

# 深層強化学習によるテキスト平易化における文難易度報酬の適用

中町礼文<sup>†</sup>, 梶原智之<sup>‡</sup>, 荒瀬由紀<sup>†</sup>

<sup>†</sup>大阪大学大学院情報科学研究科, <sup>‡</sup>大阪大学データリテリフロンティア機構

{nakamachi.akifumi, arase}@ist.osaka-u.ac.jp, kajiwara@ids.osaka-u.ac.jp

## 1 はじめに

テキスト平易化 [1] は、テキストの意味を保持しつつ、難解な文法や語彙を平易に変換する言い換え生成タスクの一種である。この技術は言語学習者の学習支援や読解支援に役立つだけでなく、機械翻訳などの自然言語処理応用タスクの前処理としても有用である。

テキスト平易化の理想的な出力は、以下を満たす。

- 同義性：入出力間で意味が等価である
- 文法性：出力文が文法的に正しい
- 平易性：入力文に対して出力文が平易である

Zhang and Lapata [2] は、これらのそれぞれに対応する報酬を用いた深層強化学習に基づくテキスト平易化モデルを提案した。しかし、平易性に関する報酬に対応する SARI [3] は文の難易度を直接的に評価する尺度ではないため、平易性についての人手評価とは弱い相関しか得られていない。

本研究では、上記の 3 項目に対応する報酬を BERT [4] を用いた回帰モデルとして設計し、図 1 に示すように深層強化学習の枠組みでテキスト平易化モデルを訓練する。Zhang and Lapata [2] の先行研究とは異なり、文の難易度を直接的に評価する報酬の導入によって、提案手法はより平易なテキストを生成することが期待できる。Newsela [5] を用いた実験の結果、我々の提案する報酬に基づく深層強化学習によって、テキスト平易化の性能が改善することを確認した。

## 2 関連研究

テキスト平易化は同一言語の翻訳問題として解くことができる。そのため、フレーズベースの統計的機械翻訳モデルを用いる PBMT-R [6] や注意機構に基づくニューラル機械翻訳モデルを用いる EncDecA [2, 7] などが提案されている。本研究では、単純な翻訳問題としてではなく、テキスト平易化タスクの特徴であるテキストの難易度を考慮してモデルを設計する。

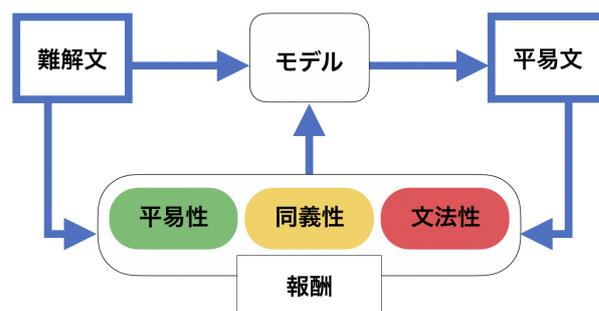


図 1: 深層強化学習に基づくテキスト平易化

テキストの難易度を考慮するテキスト平易化モデルとしては、クロスエントロピー損失を単語の難易度によって重み付けする手法 [8, 9] が提案されている。しかし、これらの手法では文全体の大域的な難易度を考慮することができない。深層強化学習の報酬として同義性・文法性・平易性を考慮する DRESS [2] は、本研究と最も関連する先行研究である。DRESS では、入力文と出力文をそれぞれ自己符号化器でベクトル化し、ベクトル間の余弦類似度によって同義性を評価する。また、文法性は出力文の言語モデル尤度、平易性は入力文・出力文・参照文の 3 つ組を用いて SARI [3] で評価する。しかし、SARI は入力文に対する出力文の語句の追加・保持・削除の F 値に基づく評価尺度であり、テキストの難易度を直接的に評価しない。本研究では、テキストの難易度を直接的に評価する報酬を用いて深層強化学習の枠組みでモデルを訓練する。

## 3 提案手法

本研究では、入出力間の同義性・出力文の文法性・出力文の平易性の 3 報酬に基づく深層強化学習によって、事前訓練済みのテキスト平易化モデルを再訓練する。

### 3.1 報酬推定器の構築

テキスト平易化モデルの強化学習に用いる同義性・文法性・平易性の報酬推定器を構築するために、それぞれの関連タスクのデータセットを用いて回帰モデルを訓練する。各報酬器の訓練に用いたデータセットの統計情報を表 1 に示す。

同義性に関する報酬器を訓練するために、文間の意味的類似度推定タスクである STS-B データセット<sup>1</sup> [10] を用いる。STS-B は、複数のアノテータが入力の 2 文の意味的類似度を [0, 5] の範囲で判定したデータセットである。本研究では、2 文を入力として受け取り、各アノテータが付与したスコアの平均値を推定する回帰モデルを訓練する。ただし、推定性能を高めるために、(文 1、文 2、類似度) の形式の各データから、(文 2、文 1、類似度)、(文 1、文 1、5.0)、(文 2、文 2、5.0) の疑似データを生成し、データ拡張を行う。

文法性に関する報酬器を訓練するために、文法誤り訂正タスクで利用される GUG データセット<sup>2</sup> [11] を用いる。GUG は、1 人の専門家が文の文法性を [1, 4] の範囲で判定したデータセットである。本研究では、1 文を入力として受け取り、文法性スコアを推定する回帰モデルを訓練する。

平易性に関する報酬器を訓練するために、テキスト平易化タスクで利用される Newsela データセット<sup>3</sup> [5] を用いる。Newsela は、ニュース記事に対して米国学校制度の学年ラベル (2 年生から 12 年生まで) が付与されたデータセットである。テキスト平易化タスクでは、これらのニュース記事を専門家が人手で平易に書き換えたものに対して自動的に文アライメントをとったパラレルコーパス [5] が利用されており、本研究でもこれを用いる。本研究では、1 文を入力として受け取り、その文が含まれる文書の学年ラベルを推定する回帰モデルを訓練する。

### 3.2 テキスト平易化モデルの事前訓練

テキスト平易化モデルの事前訓練として、注意機構に基づくニューラル機械翻訳モデル [12] のクロスエントロピー損失を最小化する。このモデルは、各単語を one-hot 表現で表した入力系列  $\mathbf{x} = \{x_1, \dots, x_n\}$  から

<sup>1</sup><http://ixa2.si.ehu.es/stswiki/index.php/STSBenchmark>

<sup>2</sup><https://github.com/EducationalTestingService/gug-data>

<sup>3</sup><https://newsela.com/data/>

	Train	Dev	Test	Label	Pearson r
同義性	5,749	1,500	1,379	[0, 5]	0.897
文法性	1,518	747	754	[1, 4]	0.655
平易性	94,208	1,129	1,077	[2, 12]	0.762

表 1: 報酬推定器の構築

中間表現を生成する符号化器と、符号化器の生成する中間表現および出力系列  $\mathbf{y} = \{y_1, \dots, y_m\}$  から次の出力単語分布を生成する復号器によって構成される。

符号化器  $f_{\text{enc}}$  は、入力系列  $\mathbf{x}$  から中間表現  $\mathbf{h} = \{h_1, \dots, h_n\}$  を式 (1) のように生成する。

$$h_j = f_{\text{enc}}(h_{j-1}, x_j) \quad (1)$$

復号器は、符号化器にて生成された中間表現および出力系列  $\mathbf{y} = \{y_1, \dots, y_m\}$  から出力予測の対数尤度を式 (2) のように計算する。

$$\log p(\mathbf{y}|\mathbf{x}) = \sum_{i=1}^m \log p_{\text{dec}}(y_i | y_{1:i-1}, \mathbf{h}) \quad (2)$$

事前訓練では、任意の長さの入力文  $\mathbf{x}$  および長さ  $M$  以下の出力文  $\mathbf{y}$  からなるデータ集合  $D = \{(\mathbf{x}^1, \mathbf{y}^1), \dots, (\mathbf{x}^K, \mathbf{y}^K)\}$  に対して、以下のクロスエントロピー損失を最小化する。

$$\mathcal{L}_C = - \sum_{k=1}^K \sum_{i=1}^M y_i^k \log p_{\text{dec}}(y_i^k | y_{1:i-1}^k, \mathbf{h}^k) \quad (3)$$

### 3.3 強化学習と報酬設計

強化学習に基づくテキスト平易化モデルの再訓練のために、REINFORCE [13] を用いる。REINFORCE では、クロスエントロピー損失の各時刻に対して推定される報酬  $R_t$  を重み付けした目的関数を最適化する。

$R_t$  は、 $\mathbf{x}, \mathbf{y}, \hat{\mathbf{y}}$  などの系列全体から計算される評価値  $\mathcal{S}(\cdot)$  に対して、ニューラルネットワーク  $b_t$  を用いて次のように定義される [14]。

$$R_t(\cdot) = \mathcal{S}(\cdot) - b_t \quad (4)$$

時刻  $t$  の報酬を適切に推定するため、強化学習の中で並行して  $b_t$  を目的関数  $\|b_t - \mathcal{S}(\cdot)\|^2$  で最小化する。

評価値は、 $\mathcal{S}(\cdot) = \alpha r_{\text{mean}} + \beta r_{\text{gram}} + (1 - \alpha - \beta) r_{\text{simp}}$  と設計する。

	自動評価尺度			報酬推定器による評価			
	BLEU ↑	SARI ↑	FKGL ↓	同義性 ↑	文法性 ↑	平易性 ↑	平均報酬 ↑
DRESS [2]	<b>23.21</b>	<b>27.37</b>	<b>4.13</b>	0.732	<b>0.952</b>	0.606	0.763
EncDecA	22.40	25.61	5.50	0.734	0.916	0.617	0.756
BERT-RL	23.02	25.31	4.86	<b>0.785</b>	0.881	<b>0.641</b>	<b>0.769</b>

表 2: テキスト平易化モデルの性能評価

- $r_{\text{mean}}(\mathbf{x}, \hat{\mathbf{y}})$ : STS-B の回帰モデルの出力報酬
- $r_{\text{gram}}(\hat{\mathbf{y}})$ : GUG の回帰モデルの出力報酬
- $r_{\text{simp}}(\hat{\mathbf{y}})$ : Newsela の回帰モデルの出力報酬

学習を安定させるために、クロスエントロピー損失との線形補完により、最終的な損失を計算する [15]。

$$\mathcal{L} = \lambda \mathcal{L}_C + (1 - \lambda) \mathcal{L}_R \quad (5)$$

## 4 評価実験

各タスクにおける報酬推定器の性能および Newsela におけるテキスト平易化モデルの性能を評価する。

### 4.1 報酬推定器の性能評価

本実験では、報酬推定のための回帰モデルとして事前訓練済みの BERT [4] を各タスクで再訓練する。BERT は大規模コーパスを用いて事前訓練された文の符号化器であり、対象タスクにおける再訓練によって多くのタスクで高い性能を達成している。

BERT には bert-base-uncased<sup>4</sup> を用いる。これは 12 の注意ヘッドに基づく 768 次元の 12 層モデルであり、各層には確率 0.1 でドロップアウトを適用している。単語分割には語彙数 30,522 の SentencePiece<sup>5</sup> を用いる。バッチサイズは 32、エポック数 3 で、学習率  $1e^{-5}$ 、epsilon  $1e^{-8}$  の Adam [16] で最適化する。

報酬推定器の出力は次のように  $[0, 1]$  に正規化する。

- $r_{\text{mean}}(\mathbf{x}, \hat{\mathbf{y}}) = \text{BRRT}_{\text{STS-B}}(\mathbf{x}, \hat{\mathbf{y}}) / (5 * 3)$
- $r_{\text{gram}}(\hat{\mathbf{y}}) = \text{BERT}_{\text{GUG}}(\hat{\mathbf{y}}) / (4 * 3)$
- $r_{\text{simp}}(\hat{\mathbf{y}}) = (13 - \text{BERT}_{\text{Newsela}}(\hat{\mathbf{y}})) / (12 * 3)$

各データセットのテストセットを用いて評価したピアソンの相関係数を表 1 に示す。いずれのタスクに

おいても、BERT に基づく報酬推定が人手評価との高い相関を持つことがわかる。ただし、小規模なデータセットを用いて訓練した文法性の報酬推定器は、人手評価との相関が比較的低い。

### 4.2 テキスト平易化モデルの性能評価

テキスト平易化モデルには、DRESS [2] の設定に合わせて 256 次元の 2 層 LSTM を使用し、隠れ層に確率 0.2 でドロップアウトを適用する。語彙は訓練データのソース側とターゲット側のそれぞれで出現頻度が 3 回以上の単語のみを用いる。なお、安定した報酬計算のために、モデルの最大出力長を 30 単語<sup>6</sup> に制限する。

バッチサイズは 32 で、事前訓練および強化学習による再訓練の両方で BLEU [17] による early stopping (エポック数 20) によって訓練を停止する。テキスト平易化モデルの事前訓練では、学習率 0.001, epsilon 0.999 の Adam [16] で最適化する。強化学習による再訓練では、各報酬の重みを  $\alpha = \beta = 0.33$ ,  $\lambda = 0.2$  とし、学習率 0.01 の SGD で最適化する。

本実験では、深層強化学習に基づくテキスト平易化モデルである DRESS [2]、我々が事前訓練したテキスト平易化モデルである EncDecA、強化学習による再訓練を行った BERT-RL の 3 モデルの性能を比較する。性能評価には、テキスト平易化タスクで一般的な BLEU [17] · SARI [3] · FKGL [18] とともに、我々が構築した報酬推定器を使用する。

実験結果を表 2 に示す。提案手法である BERT-RL が、ベースラインである EncDecA よりも BLEU および FKGL の評価において高い性能を達成した。表 2 の報酬推定器による評価では、各報酬推定器の出力を文ごとに  $[0, 1]$  に正規化した値の平均値を示している。提案手法は同義性および平易性の報酬値において最高性能を達成したが、文法性の報酬値はベースラインより

<sup>4</sup><https://github.com/huggingface/transformers>

<sup>5</sup><https://github.com/google/sentencepiece>

<sup>6</sup>訓練データにおける参照文の 95%以上の文を網羅できる。

入力文	They are tired and it shows in their voices , but they 're still on the freedom highway .
参照文	Their voices sound tired .
EncDecA	They are tired and it shows in their voices , but they 're still on the freedom .
BERT-RL	They are tired and it shows in their voices .

表 3: モデル出力の例

も低下した。これは先述の通り、小規模なデータセットで訓練した文法性の報酬推定器が比較的低い人手評価との相関を持つことに起因すると考えられる。

表 3 に、モデル出力の例を示す。EncDecA のベースラインよりも提案手法の BERT-RL が積極的に削除や書き換えを行い、難解な表現を平易化していた。

## 5 おわりに

本研究では、深層強化学習によるテキスト平易化において、BERT を利用した入力文と出力文の同義性、出力文の文法性、出力文の平易性の各報酬推定器を構築し、適用した。既存手法が SARI に対して最適化を行っているのに対して、提案手法では文の難易度を直接的に考慮した最適化を行うことで、テキスト平易化の性能を改善できることを示した。

## 謝辞

本研究は JST (ACT-I、課題番号: JPMJPR18UB) の支援を受けたものである。

## 参考文献

- [1] Matthew Shardlow. A Survey of Automated Text Simplification. *International Journal of Advanced Computer Science and Applications, Special Issue on Natural Language Processing 2014*, pp. 58–70, 2014.
- [2] Xingxing Zhang and Mirella Lapata. Sentence Simplification with Deep Reinforcement Learning. In *Proc. of EMNLP*, pp. 584–594, 2017.
- [3] Wei Xu, Courtney Napoles, Ellie Pavlick, Quanze Chen, and Chris Callison-Burch. Optimizing Statistical Machine Translation for Text Simplification. *TACL*, Vol. 4, pp. 401–415, 2016.
- [4] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proc. of NAACL*, pp. 4171–4186, 2019.
- [5] Wei Xu, Chris Callison-Burch, and Courtney Napoles. Problems in Current Text Simplification Research: New Data Can Help. *TACL*, pp. 283–297, 2015.
- [6] Sander Wubben, Antal van den Bosch, and Emiel Krahmer. Sentence Simplification by Monolingual Machine Translation. In *Proc. of ACL*, pp. 1015–1024, 2012.
- [7] Sergiu Nisioi, Sanja Štajner, Simone Paolo Ponzetto, and Liviu P. Dinu. Exploring Neural Text Simplification Models. In *Proc. of ACL*, pp. 85–91, 2017.
- [8] Reno Kriz, João Sedoc, Marianna Apidianaki, Carolina Zheng, Gaurav Kumar, Eleni Miltsakaki, and Chris Callison-Burch. Complexity-Weighted Loss and Diverse Reranking for Sentence Simplification. In *Proc. of NAACL*, pp. 3137–3147, 2019.
- [9] Daiki Nishihara, Tomoyuki Kajiwara, and Yuki Arase. Controllable Text Simplification with Lexical Constraint Loss”. In *Proc. of ACL*, pp. 260–266, 2019.
- [10] Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. GLUE: A Multi-Task Benchmark and Analysis Platform for Natural Language Understanding. In *Proc. of EMNLP*, pp. 353–355, 2018.
- [11] Michael Heilman, Aoife Cahill, Nitin Madnani, Melissa Lopez, Matthew Mulholland, and Joel Tetreault. Predicting Grammaticality on an Ordinal Scale. In *Proc. of ACL*, pp. 174–180, 2014.
- [12] Thang Luong, Hieu Pham, and Christopher D. Manning. Effective Approaches to Attention-based Neural Machine Translation. In *Proc. of EMNLP*, pp. 1412–1421, 2015.
- [13] Ronald J. Williams. Simple Statistical Gradient-following Algorithms for Connectionist Reinforcement Learning. *Machine Learning*, pp. 229–256, 1992.
- [14] Marc’Aurelio Ranzato, Sumit Chopra, Michael Auli, and Wojciech Zaremba. Sequence Level Training with Recurrent Neural Networks. In *Proc. of ICLR*, 2016.
- [15] Kazuma Hashimoto and Yoshimasa Tsuruoka. Accelerated Reinforcement Learning for Sentence Generation by Vocabulary Prediction. In *Proc. of NAACL*, pp. 3115–3125, 2019.
- [16] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *Proc. of ICLR*, 2014.
- [17] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. BLEU: a Method for Automatic Evaluation of Machine Translation. In *Proc. of ACL*, pp. 311–318, 2002.
- [18] J. Peter Kincaid, Robert P. Fishburne Jr., Richard L. Rogers, and Brad S. Chissom. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. *Technical report, Defence Technical Information Center Document*, pp. 8–75, 1975.