ニューラル機械翻訳のための強化学習における様々な評価指標報酬の調査

Studies on Evaluation Metrics as Rewards for Reinforcement Learning in Neural Machine Translation

愛媛大学　大学院理工学研究科教授
二宮　崇

2001 年東京大学大学院理学系研究科情報科学専攻博士課程修了。博士 (理学)。2017 年より現職。自然言語処理の研究に従事。

愛媛大学　大学院理工学研究科
中谷　祐貴

2021 年愛媛大学工学部情報工学科卒業。2023 年同大学院理工学研究科博士前期課程修了。自然言語処理の研究に従事。

愛媛大学　大学院理工学研究科助教
梶原　智之

2018 年首都大学東京大学院システムデザイン研究科博士後期課程修了。博士 (工学)。大阪大学データビリティフロンティア機構の特任助教を経て、2021 年より現職。言い換え生成や品質推定など自然言語処理の研究に従事。

1　はじめに

深層学習に基づく系列変換モデル [1] は流暢な文生成が可能であり、機械翻訳やテキスト平易化など、多くのテキスト生成タスクで成功を収めている。テキスト生成に関する先行研究の多くは、出力文と参照文の間のクロスエントロピー損失を用いて、トークン単位での最尤推定によって系列変換モデルを訓練している。クロスエントロピー損失の微分可能性は、教師あり学習の枠組みでの勾配計算に有効ではあるが、柔軟性に欠ける。つまり、意味的に妥当な出力文が、参照文との表層的な不一致によって、不当に低評価を受ける場合がある。

このような Loss-Evaluation Mismatch 問題 [2] [3] は、強化学習 [4] による評価指標への直接的な最適化によって対処できる。強化学習の報酬には、系列変換モデルのパラメタで微分不可能な関数を用いることができるため、単語 N-gram 単位の評価指標である BLEU [5] や文単位の評価指標である BLEURT [6] など、任意の評価指標を採用できる。強化学習を用いることで、機械翻訳 [2] [7] [8] やテキスト平易化 [9] [10] などの深層学習に基づくテキスト生成の性能改善が報告されている。

機械翻訳においては、多くの先行研究 [2] [11] [7] [12] が強化学習の報酬計算に BLEU を用いているが、BLEU は人手評価との相関が充分に高いわけではない。機械翻訳の自動評価タスク [13] では、表層マッチングに基づく chrF [14] や BERT [15] に基づく手法 [16] [17] [6] など、BLEU よりも高い人手評価との相関を持つ評価指標が提案されている。そのため、これらの評価指標を用いた報酬計算によって、機械翻訳の強化学習を更に改善できる可能性がある。

本稿は、国際ワークショップ「The 9th Workshop on Asian Translation（WAT 2022)」において我々が提案したニューラル機械翻訳強化学習のための様々な評価指標報酬に関する研究 [18] について解説する。本研究では、図 1 に示す強化学習の枠組みで Transformer ベースの機械翻訳モデル [1] を訓練する。ただし、機械翻訳の強化学習では、数万トークンからなる語彙を扱うために、行動空間が非常に大きい。そのため、先行研究

[2] [7] と同様に、クロスエントロピー損失最小化の事前訓練を経た機械翻訳モデルに対しての再訓練として強化学習を適用する。そして、報酬計算と性能評価の両方において複数の評価指標を検討し、機械翻訳の強化学習に適した報酬関数について調査する。

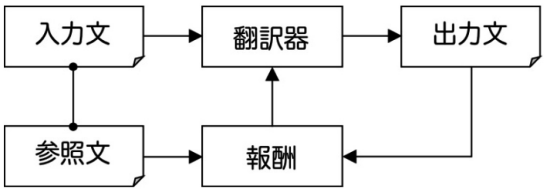

図 1　強化学習に基づく機械翻訳

IWSLT-2014 の独英翻訳タスク [19] における評価実験の結果、BLEU を報酬とする強化学習では BLEU および chrF の表層マッチングに基づく評価指標しか改善できないことが明らかになった。一方で、文間の意味的類似度推定（STS：Semantic Textual Similarity）タスク [20] において訓練した BERT [8] や BLEURT [6] など、BERT に基づく一部の評価指標を報酬とする機械学習では、様々な評価指標が改善された。

2　機械翻訳の強化学習

本研究では、様々な評価指標を報酬とする深層強化学習によって、事前訓練済みの機械翻訳モデルを再訓練する。まず 2.1 節で機械翻訳モデルの事前訓練について説明し、次に 2.2 節で強化学習による再訓練について述べ、最後に 2.3 節で強化学習に用いる報酬としての機械翻訳の評価指標を概説する。

2.1.　事前訓練

ニューラル機械翻訳モデルは、入力文を符号化するエンコーダと出力文を生成するデコーダからなる系列変換モデルとして構成される。エンコーダは、入力文のトークン列 $x = (x_1,\cdots,x_K)$ が与えられ、隠れ状態 $h = (h_1,\cdots,h_K)$ を出力する。デコーダは、エンコーダによって生成された隠れ状態 h が与えられ、出力文のトークン列 $y = (y_1,\cdots,y_M)$ を 1 つずつ出力する。トークン y_t の生成確率は、x および $y_{<t}=(y_1,\cdots,y_{t-1})$ を条件として最大化され、N 個の対訳文 $(x^i, y^i)_{i=1,\cdots,N}$ が与えられると、各対訳文対 (x, y) の対数尤度を次のように計算する。

$$\log p(y|x) = \sum_{t=1,\cdots,M} \log p(y_t|y_{<t}, x)$$

事前訓練では、入力文 x および長さ M 以下の出力文 y からなる N 個の対訳文 $(x^i, y^i)_{i=1,\cdots,N}$ に対して、各対訳文対のクロスエントロピー損失の合計を最小化する。各対訳文対 (x, y) のクロスエントロピー損失は次式で表せられる。

$$L_X = -\sum_{t=1,\cdots,M} y_t \log p(y_t|y_{<t}, x)$$

2.2.　再訓練

強化学習に基づく機械翻訳モデルの再訓練のために、REINFORCE [4] を用いる。REINFORCE は方策勾配アルゴリズムの一種であり、機械翻訳モデルが報酬の期待値を最大化するように訓練される。

再訓練の損失関数は、対数尤度を報酬で重み付けすることで求められる。

$$L_R = \sum_{t=1,\cdots,T} (R(z_1,\cdots,z_T) - R_b)\log p(z_t|z_{<t})$$

ここで、R は報酬関数、R_b はベースライン報酬、$z_1,\cdots,z_T$ はデコーダからの出力文である。本研究では、ベースライン報酬としてミニバッチ内の平均報酬を使用する。

また、訓練を安定させるために、先行研究 [7] と同様に強化学習の際に以下の損失関数を使用する。

$$L = \lambda L_X + (1 - \lambda)L_R$$

2.3.　強化学習の報酬

本研究では、以下の評価指標を強化学習の報酬として用いる。

- BLEU[1] [5]：単語 N-gram の一致率を用いて、出力文と参照文の表層的な類似度を評価する。
- chrF[1] [14]：文字 N-gram と単語 N-gram の F 値を用いて、出力文と参照文の表層的な類似度を評価する。
- STS BERT [8]：STS タスク [20] において再訓練した BERT [15] を用いて、出力文と参照文の意味的な類似度を評価する。
- Sentence BERT[2] [21]：自然言語推論（NLI：Natural Language Inference）タスク [22] において再訓練した BERT を用いて、出力文と参照文の意味的な類似度を評価する。

1　https://github.com/mjpost/sacrebleu

2　https://huggingface.co/sentence-transformers/all-mpnet-base-v2

表 1　IWSLT-2014 De → En タスクにおける機械翻訳の強化学習の性能
（太字は強化学習による改善，下線は最高値を示す）

報酬	BLEU	Sent. BERT	BERT Reg.	SimCSE	chrF	BERT Score	BLEURT	STS BERT
なし	33.73	75.66	0.0478	82.10	54.27	58.47	0.0639	3.654
BLEU	**34.26**	74.91	0.0202	81.93	**54.39**	58.01	0.0234	3.641
Sent. BERT	**33.78**	**75.79**	**0.0513**	**82.24**	**54.38**	**58.72**	**0.0649**	**3.656**
BERT Reg.	33.47	**75.80**	**0.0557**	**82.32**	54.25	**58.64**	**0.0681**	3.650
SimCSE	33.73	**75.84**	**0.0512**	**82.25**	**54.37**	**58.76**	**0.0669**	**3.659**
chrF	**33.90**	**75.81**	**0.0517**	**82.24**	**54.45**	**58.69**	**0.0671**	**3.657**
BERTScore	**33.96**	**75.80**	**0.0511**	**82.30**	**54.48**	**58.80**	**0.0677**	**3.658**
BLEURT	**33.85**	**75.90**	**0.0572**	**82.33**	**54.44**	**58.92**	**0.0759**	**3.660**
STS BERT	**34.09**	**76.11**	**0.0528**	**82.52**	**54.62**	**59.10**	**0.0700**	**3.684**

- SimCSE[3] [23]：NLI コーパス中の含意関係にある文対を正例として対照学習した BERT を用いて、出力文と参照文の意味的な類似度を評価する。
- BERTScore[4] [17]：RoBERTa（Large）から得られる文脈化トークン埋め込みの最大マッチングを用いて、出力文と参照文の意味的な類似度を評価する。
- BERT Regressor [16]：機械翻訳の自動評価タスク [13] において再訓練した BERT を用いて、出力文と参照文の意味的な類似度を評価する。
- BLEURT[5] [6]：折返翻訳などによって自動生成した単言語パラレルコーパス上で再訓練し、さらに機械翻訳の自動評価タスクにおいて再訓練した BERT を用いて、出力文と参照文の意味的な類似度を評価する。

3　評価実験

3.1.　実験設定

事前訓練と強化学習の再訓練の両方で IWSLT-2014 の独英タスク [19] を用いた。訓練用データは 159,392 文対、検証用データは 7,245 文対、評価用データは 6,750 文対である。

機械翻訳モデルには Transformer [1] を使用し、レイヤ数を 6、ヘッド数を 4、次元数を 256、ドロップアウト率を 0.3 とした。事前訓練では、最適化手法を Adam [24]（学習率は 0.0003）、バッチサイズを 2,048 とし、検証用データにおける BLEU による early stopping によって訓練を停止した。強化学習では、最適化手法を Adam（学習率は 0.00001）、$\lambda = 0.3$、バッチサイズを 512 とし、報酬として用いる評価指標による early stopping によって訓練を停止した。実装には Reinforce-Joey[6] [12] を用いた。

報酬計算および性能評価のための評価指標には、2.3 節のツールを用いた。ただし、STS BERT [8] および BERT Regressor [16] については、Hugging Face Transformers[7] [25] の $\text{BERT}_{\text{BASE}}$[8] を用いて実装した。

3.2.　実験結果

表 1 に実験結果を示す。1 行目の“報酬なし”は、強化学習を行わずに事前訓練のみを行ったベースラインである。このベースラインと 2 行目以降の強化学習の比較から、報酬に用いるのと同じ評価指標で性能を評価した際には、どの手法でも強化学習によって性能が向上することがわかる。

報酬として BLEU を用いた場合には、強化学習によって BLEU および chrF の表層マッチングに基づく評価指標しか改善されておらず、その他の BERT ベースの評価指標では性能が悪化している。一方で、同じく表層マッチングに基づく chrF を報酬として用いた場合には、強化学習によって全ての評価指標が改善されている。

BERT ベースの報酬のうち、Sentence BERT による強化学習は全体的にベースラインモデルとの差分が少なく、効果が少ないことがわかる。また、SimCSE

3　https://huggingface.co/princeton-nlp/sup-simcse-roberta-large
4　https://github.com/Tiiiger/bert_score
5　https://storage.googleapis.com/bleurt-oss/bleurt-large-512.zip
6　https://github.com/samuki/reinforce-joey
7　https://github.com/huggingface/transformers
8　https://huggingface.co/bert-base-uncased

表 2　WMT-2017 Metrics タスクにおける人手評価とのピアソン相関（太字は最高値）

	cs-en	de-en	fi-en	lv-en	ru-en	tr-en	zh-en	平均
BLEU	0.412	0.413	0.565	0.393	0.460	0.531	0.524	0.471
chrF	0.517	0.531	0.671	0.525	0.599	0.607	0.591	0.577
STS BERT	0.535	0.597	0.667	0.637	0.611	0.589	0.608	0.606
Sent. BERT	0.632	0.621	0.692	0.685	0.690	0.657	0.635	0.659
SimCSE	0.696	0.628	0.684	0.696	0.713	0.660	0.672	0.678
BERTScore	0.710	0.745	0.833	0.756	0.746	0.751	0.775	0.759
BERT Reg.	0.712	0.732	0.858	0.804	0.775	0.789	0.765	0.776
BLEURT	**0.845**	**0.845**	**0.870**	**0.865**	**0.861**	**0.846**	**0.860**	**0.856**

表 3　評価指標同士でのピアソンの相関係数

	BLEU	STS BERT	chrF	SimCSE	Sent. BERT	BERT Reg.	BLEURT	BERT Score	平均
BLEU	–	0.449	0.788	0.417	0.428	0.517	0.496	0.641	0.534
STS BERT	0.449	–	0.671	0.772	0.788	0.648	0.665	0.636	0.661
chrF	0.788	0.671	–	0.616	0.635	0.608	0.613	0.715	0.664
SimCSE	0.417	0.772	0.616	–	0.856	0.653	0.717	0.664	0.671
Sent. BERT	0.428	0.788	0.635	0.856	–	0.674	0.712	0.662	0.679
BERT Reg.	0.517	0.648	0.608	0.653	0.674	–	0.866	0.798	0.681
BLEURT	0.496	0.665	0.613	0.717	0.712	0.866	–	0.805	0.696
BERTScore	0.641	0.636	0.715	0.664	0.662	0.798	0.805	–	0.703

を報酬とする強化学習では BLEU の改善が見られず、BERT Regressor を報酬とする強化学習ではベースラインモデルよりも BLEU が悪化した。

BERT ベースの報酬のうち、BERTScore、BLEURT、STS BERT を用いることで、今回検証した全ての評価指標において性能が向上することを確認できた。特に、STS BERT は過半数の評価指標において最高性能を達成しており、機械翻訳の強化学習に最も適した報酬関数であると言える。

4　分析

4.1.　評価指標のメタ評価

本節では、表 1 の実験において強化学習の報酬として有効であった評価指標が、機械翻訳の人手評価と高い相関を持つのかを検証する。本分析では、WMT-2017 の自動評価タスク [13] における to-English 言語対を対象に、評価指標と人手評価のピアソン相関を調査する。本タスクは、cs-en・de-en・fi-en・lv-en・ru-en・tr-en・zh-en の 7 言語対が対象で、各 560 文対（出力文と参照文）に人手評価が付与されている。

分析の結果を表 2 に示す。BERT ベースの評価指標が、BLEU および chrF の表層マッチングの評価指標よりも人手評価との高い相関を持つことがわかる。特に、BLEURT が全ての言語対において人手評価との最高の相関を示している。しかし、想定とは異なり、強化学習の報酬として最適であった STS BERT は、人手評価との相関は低かった。

4.2.　評価指標間の相関関係

本節では、評価指標間の相関関係が強化学習の性能評価に影響を与えているのかを検証する。4.1 節と同様に WMT-2017 の自動評価タスクにおける to-English 言語対を対象として、本節では評価指標間のピアソン相関を調査する。

分析の結果を表 3 に示す。まず、BLEU と他の評価指標の相関が低いことがわかる。単語 N-gram を扱う chrF やトークン埋め込みのマッチングに基づく BERTScore との相関は比較的高いものの、文単位で評価を行う他の評価指標との相関が低いことから、BLEU が文単位での大域的な評価に適していない可能性が示唆される。この特性が、表 1 および表 2 において BLEU の性能が低いことに影響を与えている可能性がある。

STS BERT は、他の評価指標との相関が比較的低い傾向が見られた。つまり、表 1 の実験において STS

BERT を報酬とする強化学習が多くの評価指標から高評価を得たことは、報酬と評価指標の相性の問題ではないと考えられる。

5 おわりに

本研究では、報酬計算と性能評価の両方において複数の評価指標を検討することで、機械翻訳の強化学習に適した報酬について調査した。IWSLT-2014 の独英翻訳における実験の結果、STS タスクにおいて訓練した BERT を報酬とする強化学習により、多くの評価指標において性能を改善できることが明らかになった。ただし、STS BERT は WMT-2017 の自動評価タスクにおける人手評価との相関は比較的低く、この観点からは良い評価指標とは言えない。また、STS BERT と他の評価指標との相関も比較的低いため、報酬と評価指標の相性が良いとも言えない。なお、BERTScore および BLEURT は、人手評価との相関も良好で、他の評価指標との相関も比較的高く、かつ強化学習の報酬にも適した評価指標であると言える。

謝辞

本稿は、国際ワークショップ「The 9th Workshop on Asian Translation (WAT 2022)」に採択された論文 [18] に基づいて、それらの論文を再構成し、解説したものである。

これらの研究成果は JSPS 科研費（若手研究、課題番号：JP20K19861）および国立研究開発法人情報通信研究機構（NICT）の委託研究（課題番号：225）により得られたものである。ここに謝意を表する。

参考文献

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser and I. Polosukhin. (2017). Attention is All you Need, in Advances in Neural Information Processing Systems 30, pp. 5998–6008.

[2] M. Ranzato, S. Chopra, M. Auli and W. Zaremba. (2016). Sequence Level Training with Recurrent Neural Networks, in Proceedings of the 4th International Conference on Learning Representations.

[3] S. Wiseman and A. M. Rush. (2016). Sequence-to-Sequence Learning as Beam-Search Optimization, in Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 1296–1306.

[4] R. J. Williams. (1992). Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning, Machine Learning, Vol. 8, pp. 229–256.

[5] K. Papineni, S. Roukos, T. Ward and W.-J. Zhu. (2002). BLEU: a Method for Automatic Evaluation of Machine Translation, in Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 311–318.

[6] T. Sellam, D. Das and A. Parikh. (2020). BLEURT: Learning Robust Metrics for Text Generation, in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pp. 7881–7892.

[7] K. Hashimoto and Y. Tsuruoka. (2019). Accelerated Reinforcement Learning for Sentence Generation by Vocabulary Prediction, in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 3115–3125.

[8] G. Yasui, Y. Tsuruoka and M. Nagata.

(2019). Using Semantic Similarity as Reward for Reinforcement Learning in Sentence Generation, in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pp. 400–406.

[9] X. Zhang and M. Lapata. (2017). Sentence Simplification with Deep Reinforcement Learning, in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 584–594.

[10] A. Nakamachi, T. Kajiwara and Y. Arase. (2020). Text Simplification with Reinforcement Learning Using Supervised Rewards on Grammaticality, Meaning Preservation, and Simplicity, in Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing: Student Research Workshop, pp. 153–159.

[11] L. Wu, F. Tian, T. Qin, J. Lai and T.-Y. Liu. (2018). A Study of Reinforcement Learning for Neural Machine Translation, in Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 3612–3621.

[12] S. Kiegeland and J. Kreutzer. (2021). Revisiting the Weaknesses of Reinforcement Learning for Neural Machine Translation, in Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 1673–1681.

[13] O. Bojar, Y. Graham and A. Kamran. (2017). Results of the WMT17 Metrics Shared Task, in Proceedings of the Second Conference on Machine Translation, pp. 489–513.

[14] M. Popović. (2017). chrF++: words helping character n-grams, in Proceedings of the Second Conference on Machine Translation, pp. 612–618.

[15] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171–4186.

[16] H. Shimanaka, T. Kajiwara and M. Komachi. (2019). Machine Translation Evaluation with BERT Regressor, arXiv:1907.12679.

[17] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger and Y. Artzi. (2020). BERTScore: Evaluating Text Generation with BERT, in Proceedings of the 8th International Conference on Learning Representations.

[18] Y. Nakatani, T. Kajiwara and T. Ninomiya. (2022). Comparing BERT-based Reward Functions for Deep Reinforcement Learning in Machine Translation, in Proceedings of the 9th Workshop on Asian Translation (WAT 2022), pp.37-43.

[19] M. Cettolo, J. Niehues, S. Stüker, L. Bentivogli and M. Federico. (2014). Report on the 11th IWSLT evaluation campaign, in Proceedings of the 11th International Workshop on Spoken Language Translation: Evaluation Campaign, pp.2–17.

[20] D. Cer, M. Diab, E. Agirre, I. Lopez-Gazpio and L. Specia. (2017). SemEval-2017 Task 1: Semantic Textual Similarity Multilingual and Crosslingual Focused Evaluation, in Proceedings of the 11th International Workshop on Semantic Evaluation, pp. 1–14.

[21] N. Reimers and I. Gurevych. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks, in

Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, pp. 3982–3992.

[22] S. R. Bowman, G. Angeli, C. Potts and C. D. Manning. (2015). A large annotated corpus for learning natural language inference, in Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp.632–642.

[23] T. Gao, X. Yao and D. Chen. (2021). SimCSE: Simple Contrastive Learning of Sentence Embeddings, in Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 6894–6910.

[24] D. P. Kingma and J. Ba. (2015). Adam: A Method for Stochastic Optimization, in Proceedings of the 3rd Inter-national Conference on Learning Representations.

[25] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. L. Scao, S. Gugger, M. Drame, Q. Lhoest, A. Rush. (2020). Transformers: State-of-the-Art Natural Language Processing, in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, pp. 38–45.