Neural Chinese Poetry to English Translation

Yitao Zhou
College of Art & Science, New York University
yitao.zhou@nyu.edu

Abstract

In this work, I present a novel attempt to use state-of-the-art Neural Machine Translation (NMT) Transformer model to accomplish the task of Chinese poetry to English translation. My system could translate a source Chinese poem into readable corresponding semantic English meanings.

1 Introduction

1.1 Problem Description

Robert Frost once said, “Poetry is what gets lost in translation.” Despite recent developments in machine translation of regular texts, poetry translation remains a challenging problem, especially in Chinese poetry. Chinese Poetry is a culture treasure that act as records of people’s living experiences. Chinese poems are typically written with a fixed number of characters and sentences, together with specialized rules and rhythms applied to them. A famous collection of poem, named as Three Hundred Tang Poems, along with some great poems have been manually translated to English for the sake of non-Chinese speakers’ understanding for these poems as well as authors’ personal experience, yet most of the poems remain untranslated.

The difficulty of poem’s translation lies in its intrinsic complexity, however, attempts are always made since it will promote one culture’s understanding of another, especially in historical contexts. In this paper, I propose to build up a neural machine translation system to translate Chinese poems to English. Specifically, I will first implement a translation system based on Transformer NMT model (Vaswani et al., 2017). Then feed the system with Mandarin-English translation dataset to verify the feasibility of my system and adjust the hyperparameters to make it suitable for Chinese-English translation. Next, I will clean the poetry-English dataset to get each stich paired with its translation; and finally feed the system with poetry-English dataset to test the result of machine translation of poems.

1.2 Motivation

As a Mandarin speaker, during my personal learning journey, I observed that Chinese poetry is one of the most indigestible part of Chinese language. Main reasons are as follows:

1. Many characters used in ancient poetry are uncommon in Mandarin texts. Characters that are still in use nowadays may have a different meaning in poetry, and one character have diverse meanings.

2. The linguistic habits of ancient poems’ characters are different from those in Mandarin. Inverted stiches, elliptical stiches are commonly

used due to the limitation of 5 or 7 characters per stich for Chinese poems.

3. The background of ancient poetry is not clear. Rarely do preface, notes, and references exist, making it difficult for readers to grasp the poet's intention.

To understand a poem in Mandarin, the common process is finding idiomatic explanations of each character in the dictionary with similar connotations; and simultaneously taking care of ordering of words since inverted stiches happens from time to time; as well as taking into account author's personal experience at that time. Having similar processes as translating a foreign language, understanding Chinese poems is a time consuming procedure even for Mandarin speaker. Therefore, to save the time of understanding Chinese poems, I will attempt to build a machine translation model.

2 Related Works

There has been very little work related to the translation of poetry between languages. One recent work is conducted by Ghazvininejad et al. (2018) regard to neural French poem to English poem translating. They proposed to use sequence-to-sequence model with two layers of LSTM, to translate French poems with enforcing rhythm and rhyme constraints. Human evaluation ranks translation quality as acceptable 78.2% of the time. Prior to the popularity of NMT, Genzel et al. (2010) did a phrase-based MT to translate Italian and French poems to English ones with constaints on rhythm and rhyme with a report of only 12 out of 109 6-line French stanzas got generated according to specified scheme.

Despite few achievements are made in the field of poetry translation, works in applying MT techniques to Chinese poetry generation are flourishing and achieve remarkable results. Wang et al. (2016) and Yi et al. (2018) proposed models to keep the generated poems coherent and semantically consistent with the user's intent, while researches related to polishing up generation schema are continuously carried on (Yan, 2016; Deng et al., 2019).

3 Data

3.1 Mandarin-English translation dataset

I applied a dataset containing 21116 pairs of Mandarin-English translation, downloaded from `http://www.manythings.org/anki/` and considering it as reference translation. This dataset mainly contains semantic contents. An adjustment I made to this dataset is converting all Chinese inputs into simplified Chinese to reduce the number of entries in encoder vocabulary. After that, the Chinese vocabulary includes 2700 unique characters and English vocabulary includes 6156 words (including plural, tense etc.).

3.2 poetry-English translation dataset

I obtained the translation masterpiece of great Chinese translator Yuanchong Xu, who manually translated Three Hundred Tang Dynasty Poems into English and published in a book. I converted this book into machine readable files, cleaned the data, and paired each stich with translation.

After cleaning, this poetry-English dataset contains 3130 stiches with 2465 unique Chinese characters and 4173 English words (including plural, tense etc.) from 318 poems.

Dataset Information		
Dataset language	Mandarin-English	poetry-English
Pairs	21116	3130
Chinese characters	2700	2465
English characters	6156	4173

4 Methodology

4.1 Neural Machine Translation

The architecture of my poetry translation system is an encoder-decoder based self-attention Transformer model. The advantages of Neural Machine Translation model include learning long-range dependencies and lower morphology, lexical errors according to a comparison case study (Bentivogli et al., 2016). As a branch of Statistical Machine Translation, neural machine translation, especially Transformer, achieved state-of-the-art results in translation tasks with producing a nice result similar to phrase-based translation systems.

The reason for not choosing to use phrase-based systems is that my idea is hard to be implemented because Chinese poems are normally understood character by character rather than phrases. My purpose is that the machine system can follow the similar procedure as human's learning procedure, and character by character encoding is exactly one of the features of most NMT models.

4.2 Attention

Self-attention mechanism is the main technique I adopted to my system. It actually abandons using RNNs or CNNs, breaking one bottleneck of Neural Machine Translation which is fixed-length vector (Bahdanau et al., 2014). Prior to the propose of this mechanism, the encoder-decoder based neural machine translation model needs to encode the source sentence into a vector, normally as long as how many pieces the sentence is split into. However, in self-attention mechanism, each time when predicting a target word, this mechanism will search where the most relevant information is concentrated in source sentence, allowing the model to predict the target word based on the whole context vector generated from source rather than specific hidden states. In a word, the self-attention weights the relevancy of encoder hidden states to provide additional information during prediction.

4.3 Transformer Model

Among various neural translation models, the transformer model based on self-attention mechanism shows to produce promising results in both translation performance and speed, because self-attention can capture the relationship between two tokens regardless of their distance, and it does not apply recurrent units for modeling which makes parallel computing possible to improve speed (Vaswani et al., 2017).

The Transformer model I implemented basically contains these components:

1. Word segmentation. The first step is separating the sentences into isolated word and punctuation with every English word in a lower case and Chinese sentence in characters. Alternative methods in decoding English inputs include byte pair encoding (Sennrich et al., 2015), meaning separating English word also in character based form since it could possibly solve the problem of rare words, tenses and prefix, suffixes in English.

2. Input setup. Generating dictionary, mapping tokens in both encoder and decoder part to indices in dictionary in order to generate tensors with respect to every sentence.

3. Encoder. On the encoder layer, I first embed the given tensors. Then I use 6 identical stacked layers, each of them composed of a self-attention sub-layer and a feed-forward sub-layer, which will calculate different linear transformations of the input. Meanwhile, to ease training, layer normalization and residual connection are used to avoid vanishing gradient problems. The self-attention model used in my Transformer model is Multi-Head Attention, which calculates multiple attention weighted sums, and allows the model to jointly attend to information from different representation subspaces at different positions in the sentence.

4. Decoder. Likewise, on the decoder layer, my model has another stack of 6 identical layers with an additional encoder-decoder attention sub-layer to take into account the output of encoder layer. The decoder generates one word at a time from left to right. The first word is generated based on the final representation of encoder and every word subsequent to the first word is based on all previous words together with final representation of encoder.

5. Training and testing. The final step is to train the system with training set and use the parameters got from training the model to produce translation results according to Chinese input sentences.

5 Experiments and Evaluations

5.1 Mandarin-English translation

The reason for conducting Mandarin-English translation first is that characters in mandarin are the same as those in poetry. Therefore, I can use this experiment to adjust hyperparameters to make the system suitable for translation between two languages afterwards.

I split this dataset from `manythings.org` into 80% training set, 16% validation set and 4% test set. In practice, 16892 sentences are used to train, 3378 are used to validate and 844 are used to test. With excluding OOV in input sequences during testing, the output includes 825 sentences of machine translation result.

I chose to evaluate this task with BLEU-4 score because the dataset mainly consists of semantic contents. Meanwhile, I applied BLEU score with smoothing method4: having inflated precision values for shorter translation (Chen & Cherry, 2014). The reason for adopting this smooth function is that the reference output do contain some sentence with length shorter than 4. After smoothing, the BLEU score is 26.63. As a comparison, Tencent Transformer baseline system for WMT18 achieved BLEU score of 24.3 in Chinese-English task, despite running on different datasets. Figure 1 consists three consecutive results I got from testing. Although there exists deficient related to word choices, the machine translation system shows to produce understandable results in this Mandarin-English translation task with taking care of word alignment.

5.2 poetry-English translation

With adjusting hyperparameters to produce an acceptable result on Mandarin-English translation task, I continued applying the system to train and translate Chinese poetry to English. I applied the same splitting strategy as used in the Mandarin-English translation. 2500 of the stiches worked as training set, 500 as validation set while 130 as test set. A sample machine produced output is in figure 2 and these four stiches are a quatrain in Three Hundred Tang Poems. The machine translation results do match Chinese token meanings and coherent to read in English. However, machine translation results are

	他	去	意大利	的	目的	是	学习	音乐
TOKEN TRANSLATION:	he	go(to)	Italy	`s	purpose	is	study	music
CANDIDATE(MACHINE):	he went to italy in order to study music							
REFERENCE:	He went to Italy for the purpose of studying music.							

	他	和	他	所有的	同学	相处	融洽
TOKEN TRANSLATION:	he	and	his	all	classmates	get along with	well
CANDIDATE(MACHINE):	he is on goods terms with all of his classmates						
REFERENCE:	He is getting along well with all of his classmates.						

	我	发觉	说	英语	很	简单
TOKEN TRANSLATION:	I	found	speaking	english	very	easy
CANDIDATE(MACHINE):	i found it easy to speak english fluently					
REFERENCE:	I discovered that speaking English was pretty easy.					

Figure 1: Examples of Mandarin-English Translation Results

	向	晚	意	不	适
TOKEN TRANSLATION:	toward	evening	mood	not	happy
CANDIDATE(MACHINE):	to evening, twilight grows dim				
REFERENCE:	With twilight shadows in my heart				

	驱	车	登	古	原
TOKEN TRANSLATION:	drive	carriage	mount	old	mountain
				(refering to an mountain in central-China)	
CANDIDATE(MACHINE):	they enforce a car to ascend old mountains				
REFERENCE:	I have driven up among the Leyou Tombs				

	夕阳	无限	好
TOKEN TRANSLATION:	sunset	infinitely	good
CANDIDATE(MACHINE):	sunset clings a good sunset		
REFERENCE:	To see the sun, for all his glory,		

	只	是	近	黄昏
TOKEN TRANSLATION:	only	is	close(to)	twilight(dusk)
CANDIDATE(MACHINE):	only i am always in one yellow twilight			
REFERENCE:	Buried by the coming night.			

Figure 2: Examples of poetry-English Translation Results

hard to compete or evaluate with manual translation because of their divergence in focuses. Machine translation produces a coherent semantic meaning of translation tasks, yet taking into account imagines and author's moods is almost limited as a manual task.

Unlike Mandarin-English translation, the poetry translation can be significantly different from the referred translation but offering a high quality translation of poem's semantic meanings. A good method for evaluating poetry machine translation result could be based on fluency, coherence, fitness of semantic meanings, and poeticness. To evaluate, one way is to invite some professional bilingual speakers to judge the translation results, yet this evaluation is unfortunately left undone because of resource limitation.

6 Analysis

To some extent, poetry-English machine translation presents a coherent result in retrieving inversions and omissions of subject words and prepositions. Meanwhile, the machine's English word choices also expose the similar meaning of the original Chinese characters.

However, there do exist some shortcomings in the poetry translation results. Although neural poetry translation presents a readable result on behalf of the semantic meanings as the Mandarin-English translation do, difficulty still exists for obeying rhythm and rhyme rules in Chinese poetry, taking into account writing backgrounds, pronouns, and low-frequency words.

There are two reasons for difficulty in keeping rhythm and rhyme during the translation of Chinese poetry. Firstly, of translating two languages from different language families, appropriate rhyming words are hard to find. Secondly, the rhythm in Chinese poetry does not receive attention during understanding of poetry, since the main purpose of rhythm in Chinese poetry is to make the poem catchy instead of real meanings.

In the meantime, Chinese poetry often has embedded meanings or stories which are hard to summarize in English with few words. Normally in Mandarin texts, quoting a specific intricate things is often followed with understandable illustrations, which will help understanding and translation. However, due to the sentence pattern of limited number of words in Chinese poetry, few words are applied to represent such objects. In narrative Chinese poems, historical facts and stories are quoted to mapping specific incidents at the moment of writing, while in landscape Chinese poems, ancient geographical names applied, both of them making understanding harder. One story may uniquely utilized in one poem and often requires manual case-by-case analysis, which are hard to be applied in MT.

What's more, it is an interesting fact that the mandarin-English dataset has 2700 Chinese characters from 21116 sentences and poetry-English dataset has 2465 Chinese characters from 3130 stiches. Presenting this fact is to show that in Chinese poems, more rare words are applied, making translation even harder for machine to accomplish.

7 Future Works

7.1 Clustering/Positional Encoding

A possible pre-possessing technique to improve the poetry translation result can be clustering the similar meaning characters together because some characters in Chinese poetry are referencing to the same meaning. If this technique is applied, in the vocabulary set, there will be fewer entries because entries are replaced by regions. Thus during generation process, instead of calculating the possibility for each word, the system could calculate the possibility of regions in the vocabulary set. As a result, the likelihood to get an accurate meaning for one word in poem tends to increase.

7.2 Generation Order

Although self-attention mechanism is possible to relate two distant words, during the generation process, generation ordering is a meaningful task to be considered. Especially, translation of less well-aligned language pairs may be more sensitive to generation order, such as Chinese (Chan et al., 2019) and Chinese poetry. The reason for doing so lies in that most of poems do not contain subjects, making most of the translations having diverse first word, so having the noun generated first and insert others may offer improvements. In my implementation, I used a normal left-to-right generation order, but in the empirical study from Chan et al., they argue that Insertion(order) Transformer may achieve a 2.7 BLEU score comparing to left-to-right generation order on Chinese generation. Therefore, it will be an interesting attempt to apply generation ordering problems to poetry translation problem.

7.3 Back Translation

Neural Machine Translation is a corpus-based paradigm, the translation quality strongly depends on the quality and quantity of the training data provided. In addition, compared to phrase-based MT it requires more data to provide a good performance (Poncelas et al., 2018). Meanwhile, the poetry data is not that strong since it only contains 3130 stiches, so using back translation as a linguistic tool may provide more data to train the machine translation model.

8 Conclusion

In this paper I present a novel attempt to translate Chinese poetry to English for better communication of this traditional Chinese culture treasure. My model achieve BLEU score 26.63 on Mandarin-English semantic translation task. Then I translate Three Hundred Tang Poems and the results need human evaluation. The results show that machine translation works in translating semantic meanings of poems but fails to take into account deep implications which is one of the essential part of Chinese poetry. However, further improvements can be made to enable the system to learn these implications for the sake of enhancing translation results.

References

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

Bentivogli, L., Bisazza, A., Cettolo, M., & Federico, M. (2016). Neural versus phrase-based machine translation quality: a case study. *arXiv preprint arXiv:1608.04631*.

Chan, W., Stern, M., Kiros, J., & Uszkoreit, J. (2019). An empirical study of generation order for machine translation. *arXiv preprint arXiv:1910.13437*.

Chen, B., & Cherry, C. (2014). A systematic comparison of smoothing techniques for sentence-level bleu. In *Proceedings of the ninth workshop on statistical machine translation* (pp. 362–367).

Deng, L., Wang, J., Liang, H., Chen, H., Xie, Z., Zhuang, B., ... Xiao, J. (2019). An iterative polishing framework based on quality aware masked language model for chinese poetry generation. *arXiv preprint arXiv:1911.13182*.

Genzel, D., Uszkoreit, J., & Och, F. (2010). "poetic" statistical machine translation: Rhyme and meter.

Ghazvininejad, M., Choi, Y., & Knight, K. (2018). Neural poetry translation. In *Proceedings of the 2018 conference of the north american chapter of the association for computational linguistics: Human language technologies, volume 2 (short papers)* (pp. 67–71).

Poncelas, A., Shterionov, D., Way, A., Wenniger, G. M. d. B., & Passban, P. (2018). Investigating backtranslation in neural machine translation. *arXiv preprint arXiv:1804.06189*.

Sennrich, R., Haddow, B., & Birch, A. (2015). Neural machine translation of rare words with subword units. *arXiv preprint arXiv:1508.07909*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Wang, Z., He, W., Wu, H., Wu, H., Li, W., Wang, H., & Chen, E. (2016). Chinese poetry generation with planning based neural network. *arXiv preprint arXiv:1610.09889*.

Yan, R. (2016). i, poet: Automatic poetry composition through recurrent neural networks with iterative polishing schema. In *Ijcai* (pp. 2238–2244).

Yi, X., Sun, M., Li, R., & Yang, Z. (2018). Chinese poetry generation with a working memory model. *arXiv preprint arXiv:1809.04306*.