Bridging Literary Eras: A Back-Translation Approach to Making Shakespearean Text More Accessible in Contemporary English

Mohammad Amin Ghasemi
Department of Mathematics and Computer Science
Amirkabir University of Technology
Tehran, Iran
aminghassemi7@gmail.com

Mehdi Ghatee*
Department of Mathematics and Computer Science
Amirkabir University of Technology
Tehran, Iran
ghatee@aut.ac.ir

***Abstract*—Language style transfer is the task of altering the stylistic characteristics of a text while preserving its original context. This paper focuses on style transfer tasks aimed at reducing author-specific characteristics and enhancing readability for contemporary readers. Our study centers on the transformation of Shakespearean texts for this purpose. Prior research has demonstrated that back-translation, a two-step process involving translation to an intermediate language and translating back to the original language, can be effective in achieving this objective. Leveraging translation engines, particularly their propensity to generate text resembling everyday spoken English due to their training data, we employ this approach to create a dataset for training a Sequence-to-Sequence (Seq2Seq) model enhanced with attention mechanism. Our results confirm the effectiveness of directly applying the back-translation. Furthermore, in some cases where back-translation yielded suboptimal results, the existence of similar training examples where it performed well aided our model in generating improved text compared to the target text.**

Keywords—Style transfer, Back-Translation, Seq2Seq models, Attention mechanism

I. INTRODUCTION

Many masterpieces of literature, including Victorian novels, poems, and the works of William Shakespeare, possess distinctive characteristics that often present challenges for modern-day readers seeking to comprehend them. These stylistic nuances, while contributing to their timeless appeal, can create linguistic barrier, making these texts less accessible to contemporary audiences.

In recent years, the convergence of natural language processing (NLP) and machine translation has opened doors to innovative solutions in the realm of style transfer. Style transfer involves the task of transforming text from one stylistic form to another while preserving its core semantic essence [1]. This paper introduces a pioneering solution that harnesses the capabilities of neural machine translation, specifically employing the concept of back-translation, to bridge the linguistic divide within literary works, with particular focus on the works of William Shakespeare, and bring them closer to the grasp of contemporary English readers.

SEQUENCE-TO-SEQUENCE MODEL

The Sequence-to-Sequence (Seq2Seq) model, has proven to be a pivotal architecture for a wide range of sequence generation tasks. Seq2Seq models are based on recurrent neural networks (RNNs), with Long-Short Term Memory (LSTM) cells being a popular choice due to their ability to capture long-range dependencies in sequences [2]. This model consists of two main components: an encoder and a decoder.

- **Encoder:** The encoder processes the input sequence, typically tokenized text, and compresses it into a fixed-length context vector, also known as the encoder’s hidden state. Bidirectional LSTMs, a variant of LSTMs, are often used to capture information from both the past and the future context of each word, enhancing the model’s ability to understand the input sequence.
- **Decoder:** The decoder takes the context vector produced by the encoder and generates an output sequence, word by word, typically autoregressively. LSTMs are commonly employed here as well, allowing the decoder to consider the generated words in context.

ATTENTION MECHANISMS

Attention mechanisms have revolutionized the field of natural language processing by addressing one of the fundamental limitations of Seq2Seq models: their inability to selectively focus on specific parts of the input sequence when generating the output. Attention mechanisms provide the decoder with the ability to attend to different parts of the input sequence dynamically, improving the model's translation and generation capabilities [3].

Back-Translation

Back-translation, originally developed to augment parallel corpora for machine translation [4], has found a compelling application in style transfer tasks. The technique involves a two-step process: first, translating text from the source style (Shakespearean English) to an intermediate language (e.g., German), and then translating it back to the target style (contemporary English). The magic unfolds in the return journey, as the content is reintroduced into contemporary English, examples of this are provided in Table 1. Studies have shown that author characteristics are significantly vanished by either manual and automatic machine translation [5]. This has been used in prior research to derive latent representations of sentence semantics, with are then used to generate text via separate style-specific generators [6].

- **Poem**
  Original: Stars above twinkle, as dreams we conceive
  Back-Translated: Stars twinkle above, like dreams we imagine.

- **Romance**
  Original: The world faded away, leaving only the two of them
  Back-Translated: The world disappeared and only the two of them remained

---

- **Victorian literature**
  Original: In the parlour, they discussed matters of society and literature, sipping tea from cups.
  Back-Translated: In the salon they discussed social and literal topics and drank tea from cups.

- **Shakespearean**
  Original: You are all resolved to die than to famish?
  Back-Translated: Are you all determined to die rather than starve?

Table 1: Examples of applying back-translation on different styles of literature.

We take advantage of and the fact that large scale machine translation models are trained mostly on text of every-day English to train a Seq2Seq model which its input sequences are Shakespearean text and its target or 'translated' sequences are back-translated pair using a translation engine.

II. METHODOLOGY

We begin with a dataset X comprising of stylized sentences, specifically in the Shakespearean style. Our objective is to create a corresponding dataset Y, consisting of sentences from dataset X that have undergone a two-step transformation. First, they are translated into another language, and then they are translated back into English. These transformed sentences in dataset Y, along with the original sentences in dataset X, serve as the input and target sequences for our Seq2Seq model.

We employ a Bi-Directional LSTM as the encoder in our Seq2Seq model. The encoder utilizes both forward and backward hidden states, concatenating them for each time step.

To enhance the decoding process, we incorporate an attention mechanism. During each step of decoding, the attention mechanism computes the 'relevance' of each encoder state, denoted as s_k, with respect to the current decoder state h_t. This relevance score, denoted as $score(h_t, s_k)$, is computed through an attention function. Subsequently, a probability distribution (softmax) is applied to these attention scores. The attention output or context vector, denoted as $c^{(t)}$, is computed as the weighted sum of encoder states, using the attention weights. The general computation scheme is as follows:

$$a_k^{(t)} = \frac{\exp(score(h_t, s_k))}{\sum_{i=1}^{m} \exp(score(h_t, s_i))} \qquad (1)$$

$$c^{(t)} = \sum_{k=1}^{m} a_k^{(t)} s_k \qquad (2)$$

We have incorporated Luong attention into our model. While the original paper [7], explores several model variants, the one commonly employed, and the one we have used as well, utilizes a Bilinear function to compute attention scores. This attention mechanism is applied after each decoder step at time t before making predictions. Specifically, hidden state is combined with the context vector to produce an updated hidden state, h', which is subsequently employed for prediction. The calculation is outlined below:

$$h' = \tanh(W_c\, concat(h_t, c^{(t)})) \qquad (3)$$

Here, W_c is a learned matrix.

III. Implementation Details

- **Choice of intermediate Language**: We selected German as the intermediate language for the Back-Translation process due to its superior performance when compared to languages such as French and Spanish. This choice aligns with the fact that English belongs to the Germanic language family, and it is supported by the availability of a substantial amount of German text data on the internet.
- **Translation Engine**: For the translation process, we leveraged the Google Translate API, renowned for its exceptional translation capabilities, ensuring the quality and accuracy of our transformations.
- **Model Configuration**: Our Seq2Seq model was configured with a hidden size of 512. Additionally, we incorporated recurrent dropout with a rate of 0.2 for both the encoder and decoder components. These settings were crucial for optimizing the model's performance and handling sequence generation effectively.
- **Pre-trained Word Vectors:** We used GloVe embedding vectors with dimension 100 and consider these vectors as trainable parameters during training [8].
- **Teacher-Forcing Strategy**: During the initial four epochs of training, we employed the Teacher-Forcing technique to stabilize the learning process. The Teacher-Forcing rate was initially set to one, gradually decreasing in a linear fashion as the training progressed. This strategy facilitated smoother convergence and improved training dynamics.
- **Source Dataset**: We used the TinyShakespeare dataset as the primary source of our training data.

IV. Results and Discussion

The objective of our model is essentially, to detect specific characteristics embedded within the input data and translate them into common English, which seems to align naturally with the capabilities of a Seq2Seq model. To evaluate the performance of our approach, we employ the BLEU score, a metric commonly used to measure the quality of machine translation [9], between output and target. The result, as presented in Figure 1, is juxtaposed against a threshold score typically regarded as indicative of successful performance in various tasks, reaffirming the effectiveness of the approach.

Input:	I should be arguing still upon that doubt
Output:	I should still argue about this doubt
Input:	Thou art a villain
Output:	You are a villain
Input:	It carries a brave form. But <UNK> a spirit
Output:	It bears a bold form. But it is a ghost.

Table 2: Examples of model outputs.

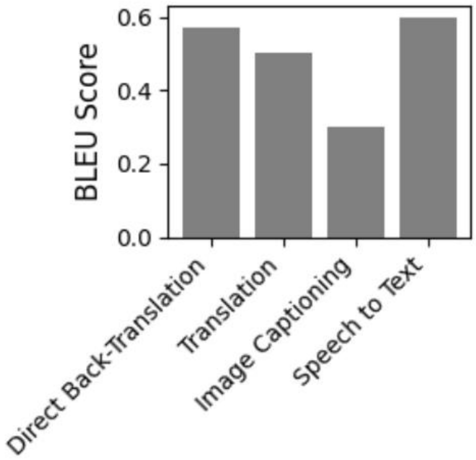

Fig 1: Comparing the obtained BLEU score of learning to directly back-translate, with a BLUE score that is considered 'good' for other tasks. (It's important to note that there is no officially recognized 'good' score for these tasks, and the numbers are presented here for the sake of argument, drawing from some state-of-the-art models)

We also observed that overfitting could occur rather easily. While training, our model exhibited the capability to generate results closely resembling the target sentences and achieve a BLEU score exceeding 0.8 for the training data. However, this high level of performance did not translate to effective generalization and resulted in less relevant test outputs.

To prevent overfitting, we found it prudent to monitor BLEU score during training and halt the training process once a specific threshold, such as 0.65, is reached. This practice helped maintain the model's generalization ability and improved its performance on unseen data.

Plotting attention maps essentially illustrates the core objective of our task: to identify and replicate elements that are already similar to contemporary language usage, while discerning the necessary transformations for those parts that require modification by aligning them with corresponding segments of the input sequence. An illustrative example is provided in Figure 2.

In our observation, we noted that in certain instances where faults were presented in the target sequences, the model demonstrated an ability to mitigate the errors with the assistance of similar examples provided during training. Examples of this phenomenon are presented in Table 2.

Furthermore, when evaluating our model against a set of Christopher Marlowe texts, an author renowned for a writing style akin to Shakespeare's, we observed promising outcomes when the test cases closely aligned with the vocabulary used during model training, as demonstrated in Table 3. However, it is noteworthy that the results exhibited a lack of coherence when sentences contained vocabulary that deviated from the training dataset. This underscores the importance of curating a comprehensive dataset encompassing a broad spectrum of vocabulary to foster the development of a more generalized model for a specific literary style.

Input: March on, since we are up in arms
Target: Keep marching, cause <UNK> in turmoil
Output: Keep marching, cause we up in arms
Input: Hath now himself met with fall of leaf
Target: Hath himself now witnessed the fall of the leaf
Output: Has now seen the fall of the leaf
Input: And lasting fealty to the new made king
Target: And enduring allegiance to the new king
Output: And lasting loyalty to the new king

Table 3: Instances of model generating a better output for a sentence than its corresponding target sentence.

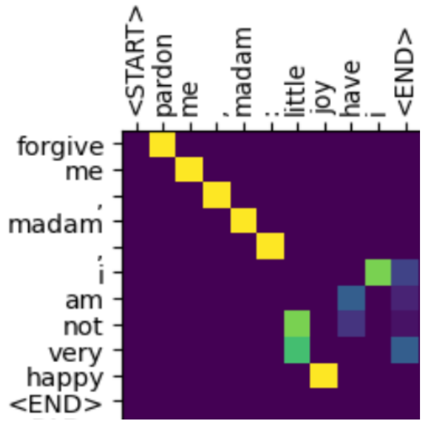

Fig 2: Example of a sentence's attention map, showcasing that the model is taking advantage of related parts of input during generation.

V. Conclusion And Future Work

In this work, we have harnessed the concept of back-translation and the capabilities of powerful translation engines to curate a dataset for training a Seq2Seq model that directly executes the two-step process of back-translation. As previously discussed, this innovative approach extends beyond Shakespearean text and can be applied to various forms of stylized literature, including poetry, Victorian novels, and even legal documents, which frequently employ esoteric vocabulary that can be substituted with everyday language equivalents.

Furthermore, we have observed that addressing partial inaccuracies introduced by translation engine becomes feasible when the model is provided with well-translated reference examples.

To create a model that can serve large-scale applications, it is imperative to accumulate a substantial volume of data encompassing a diverse range of vocabulary. Consequently, data collection emerges as a pivotal factor capable of significantly enhancing the viability of this approach. Additionally, we posit that a model could be trained to execute this transformation task across different yet stylistically similar texts. For instance, applying our approach to both Victorian and Shakespearean texts represents an intriguing avenue for further development.

References

[1] Martina Toshevska, Sonja Gievska, "A Review of Text Style Transfer using Deep Learning" arXiv:2109.15144v1 [cs.CL] 30 Sep 2021.

[2] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, "Sequence to Sequence Learning with Neural Networks" arXiv:1409.3215v3 [cs.CL] 14 Dec 2014.

[3] Dichao Hu, "An Introductory Survey on Attention Mechanisms in NLP Problems" arXiv:1811.05544v1 [cs.CL] 12 Nov 2018.

[4] Sergey Edunov, Myle Ott, Michael Auli, David Grangier, "Understanding Back-Translation at Scale" Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 489–500 Brussels, Belgium, October 31 - November 4, 2018.

[5] Ella Rabinovich, Shachar Mirkin, Raj Nath Patel, Lucia Specia, Shuly Wintner, "Personalized Machine Translation: Preserving Original Author Traits" arXiv:1610.05461v2 [cs.CL] 12 Jan 2017.

[6] Shrimai Prabhumoye, Yulia Tsvetkov, Ruslan Salakhutdinov, Alan W Black, "Style Transfer Through Back-Translation" arXiv:1804.09000v3 [cs.CL] 24 May 2018.

[7] Minh-Thang Luong, Hieu Pham, Christopher D. Manning, " Effective Approaches to Attention-based Neural Machine Translation" arXiv: 1508.04025v5 [cs.CL] 20 Sep 2015.

[8] Jeffrey Pennington, Richard Socher, and Christopher Manning. "GloVe: Global Vectors for Word Representation" In EMNLP, 2014.

[9] Kishore Papineni, Salim Roukos, Todd Ward, Wei-Jing Zhu, "BLEU: a Method for Automatic Evaluation of Machine Translation" Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp. 311-318.

The related code could be found here
Readers are welcome to use this code in their future works