

知識グラフ埋め込みを用いたニューラル機械翻訳における エンティティ表現の改良

坂井 優介 渡辺 太郎

奈良先端科学技術大学院大学

{sakai.yusuke.sr9,taro}@is.naist.jp

藤田 篤

情報通信研究機構

atsushi.fujita@nict.go.jp

1 はじめに

ニューラル機械翻訳の課題として、訓練データ中に出現しない未知語 (OOV) や訓練データに含まれていても出現頻度が低い低頻度語に対して翻訳品質が低いことが挙げられる。この課題に対処するために近年のニューラル機械翻訳ではサブワード化を行うことで OOV 問題を解決している [1, 2, 3, 4]。しかしサブワード化を行うことは単語の表層表現しか着目しておらず、単語の意味表現を考慮して OOV 問題に対処しているとはいえない。一方、知識グラフ等の単言語資源は機械翻訳で用いられる対訳文よりも豊富で入手しやすいことを活かし、単言語資源を用いてニューラル機械翻訳の精度を向上させる取り組みが存在している。その中でも本研究では知識グラフを用いた手法について着目する。知識グラフをニューラル機械翻訳に注入する手法である KG-NMT [5] を改良し、知識グラフをニューラル機械翻訳に注入する際に、エンティティをタグに置き換える手法と、エンティティをサブワードに分割にした知識グラフをニューラル機械翻訳に注入する手法の2種類提案する。WMT の newest データセットによる英独翻訳での実験の結果、提案手法ではベースラインより BLEU スコアが高いことを確認できた。

2 関連研究

2.1 知識グラフ

知識グラフとはトリプレットと呼ばれる形で記述されており、 h をヘッドエンティティ、 t をテールエンティティ、 r を h と t の関係としたとき、(奈良先端科学技術大学院大学, alias, NAIST) のように (h, r, t) の順に記される。知識グラフを用いたタスクとして h と r が与えられたとき t を予測する知識グラフ推論が存在しており、知識グラフ埋め込み

[6, 7] や強化学習 [8]、グラフニューラルネットワーク [9, 10] などによる知識グラフ推論手法が提案されている。これらの手法は、質問応答システム [11] や推薦システム [12] などでも広く利用されている。

2.2 知識グラフの機械翻訳への応用

知識グラフを用いたニューラル機械翻訳の取り組みは主に知識グラフをデータ拡張に用いる手法と知識グラフを直接ニューラル機械翻訳に注入する手法に大別できる。データ拡張に知識グラフを用いた研究として、Zhao ら [13] は訓練データ中のエンティティを知識グラフを用いて類似したエンティティに置換することで訓練データを拡張した。

一方、知識グラフを直接ニューラル機械翻訳に注入する手法として、知識グラフ埋め込みを用いる手法 [5, 14] とニューラル機械翻訳と知識グラフ推論の同時学習を行う手法 [15] が存在する。Moussalem ら [5] は KG-NMT を提案した。KG-NMT は知識グラフ埋め込みを用いて、文章中に出現するエンティティと知識グラフ中のエンティティが一致したものに対して、Sennrich らの手法 [16] を用いて該当するエンティティに、そのエンティティの知識グラフ埋め込みを結合することで、知識グラフ埋め込みの注入を行った。KG-NMT は訓練データ中に出現しないエンティティに対しても、知識グラフ埋め込みを用いているため、エンティティ間の関係も含めて注入することができている。しかしこれらの手法では知識グラフ中のエンティティと訓練データ中のエンティティが一致した場合にのみしか有効ではない。

D をニューラル機械翻訳の入力文の集合としたとき、 D に含まれるエンティティの集合を $V(D)$ と表す。また知識グラフ K を関係のインスタンス (h, r, t) の集合としたとき、知識グラフ K 中のエンティティの集合は $V(K) = \{h | (h, *, *) \in K\} \cup \{t | (*, *, t) \in K\}$ となる。

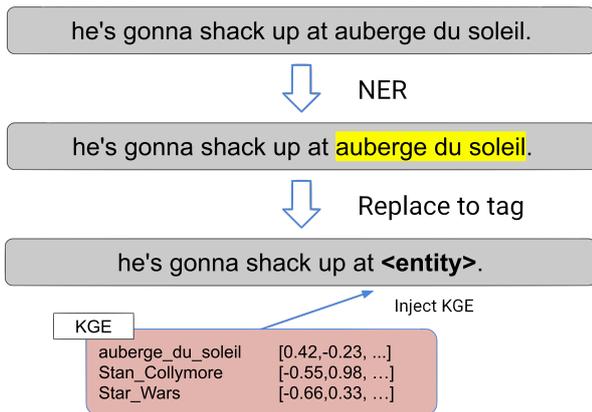


図 1 KG-tag NMT の手続きの概要

このとき実際のエンティティは、 U をすべてのエンティティの集合としたとき、ニューラル機械翻訳の訓練データにのみ含まれるエンティティの集合 ($V(D) - V(K)$)、知識グラフのみに含まれるエンティティの集合 ($V(K) - V(D)$)、どちらにも出現するエンティティの集合 ($V(D) \cap V(K)$)、どちらにも出現しないエンティティの集合 ($U - (V(D) \cup V(K))$) の計 4 種類の集合が存在する。KG-NMT では $V(D) \cap V(K)$ の集合しか扱うことができなかつた。また KG-NMT ではエンティティ単位で訓練データと知識グラフの対応付けを行うため、サブワード化ができず、扱う語彙が大きくなるという問題が生じる。

Zhao ら [15] はそれら 4 つのエンティティの集合についてもニューラル機械翻訳へ活用するために、エンティティをサブワード化したあと、ニューラル機械翻訳と知識グラフ推論の学習を同時に行った。しかし Zhao らの手法はニューラル機械翻訳と知識グラフ推論の学習を同時に行っているため、新たな言語対に対するモデルを学習する際、ニューラル機械翻訳モデルだけでなく、知識グラフ埋め込みも再度学習する必要がある。よって Zhao らの手法で N 言語間の多言語機械翻訳を行おうとすると、知識グラフ推論の学習を $N(N-1)$ 回行うことになる。そのため、Zhao らの手法は学習コストが高い上にニューラル機械翻訳と知識グラフ推論モデルを組み合わせることは容易ではない。

3 知識グラフ埋め込みを利用したニューラル機械翻訳の改良

本研究では KG-NMT を発展させ、エンティティをタグに置き換える手法と知識グラフ埋め込みの作成時にエンティティのサブワードに基づいて学習し

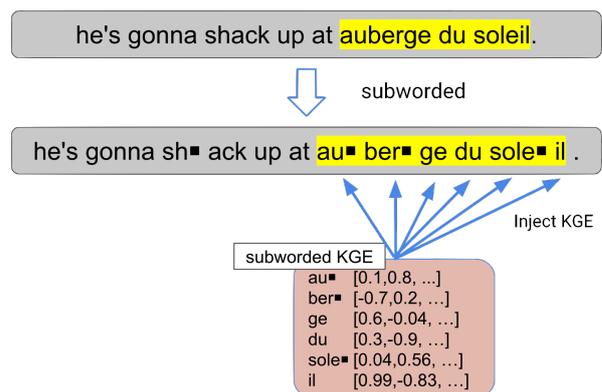


図 2 Subworded KG-NMT の手続きの概要

た知識グラフ埋め込みをニューラル機械翻訳に注入する手法の 2 種類を提案する。

3.1 KG-NMT におけるエンティティの抽象化

KG-NMT の問題を解決するために、本研究では KG-tag NMT を提案する。KG-tag NMT は未知語処理の手法である Back-off モデル [18, 19] を参考にし、データセット内の該当するエンティティに知識グラフに基づく埋め込みの情報を注入後、そのエンティティを特殊タグに置き換える。具体的な手続きの例を図 1 に示す。図 1 のように知識グラフを注入後、<entity> という特殊なタグに置き換える。KG-tag NMT では知識グラフ情報を注入後にサブワード化が行えるため、KG-NMT の課題である語彙数を削減できる。しかし KG-tag NMT では依然として $V(D) \cap V(K)$ 内のエンティティしか扱えない。

3.2 サブワードに基づく知識グラフ埋め込み

サブワードに基づく知識グラフ埋め込みを行うために Subworded KG-NMT を提案する。Subworded KG-NMT はすべてのエンティティを扱うために、知識グラフ内のエンティティ $V(K)$ のサブワード化を行う。

サブワード化した知識グラフの埋め込み表現を作成する方法として Transformer [17] を使用する。知識グラフ推論を seq2seq の枠組みで行い、作成された埋め込み表現を知識グラフ埋め込みとする。エンティティ h, t を SentencePiece [4] を用いてサブワード化したあと、 h と r が与えられた際に t を推論するタスクを、 h と r を原言語の入力文として t を目的言語の出力文を得るタスクとみなす。知識グラフ

表 1 BLEU による各モデルの比較

	news2012	news2013	news2014	news2015	news2016	news2017	news2018	WMT2017
Transformer [17]	17.81	20.36	19.53	21.28	23.20	20.25	25.24	28.66
KG-NMT [5]	16.74	19.06	18.60	19.49	20.66	18.21	22.55	22.79
KG-tag NMT	18.20	20.51	19.47	21.60	23.15	20.38	25.07	28.87
subworded KG-NMT	17.78	20.73	19.75	23.30	23.91	21.87	26.81	29.55

K が与えられた際の損失関数は式 (1) となる。

$$L(K; \theta_e, \theta_d) = \sum_{(e,r,t) \in K} \log p(t | h, r; \theta_e, \theta_d) \quad (1)$$

ここで、 θ_e と θ_d はそれぞれエンコーダとデコーダのパラメータである。

エンティティをサブワード化することにより、知識グラフ K に含まれないエンティティに対しても埋め込みを得ることができる。またニューラル機械翻訳と知識グラフ埋め込みのエンティティの対応付けはサブワード単位で行う。Subworded KG-NMT の具体的な手続きの例を図 2 に示す。図 2 のようにサブワード化した知識グラフ埋め込みをニューラル機械翻訳に注入することで、エンティティ D に対する知識グラフの注入量を向上させる。知識グラフ埋め込みの注入手法は KG-NMT [5] に倣う。

知識グラフ埋め込みをサブワード化する手法は Zhao らの手法 [15] と似ているが、Zhao らは知識グラフ推論とニューラル機械翻訳を同時に学習することによって、知識グラフをニューラル機械翻訳に活用した。一方、Subworded KG-NMT では知識グラフ推論とニューラル機械翻訳をそれぞれ独立して学習している。Subworded KG-NMT はニューラル機械翻訳と知識グラフ推論を独立して学習するため、新たな目標言語へのニューラル機械翻訳を学習する際に、以前学習した知識グラフ埋め込みの再利用が可能となる。そのため N 言語間の多言語機械翻訳を行うときは Subworded KG-NMT では知識グラフの学習を N 回行うだけでよい。この利点は Zhao らの手法 [15] の欠点を補っている。

4 実験

4.1 実験設定

実験は英独翻訳を対象とし、訓練データとして JRC-Acquis3.0 と Europarl を用い、検証用データに WMT の newstest2011、テストデータに WMT の newstest2012 から 2018 までを用いた。知識グラフは英語版の DBpedia [20] を用いた。また訓練データ

に WMT2017、検証用データに newstest2016、テストデータに newstest2017 を用いることにより同一ドメイン内での翻訳についても評価を行った。

Transformer の実装には fairseq¹⁾ と OpenNMT-py²⁾ を用いた。またサブワード化には SentencePiece³⁾ を用いた。SentencePiece の学習にはニューラル機械翻訳の訓練データを用いた。また知識グラフ中のエンティティに対するサブワード化と訓練データに対するサブワード化は同じ SentencePiece のモデルを使用した。知識グラフ埋め込みは Transformer を用いて作成した。NER は spaCy⁴⁾ を使用し、DBpedia 内のエンティティを辞書として用いた。

既存研究のベースラインとして Transformer [17] と KG-NMT [5] を用いた。なお KG-NMT についてサブワード化してしまうと、エンティティに知識グラフ情報を注入することができなくなるため、サブワード化は行わなかった。評価には SacreBLEU [21]⁵⁾ を用いた。

4.2 実験結果

表 1 に実験結果を示す。2 つの提案手法はすべてのテストデータに対して KG-NMT より BLEU スコアが向上している。これは KG-NMT では行えなかったサブワード化が提案手法では可能になったことより、未知語として処理される単語数が減少したため BLEU スコアが上昇したと考えられる。KG-tag NMT では KG-NMT の手法を改良し、エンティティをタグに置き換えることでサブワード化を行えるようにした。そのため、Transformer と同等の精度を保ちつつ、 $V(K) \cap V(D)$ エンティティについては知識グラフを活用することができるため、BLEU スコアが向上したと考えられる。

一方 Subworded KG-NMT では、翻訳の訓練データ中のエンティティ $V(D)$ と知識グラフ中のエンティ

1) <https://github.com/pytorch/fairseq>

2) <https://github.com/OpenNMT/OpenNMT-py>

3) <https://github.com/google/sentencepiece>

4) <https://spacy.io/>

5) <https://github.com/mjpost/sacrebleu>

表2 テストセットのエンティティ $V(D)$ に対する $V(K) \cap V(D)$ エンティティの数と割合 $V(K) \cap V(D)/V(D)$

	# $V(D)$	# $V(K) \cap V(D)$	$V(K) \cap V(D)/V(D)$ (%)
2012	9404	198	2.11
2013	8849	204	2.31
2014	8852	115	1.30
2015	7367	120	1.63
2016	8896	130	1.46
2017	8796	178	2.02
2018	9413	107	1.14

ティ $V(K)$ が完全に一致していなくてもサブワード単位で一致していれば知識グラフ埋め込み情報を注入できる。そのためエンティティ $V(D)$ に対して知識グラフを注入可能な範囲が拡大するので、KG-tag NMT よりも多くの知識グラフをニューラル機械翻訳に注入することができる。このことから全体としての知識グラフの注入量が向上した結果、Subworded KG-NMT の BLEU スコアが向上したと考えられる。

4.3 エンティティの分析

提案手法と Transformer [17] の BLEU スコアを比較すると、KG-tag NMT ではスコアの向上は見られるものの、知識グラフ埋め込みを注入したことにより、むしろスコアが低下しているテストデータが存在している。そこで KG-tag NMT によってエンティティ $V(K)$ とエンティティ $V(D)$ の対応付けが行えたエンティティ $V(K) \cap V(D)$ の数を調べた。その結果を表2に示す。表2よりエンティティ $V(K) \cap V(D)$ はエンティティ $V(D)$ の 1~2%ほどしか存在していないことがわかる。また既存手法の KG-NMT と Transformer を比較すると、 $V(K) \cap V(D)/V(D)$ の割合によって BLEU スコアの差が大きくなる傾向がある。これは KG-NMT がサブワード化を行うことによって OOV 問題を解決するのではなく、知識グラフを用いて OOV 問題の解決を図っていることに起因している。そのため知識グラフの注入量と翻訳品質には相関があると考えられる。

5 Transformer による知識グラフ推論の精度

Transformer による知識グラフ推論の精度について考察するための実験を行った。知識グラフ推論用データセットに YAGO3-10 [22] を使用し、知識グラフ推論の比較対象モデルを ConvKB [6] と ComplEx [7] とした。Transformer による知識グラフ推論では

表3 知識グラフ推論精度の比較

	HITS@1	HITS@3	HITS@10
ConvKB [6]	.213	.349	.500
ComplEx [7]	.260	.391	.562
Transformer (w/o subword)	.078	.130	.564
Transformer (with subword)	.102	.168	.291

エンティティのサブワード化の有無による比較も行った。実験結果を表3に示す。

HITS@ T はそれぞれ推論したエンティティのうち上位 T 番目以内に正解エンティティが含まれる割合を示している。表3の結果より、seq2seq での知識グラフ推論はエンティティをサブワード化しない場合、既存手法である ConvKB [6] や ComplEx [7] と同等の性能であることがわかった。またサブワード化を行った場合でも、上位 10 位以内に正解エンティティが含まれる割合が3割ほどあり、サブワード化を行っても知識グラフを学習できているといえる。そのため Transformer による seq2seq での知識グラフ推論は有効であるといえる。

知識グラフ推論の既存手法ではデータセット中に含まれていない未知のエンティティを推論することができないが、サブワード化し、seq2seq で知識グラフ推論を行った場合エンティティを生成するため、未知のエンティティを推論することができる利点がある。そのため、潜在的に $U - (V(D) \cup V(K))$ に対しても知識グラフ情報を学習していると予測できるので、今後その予測を検証していく。

6 まとめ

本研究ではニューラル機械翻訳の品質向上を目的とし、OOV 問題に対処するために知識グラフを注入する手法を2種類提案した。実験を行った結果、提案手法では BLEU スコアが既存手法よりも向上していることを確認した。また知識グラフの注入量によって翻訳精度が変化するため、単言語データである知識グラフの注入量を増やすことにより、翻訳精度が向上する可能性があることがわかった。さらに seq2seq による知識グラフ推論の有用性も確認した。その結果、未知のエンティティに対する推論可能性をもちつつ、既存手法と同等の精度で推論することができた。今後の課題として、多言語ニューラル機械翻訳への適用を予定している。

参考文献

- [1] 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)*, pages 1715–1725, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [2] Minh-Thang Luong and Christopher D. Manning. Achieving open vocabulary neural machine translation with hybrid word-character models. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1054–1063, Berlin, Germany, August 2016.
- [3] Junyoung Chung, Kyunghyun Cho, and Yoshua Bengio. A character-level decoder without explicit segmentation for neural machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1693–1703, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [4] Taku Kudo and J. Richardson. Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing. In *EMNLP*, 2018.
- [5] Diego Moussallem, Axel-Cyrille Ngonga Ngomo, Paul Buitelaar, and Mihael Arcan. Utilizing knowledge graphs for neural machine translation augmentation. *K-CAP '19*, page 139–146, New York, NY, USA, 2019. Association for Computing Machinery.
- [6] Dai Quoc Nguyen, Tu Dinh Nguyen, Dat Quoc Nguyen, and Dinh Phung. A novel embedding model for knowledge base completion based on convolutional neural network. In *Proceedings of the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 327–333, 2018.
- [7] C. Demir and Axel-Cyrille Ngonga Ngomo. Convolutional complex knowledge graph embeddings. *ArXiv*, abs/2008.03130, 2020.
- [8] Wenhan Xiong, Thien Hoang, and William Yang Wang. DeepPath: A reinforcement learning method for knowledge graph reasoning. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 564–573, Copenhagen, Denmark, September 2017. Association for Computational Linguistics.
- [9] Deepak Nathani, Jatin Chauhan, Charu Sharma, and Manohar Kaul. Learning attention-based embeddings for relation prediction in knowledge graphs. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2019.
- [10] Rui Wang, Bicheng Li, Shengwei Hu, Wenqian Du, and Min Zhang. Knowledge graph embedding via graph attenuated attention networks. *IEEE Access*, PP:1–1, 12 2019.
- [11] Antoine Bordes, Nicolas Usunier, Sumit Chopra, and Jason Weston. Large-scale simple question answering with memory networks. *CoRR*, abs/1506.02075, 2015.
- [12] Fuzheng Zhang, Nicholas Jing Yuan, Defu Lian, Xing Xie, and Wei-Ying Ma. Collaborative knowledge base embedding for recommender systems. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, page 353–362, New York, NY, USA, 2016. Association for Computing Machinery.
- [13] Yang Zhao, Jiajun Zhang, Yu Zhou, and Chengqing Zong. Knowledge graphs enhanced neural machine translation. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 4039–4045. International Joint Conferences on Artificial Intelligence Organization, 7 2020. Main track.
- [14] Yu Lu, Jiajun Zhang, and Chengqing Zong. Exploiting knowledge graph in neural machine translation. In Jiajun Chen and Jiajun Zhang, editors, *Machine Translation*, pages 27–38, Singapore, 2019. Springer Singapore.
- [15] Yang Zhao, Lu Xiang, Junnan Zhu, Jiajun Zhang, Yu Zhou, and Chengqing Zong. Knowledge graph enhanced neural machine translation via multi-task learning on sub-entity granularity. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4495–4505, Barcelona, Spain (Online), December 2020. International Committee on Computational Linguistics.
- [16] Rico Sennrich and Barry Haddow. Linguistic input features improve neural machine translation. In *Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers*, pages 83–91, Berlin, Germany, August 2016. Association for Computational Linguistics.
- [17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17*, page 6000–6010, Red Hook, NY, USA, 2017. Curran Associates Inc.
- [18] S. Jean, Kyunghyun Cho, R. Memisevic, and Yoshua Bengio. On using very large target vocabulary for neural machine translation. *ArXiv*, abs/1412.2007, 2015.
- [19] Thang Luong, Ilya Sutskever, Quoc Le, Oriol Vinyals, and Wojciech Zaremba. Addressing the rare word problem in neural machine translation. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 11–19, Beijing, China, July 2015. Association for Computational Linguistics.
- [20] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In *The Semantic Web*, pages 722–735, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.
- [21] Matt Post. A call for clarity in reporting BLEU scores. In *Proceedings of the Third Conference on Machine Translation: Research Papers*, pages 186–191, Brussels, Belgium, October 2018. Association for Computational Linguistics.
- [22] F. Mahdisoltani, J. Biega, and Fabian M. Suchanek. Yago3: A knowledge base from multilingual wikipedias. In *CIDR*, 2015.