは目的言語文内の単語を削除しないため、NE の構成単語を考慮した翻訳が可能である。

WMT2014 の英語とドイツ語間のニュース翻訳タスクにおいて提案モデルを評価した結果、目的言語文の NE を考慮しないベースライン NMT に比べて、英独翻訳では最大 0.09 ポイント、独英翻訳では最大 0.35 ポイント BLUE が向上した。

2 従来モデル

本節では、目的言語文の NE を考慮する従来の NMT モデル [2] を説明する。従来モデルは、原言語

1) 数値・時間表現を表す「N/T」、地名を表す「LOC」、人名を表す「PER」の3種類の NE 種別記号を使用している。

と目的言語間に対応付いた NE を NE 種別記号で置き換えた文を翻訳する文翻訳機と、NE を独立に翻訳する NE 翻訳機で構成される。文翻訳機と NE 翻訳機は、それぞれ、単語レベルと文字レベルの RNN に基づく NMT[1] を用いている。

学習時には、教師データである対訳文の原言語文及び目的言語文に対して NER を行い、認識された NE を原言語と目的言語間で自動で対応付ける。そして、対応付いた NE を NE 種別記号で置き換えた対訳文を文翻訳機の教師データ、対応付いた NE の対を NE 翻訳機の教師データとして使用する。

推論時には、原言語文に対して NER を行い、認識された NE を NE 種別記号に置き換えた文を文翻訳機で翻訳する。また、認識された NE を NE 翻訳機で翻訳する。そして、文翻訳機の結果と NE 翻訳機の結果を統合することで目的言語文を生成する。

この従来モデルには大きく 2 つの問題点がある。1 つ目は、考慮できる NE が限られることである。従来モデルは、原言語と目的言語間に対応付いた NE しか活用できないため、原言語側と種類や粒度が異なる目的言語文の NE の情報は失われてしまう。また、原言語の NE タガーが用意できない場合、目的言語の NE タガーが用意できたとしても目的言語の NE を活用することはできない。2 つ目は、NE 部分を NE 種別記号に置き換えてしまうため、文翻訳機では NE の構成単語を考慮した翻訳を行えないことである。翻訳する際の手がかりから NE の構成単語の情報が失われるので、特に NE を正しく認識できなかった場合、NE 種別記号に置き換えることで翻訳性能が悪化する可能性がある。

3 提案モデル

本研究では、目的言語文中の NE の前後に、NE の種類と開始/終了情報を含む NE タグを付与して翻訳を行う、Transformer NMT モデルを提案する。図 1 に提案モデルの概要を示す。

提案モデルでは、学習時に、教師データである対訳文の目的言語文に対して NER を行い、目的言語文内の NE を特定する。そして、特定した NE の前に「<NE 種別>」、後に「</NE 種別>」という NE タグを付与する。図 1 では、「Patrick Fresacher」という NE の前後に、種別が人名である NE の開始と終了を表す「<PER>」と「</PER>」がそれぞれ付与されている。この NE タグを付与した目的言語文を用いて Transformer NMT[10] を学習する。推論時は、学習し

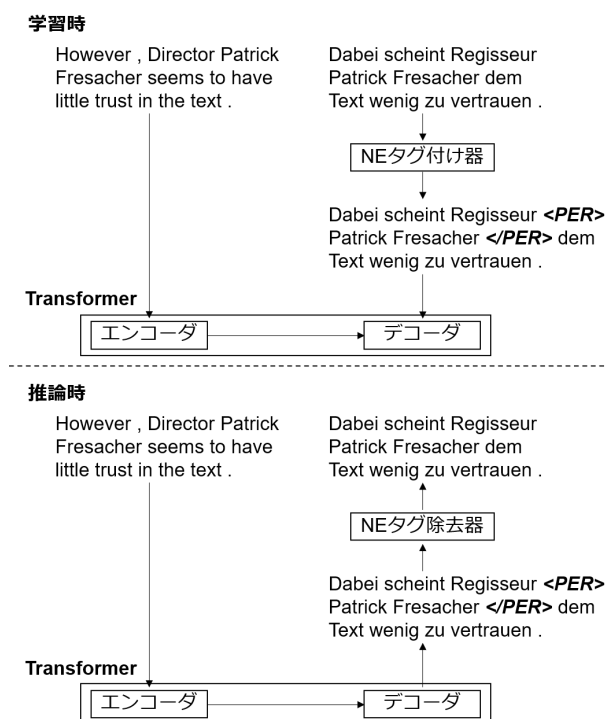


図 1 提案モデルの概要図

た Transformer NMT で原言語文を翻訳する。その結果、生成された文に NE タグが含まれている場合は、NE タグを削除した文を目的言語文として出力する。

提案モデルで原言語側の NE を活用する際は、学習時に、対訳文中の原言語文にも目的言語文と同様に NE タグを付与し、原言語文と目的言語文共に NE タグが付与された対訳文対から Transformer NMT を学習する。推論時には、原言語文に対して NER を行い NE タグを付与してから、学習した Transformer NMT で翻訳を行う。そして、生成した文に NE タグが含まれている場合には、NE タグを削除した文を目的言語文とする。

このように、提案モデルでは原言語の NER に依らず、目的言語の NER で識別された全ての NE を考慮できる。また提案モデルでは、元々の目的言語文の情報に NE の情報が追加される (NE の構成単語の情報が失わない) ため、NER の認識誤りによる悪影響を抑えることができると考えられる。

4 実験

4.1 実験設定

本実験では、提案モデルの有効性を WMT2014 の英語とドイツ語間のニュース翻訳タスク²⁾において

2) <https://www.statmt.org/wmt14/translation-task.html>

表1 教師データ中の NE の統計量

	英語	ドイツ語
文数	3,900,502	3,900,502
NE を含む文数	2,164,064	1,730,478
総サブワード数 (延べ数)	110,867,266	113,526,532
NE 部分の総サブワード数	10,089,342	8,225,938
LOC	27.06%	34.80%
ORG	38.35%	29.35%
PER	15.27%	18.35%
MISC	19.32%	17.50%

検証する。提案モデルのベースとする Transformer NMT は, Fairseq[6] の Transformer Base を使用した。この NE を考慮しないベースモデルと, ベースモデルで原言語の NE のみを考慮するモデル, 目的言語の NE のみを考慮するモデル, 原言語と目的言語の両言語の NE を考慮するモデルの 4 つのモデルの性能を評価し, 比較する。翻訳性能の評価指標は BLEU (%) を用いる。ハイパーパラメータは Vaswani ら [10] の設定に従った。英語文とドイツ語文は BPE でサブワード分割した。語彙は両言語で共有し語彙サイズは 40,000 とした。また学習は, 開発データに対する性能が 5 エポック連続で向上しなくなったら終了させた。

NE タガーは Stanza[7] を使用した。英語の NER モデルは CoNLL-2003 のデータから学習したモデル, ドイツ語の NER モデルは GermEval 2014 のデータから学習したモデルを使用した。各モデルが認識する NE は, 「LOC (地名全般)」, 「ORG (企業, 団体, 組織など)」, 「PER (架空の人物も含む人名)」, 「MISC (その他の固有表現)」の 4 種類である。表 1 に, 本実験で使用する教師データにおいて NE タガーで認識された NE の統計量を示す。表 1 より, 認識した NE の数は英語の方がドイツ語よりも多いことが分かる。

4.2 実験結果

実験結果を表 2 に示す。表 2 より, 英独翻訳と独英翻訳の両方で, NE を考慮しないモデルと原言語文の NE のみを考慮するモデルのどちらに対しても, 目的言語文の NE を考慮することで BLEU が改善できることが分かった。また, 原言語と目的言語の両言語の NE を考慮したモデルの方が, 目的言語の NE のみを考慮したモデルよりも BLEU が高くなった。この結果から, 目的言語の NE の情報と原

表2 英独/独英翻訳性能

	原言語文	目的言語文	BLEU (%)
英独	無し	無し	27.59
	無し	GermEval14	27.68
	CoNLL03	無し	27.73
	CoNLL03	GermEval14	27.81
独英	無し	無し	31.70
	無し	CoNLL03	31.77
	GermEval14	無し	31.81
	GermEval14	CoNLL03	32.16

表3 NEの有無別翻訳性能

	原言語文	目的言語文	NE 有	NE 無
英独	無し	無し	29.15	24.64
	無し	GermEval14	29.45	24.35
	CoNLL03	無し	29.33	24.70
	CoNLL03	GermEval14	29.53	24.58
独英	無し	無し	32.81	29.32
	無し	CoNLL03	32.81	29.53
	GermEval14	無し	32.76	29.74
	GermEval14	CoNLL03	33.12	30.10

言語の NE の情報を組み合わせることで翻訳性能をより改善できることが分かった。

5 考察

5.1 評価データでの NE の有無による違い

評価データの参照文に対して NER を行い, 評価データを NE を含む文と含まない文に分け, それぞれに対する翻訳性能を調べた。結果を表 3 に示す。

評価の結果, NE を含まない文に対しては, 目的言語の NE を考慮することにより, 独英翻訳では BLEU は改善したが英独翻訳では BLEU は低下した。一方, NE を含む文に対しては, 英独翻訳と独英翻訳の両方で, 提案モデルはベースラインモデルと同等もしくは高い BLEU となった。このことから, 提案モデルは狙い通り NE を含む文に対して有効であることが確認できた。しかし一方で, NE を含まない文に対しては悪影響を及ぼす場合があることが分かった。

5.2 NE の再現率

NE に対する翻訳性能を調べるために, 各モデルが生成した目的言語文に対して, 参照文で出現して

表 4 NE の再現率

	モデル		構成サブワード数				全ての NE
	原言語文	目的言語文	1	2	3	4 以上	
英独	無し	無し	81.59%	81.77%	79.03%	69.11%	77.83%
	無し	GermEval14	84.06%	81.28%	78.60%	67.67%	77.89%
	CoNLL03	無し	82.65%	81.65%	78.03%	68.15%	77.61%
	CoNLL03	GermEval14	82.65%	82.39%	79.46%	68.39%	78.17%
独英	無し	無し	81.67%	77.64%	74.62%	64.67%	74.95%
	無し	CoNLL03	83.01%	76.52%	74.11%	62.47%	74.37%
	GermEval14	無し	81.67%	77.54%	75.13%	64.43%	74.98%
	GermEval14	CoNLL03	83.42%	77.34%	74.23%	63.86%	75.06%

表 5 英語 NE タグとして OntoNotes モデル使用時の独英翻訳性能

原言語文	目的言語文	BLEU (%)
無し	無し	31.70
無し	CoNLL03	31.77
無し	OntoNotes	31.87
GermEval14	無し	31.81
GermEval14	CoNLL03	32.16
GermEval14	OntoNotes	31.83

いる NE の再現率を調査した。再現率を計算する際は、参照文の NE の文字列がモデルの出力文に含まれている場合に再現できたとみなした。結果を表 4 に示す。表 4 では、NE を構成するサブワード数毎の再現率も示す。

表 4 より、NE 全体の再現率について、英独翻訳においては目的言語文の NE を考慮することで再現率が改善したことが分かる。一方、独英翻訳においては、原言語文の NE のみを考慮したモデルと比較すると目的言語文の NE を考慮することで再現率が低下した。NE を構成するサブワード数毎の再現率をみると、1つのサブワードで構成される NE に関しては、全ての設定で、目的言語の NE を活用することにより再現率が同等あるいは改善している。そして、構成するサブワード数が増えるほど再現率自体が低下し、また、目的言語の NE を考慮することで悪化しやすくなる傾向があった。これらのことから、提案モデルは、一つのサブワードで構成される NE の翻訳には有効であることが確認できた。

5.3 使用する NER モデルの影響

本節では、提案モデルの有効性が、使用する NE タグによりどのように変化するかを考察する。

具体的には、英語の NE タグとして、OntoNotes コーパスから学習した Stanza の NER モデルを使用した際の提案モデルの独英翻訳性能を評価する。OntoNotes モデルは、CoNLL03 モデルが認識する「LOC」、「ORG」、「PER」に加えて、数を表す「CARDINAL」や日付表現を表す「DATE」など、合計 18 種類の NE を認識する。本実験の教師データに対して認識された NE 部分の総サブワード数は 14,310,222、NE を含む文数は 2,532,606 であり、CoNLL03 モデルと比較して認識された NE の割合は大きくなった。

評価結果を表 5 に示す。表 5 より、英語の NE タグとして OntoNotes モデルを使用した場合でも、目的言語文の NE を考慮する有効性を確認できた。このことから、提案モデルは NER で使用する NE の種類によらず、翻訳性能の改善に効果があることが分かった。ただし、OntoNotes モデルに変更することで認識された NE の割合は増えたにもかかわらず、原言語の NE を考慮した場合においては CoNLL03 モデルを使用した場合の方が有効であった。使用する NE タグと提案モデルの有効性の関係に関するより詳細な調査は今後行っていきたい。

6 おわりに

本研究では、目的言語文の NE の情報をタグの付与により考慮する Transformer に基づく NMT モデルを提案した。WMT2014 の英独/独英ニュース翻訳タスクにおいて、提案モデルにより翻訳性能が改善できることを確認した。また、実験結果を分析することで、提案モデルは、一つのサブワードで構成される NE の翻訳に有効であることを確認した。今後は複数のサブワードで構成される NE に対しても有効なモデルに改良していきたい。

参考文献

- [1] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *Proceedings of the 3rd International Conference on Learning Representations*, 2015.
- [2] Xiaoqing Li, Jinghui Yan, Jiajun Zhang, and Chengqing Zong. Neural name translation improves neural machine translation. In *Proceedings of the 14th China Workshop on Machine Translation*, pp. 93–100. Springer, 2018.
- [3] Zhongwei Li, Xuancong Wang, Ai Ti Aw, Eng Siong Chng, and Haizhou Li. Named-entity tagging and domain adaptation for better customized translation. In *Proceedings of the Seventh Named Entities Workshop*, pp. 41–46, 2018.
- [4] Maciej Modrzejewski, Miriam Exel, Bianka Buschbeck, Thanh-Le Ha, and Alexander Waibel. Incorporating external annotation to improve named entity translation in NMT. In *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*, pp. 45–51, 2020.
- [5] Graham Neubig, Zi-Yi Dou, Junjie Hu, Paul Michel, Danish Pruthi, and Xinyi Wang. compare-mt: A tool for holistic comparison of language generation systems. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pp. 35–41, 2019.
- [6] Myle Ott, Sergey Edunov, Alexei Baevski, Angela Fan, Sam Gross, Nathan Ng, David Grangier, and Michael Auli. fairseq: A fast, extensible toolkit for sequence modeling. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*, pp. 48–53, 2019.
- [7] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp. 101–108, 2020.
- [8] Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1715–1725, 2016.
- [9] Arata Ugawa, Akihiro Tamura, Takashi Ninomiya, Hiroya Takamura, and Manabu Okumura. Neural machine translation incorporating named entity. In *Proceedings of the 27th International Conference on Computational Linguistics*, pp. 3240–3250, 2018.
- [10] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pp. 5998–6008, 2017.