

著者識別に基づく文体スタイル変換への取り組み

理学専攻・情報科学コース

2240679

施 佳坤

1 はじめに

テキストは最も一般的なデータの一つであり、人間の言語表現の一形態として、表現したい内容だけでなく、独自の属性特性も含んでいる。本研究はテキスト変換に注目する。Text Style Transfer (TST) の目的は、テキストの内容を保持しつつ、そのスタイルを変更することである [1]。異なる著者の執筆スタイルは異なる。例えば、太宰治のスタイルは内向的で悲壮感があり、一方で東野圭吾は緊密でサスペンスに富んでいる。特定の作家が書いた雰囲気や文章を読みたい読者の好みを満たすため、本稿で研究するテキストスタイル変換は、通常の日本語を特定の作家のスタイルに自動的に変換し、ふつうの平文を特定の作家の文体スタイルに変換する。

テキストスタイル変換は教師あり学習の一種であり、原文テキストと目標テキストのデータが必要である。従来のテキストスタイル変換は、データセットのスタイルが限られていたため、特定のデータセット上でのみスタイル変換が可能であった。しかし近年、ChatGPT [2] のような大規模言語モデルが登場し、任意のテキストスタイル変換データセットを大規模言語モデルにより生成し、異なる作家のスタイルのテキストを得るためのテキストスタイル変換の学習が可能になった。

2 関連研究

本節では、従来のテキストスタイル変換方法と大規模言語モデルの発展の二つの観点から関連研究を紹介する。

2.1 従来のテキストスタイル変換方法

テキストスタイル変換は、教師あり学習の一種であり、多くの研究が Sequence-to-Sequence (Seq2Seq) モデル [7] に基づいて展開されている。このモデルはエンコーダとデコーダを含む。エンコーダには原文テキストが入力され、デコーダではスタイル変換後のテキストが生成される。Seq2Seq モデルは、大量の平行データに依存しており、最も基本的なモデルからの学習が必要であった。しかし、自然言語処理と大規模言語モデルの発展に伴い、テキストスタイル変換は既存大規模言語モデルのファインチューニングへと進化し、データ依存度を大幅に低減してきた。

2.2 大規模言語モデルの発展

近年、BERT [4]、T5 [5]、GPT [6] などの言語モデルが急速に発展している。BERT (Bidirectional Encoder Representations from Transformers) は双方向の Transformer エンコーダモデルで、主に言語表現を事前学習するために使用されています。「Masked Language Model」(MLM) の事前学習方式を採用している。つまり、入力テキストの中でランダムにいくつかの単語を選択して覆い、モデルを通じてこれらの覆い隠された単語を予測する。Bert は自然言語処理に広く応用され、問答システム、感情分析、テキスト分類な

どのタスクに使用できる。

T5 (Text-to-Text Transfer Transformer) はテキストからテキストへの変換モデルであり、入出力形式を統一することで、複数の NLP タスクを同じ方法で処理することができる。T5 では、事前学習方式を採用し、大量の NLP タスクを類似のテキスト変換問題に変換して学習した。T5 は汎用性が高く、機械翻訳、要約生成、質問回答などの各種 NLP タスクに応用できる。

GPT (Generative Pre-trained Transformer) は、主に大規模な事前学習とテキスト生成のための Transformer に基づく生成モデルである。GPT は、モデルが一貫性のあるテキストを生成できるように、教師無しプリトレーニングにより大規模なテキスト語彙上で予測する。GPT は主にテキスト生成に用いられ、一貫したテキスト段落、文章、または会話を生成することができる。

3 提案手法

図 1 に示すように、本研究では ChatGPT を用いて著者のオリジナルテキストスタイルを一般的なテキストに変換したデータセットを生成し、事前学習された T5 モデルのファインチューニングを行い、著者のテキストスタイル変換モデルを得る。また、本研究では BERT を利用したスタイル判別モデルを構築し、テキストスタイル変換の結果を評価する。

3.1 データ生成

異なる著者のテキストスタイル変換データセットを取得するために、ChatGPT を使用してデータを生成する。具体的な方法としては、自然言語処理の分割方法を用いて著者の作品を一文ごとにに分ける。これらの文を ChatGPT に入力し、標準的な日本語スタイルのテキストを出力させる。

本研究では、図 1(A) に示されるように、二つの Prompt 方式を使用し、prompt の違いによる標準的な日本語変換の差異を確認する。Zero-Shot の Prompt 方式では、入力された作家のテキストを直接 ChatGPT が標準的な日本語に変換する。One-Shot の Prompt 方式では、最初に一つの変換の例を与え、ChatGPT に日本語への変換方法を示した後、作家のテキストを入力して標準的な日本語を出力させる。

3.2 モデルの学習

図 1(B) に示されるように、BERT を用いた分類器を学習し、著者のスタイルかどうかを識別する分類器を使用して、文の特徴ベクトルからスタイルのカテゴリを得る。特定の著者のスタイルに属する文には 1 のラベルが付けられ、それ以外は 0 のラベルが付けられる。クロスエントロピー損失を用いて分類損失を計算する。

次に、T5 を用いたスタイル変換モデルを学習する。T5 モデルは、ChatGPT によって生成された標準的な日本語を入力として受け取り、目標の著者スタイルの

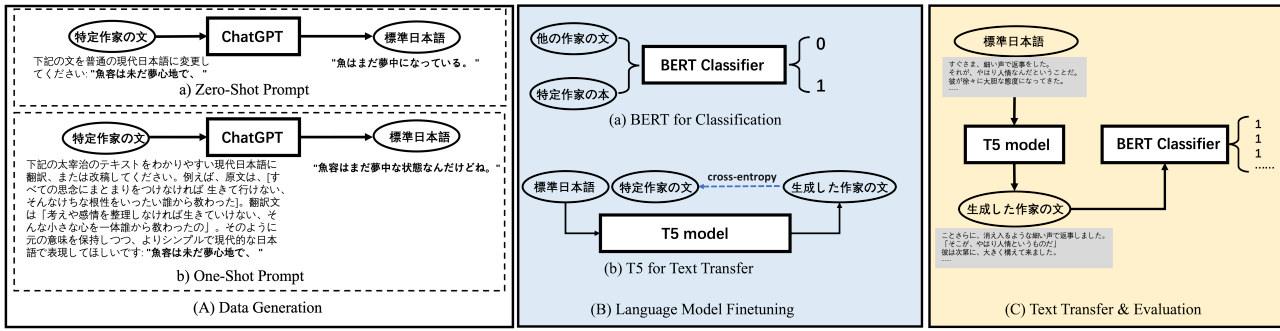


図 1: 提案手法の全体図

表 1: Bert 分類器の精度

モデル	精度
Bert_Three_Author	94.1%

表 2: T5 スタイル変換モデル精度

	モデル	精度
Zero-Shot	T5-Base	69.21%
	T5-ECC	73.55%
One-Shot	T5-Base	73.31%
	T5-ECC	78.18%

テキストを出力する。生成されたテキストと特定の著者のテキストとの類似度を判断するためにクロスエントロピー損失関数を使用する。繰り返しの学習を通じて、T5 モデルは目標の著者のスタイルのテキストを生成する方法を学習する。

最後に、T5 モデルによって生成されたスタイルのテキストを構築した著者分類器に入力して検証し、分類器がそのテキストを特定の作家のスタイルに属するものと判断すれば、スタイル変換が成功したとみなされる。

4 実験

まずは、BERT で三値分類器を学習する。具体的には、青空文庫から太宰治、宮沢賢治、芥川龍之介三人の作家のテキストデータをそれぞれ37878, 37367, 30849 収集し、それぞれ 30000 件のテキストを学習データセットとして使用し、残りをテストセットとした。50 エポックの学習を行い、テストセットの結果は表 1 に示され、94%の精度で T5 モデルのスタイル変換結果を評価した。

太宰治のテキストデータから 30 文字以下の 6059 件を選択し、ランダムに 5000 件を訓練セットとして使用し、残りの 1059 件をテストセットとした。次に、Zero-Shot および One-Shot の方法で ChatGPT を使用して標準的な日本語を生成した。HuggingFace プラットフォーム上の事前学習された T5 モデル T5-base-japanese および T5-finetuned-ECC をベースに、それぞれファインチューニングを行った。各実験は 50 エポックの学習を行った、実験結果を表 2 に示す。」

実験の結果から、ChatGPT を利用してテキストスタイル変換データセットを構築する方法は実行可能であり、任意の作家のスタイルを持つテキスト生成器を学習することができることがわかった。また、OneShot

方式で得られたテキストデータは、より良いテキストスタイル変換効果を得ることができた。

5 おわりに

本稿では、ChatGPT を使用してテキストスタイル変換データセットを生成し、著者のテキストスタイル変換の実験を行った。作家のテキストスタイルを標準的な日本語テキストへスタイル変換する際に、Zero-Shot および One-Shot の Prompt を用いたが、One-Shot の Prompt を用いて生成したデータセットを用いてスタイル変換を行ったほうが高い精度となった。実験結果は、提案手法の有効性を示している。今後は、提案手法の精度向上を目指し、判別器、スタイル変換器、学習データに対する改善に取り組むつもりである。

参考文献

- [1] Di Jin, Zhijing Jin, Zhiting Hu, Olga Vechtomova, Rada Mihalcea; Deep Learning for Text Style Transfer: A Survey. Computational Linguistics 2022; 48 (1): 155-205. doi: https://doi.org/10.1162/coli_a.00426
- [2] OpenAI, Chatgpt. accessed: 2023-02-08., url:<https://openai.com/blog/chatgpt/> (cit. on pp. 1, 16)
- [3] Xu R, Ge T, Wei F. Formality style transfer with hybrid textual annotations[J]. arXiv preprint arXiv:1903.06353, 2019.
- [4] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.
- [5] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer[J]. The Journal of Machine Learning Research, 2020, 21(1): 5485-5551.
- [6] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training[J]. 2018.
- [7] Sutskever, Ilya, Oriol Vinyals, and Quoc V. Le. "Sequence to sequence learning with neural networks." Advances in neural information processing systems 27 (2014).