

Vietnamese Poem Generation & The Prospect Of Cross-Language Poem-To-Poem Translation

HUYNH MINH TRIET AND LE BAO QUAN

https://github.com/Anshler/poem_generator

Compiled March 25, 2023

Poetry generation has been a challenging task in the field of Natural Language Processing, as it requires the model to understand the nuances of language, sentiment, and style. Our proposed system will use Language Model and deep learning techniques to generate poems from natural language prompt that is both intuitive and offers greater control over the generated content. The model will be trained on a large corpus of poems written in a Vietnamese language, and will be evaluated using custom metrics tailored specifically for Vietnamese poem (which will be discussed in detail in later section)

Our team hopes that the results of this research will contribute to the ongoing efforts to advance the field of AI-generated poetry, especially in Vietnamese, and have potential applications in creative writing, education, and entertainment. © 2023 Optica Publishing Group

<http://dx.doi.org/10.1364/ao.XX.XXXXXX>

1. INTRODUCTION

With the recent advancement of the Causal Language Model for conversational chatbot, our team realises there is still a lack of research done in the creative field of language model, namely poetry generation.

The current implementation of poetry generation in Vietnamese is very limited in term of input flexibility, generating from a few starting words with little oversee for the body. Currently there is still no prominent model trained specifically in Vietnamese poems using the up-to-date GPT-3.

Thus, we set out to create a model that can take more complex inputs (i.e. prompts) such as instructions about theme, style or content, then the model will generate a poem accordingly, resulting in a creative, novel and unique poem with sentiment as intended

In the next section, we shall briefly take a look at some relevant papers about causal language model and poem generation, discussing their approaches, achievements and limitations.

2. LITERATURE REVIEW

Creative and novelty language models have been an interesting topic to pursue by many researchers. With recent advancement in Language Models have seen great results

and taking the world's attention by storm in the forms of OpenAI's ChatGPT and GPT-3 model [1], it is a great time to revisit this topic again since GPT-3 (Generative Pretrained Transformers) is proven to be excellent at understanding the abstract and complexity in casual text.

Most initial research on poetry generation facing a similar problem: the generative poems still lacking the deep comprehension of the content, but instead relies on templates [2], summarization [3] or machine translation [4]

More recent research have shown great advancement in this field, both in fluency and meaning of the poems. As can seen in [5] using RNN approach or [6] using attention based mechanisms [7][8]

Regardless, Tuan Nguyen et al.[9] with their SP-GPT2 is the only recent progress in Vietnamese poetry generation, using GPT-2 as the backbone of the model. In this paper, we aim to improve further by refining their method and combine with the latest GPT-3 model.

3. METHODOLOGY

The first thing of every AI project is to prepare an adequate dataset. Tuan Nguyen et al. already released their dataset with over 170000 poems of various genres. We then proceed to filter the 'good' poems, reducing the total sample count to 50000, by which, we develop a scoring system. This scoring system is also based partly on their

evaluation algorithms, expanded to encompass all genres, and shall be used for both the filtering process and post-training evaluation. How exactly we filter the dataset and how the scoring system works shall be discussed further in the Dataset section.

As for the backbone model for training this task, we tested various models from the state-of-the-art GPT-3 to open source and free to use model such as BLOOM [10]. For GPT-3 model, due to limited financial and computing resource, we only train a limited data on Davinci model (175B parameters), while mostly use Babbage (6B parameters) because it is much cheaper. And for Bloom, we use the 7.1B parameters version to be on par with Babbage.

A. Dataset

Preparing a scoring system for Vietnamese poem was a challenging task. Take Luc Bat genre for example, even though for most Vietnamese it is very familiar and easy to understand, its rule is actually quite complex. The rule is defined as follow:

- The 6th word's rhyme in the first line is relevant to the 6th word's rhyme in the second line.
- The 6th word's rhyme in the third line and the 6th word's rhyme in the fourth line is relevant to the 8th word's rhyme in the second line.
- Tones in each 6-word sentence are level, oblique and level in the 2th, 4th and 6th word's tone respectively.
- Tones in each 8-word sentence are level, oblique, level and level corresponding to the 2th, 4th, 6th and 8th word's tone.
- In each 8-word sentence, the 6th and 8th word are of different accent

Despite its complexity, and each genre has its own rule, we can still summarize that there are 3 main criteria to ensure the quality of a poem: length of each line, rhyme and tone.

We couple each pair of line of the poem to used as input, then proceed to give a score. Then we the average out this score over the whole poem and combine using the formula:

$$\text{score} = \frac{L}{10} + \frac{T \times 3}{10} + \frac{R \times 6}{10}$$

L : length score

T : tone score

R : rhyme score.

Such score could be adjusted and tinkered with depending on authors' preferences. The detail of for exact calculation for each genre of poem can be found in our repository. Even though how each genre is evaluated is slightly different, the scoring system is the same using 3 main criteria (length, tone and rhyme)

With that said, we filter all poems in the data, only taking ones with the score of 0.9 or higher, which results in cutting the amount of data by two third.

B. Method of approach

As discussed earlier, here is a summary of the whole process of this project. First, we collect and preprocess the dataset by the above mentioned method: building a filtering algorithm to score and filter high quality poems.

Then, the second step is to generate prompts to use as training input.

B.1. Text to poem model

Because the objective of this model is to let user input prompts of various length, context and requirement, we use the current state of the art GPT-3.5 API to generate the prompts. Detail of how we synthesize prompts is as follow:

$$P(X, Y, Z)$$

In this context, P is the template used to guide the creation of a poem. The genre of the poem is represented by the variable X , while the topic is represented by Y . The sequence of keywords to be included in the poem is represented by Z . So the generated result could be like:

"Write a genre X poem about Y , containing keywords Z "

Note that the actual generated prompts are in Vietnamese because we expected the model to be use in Vietnamese, both input and output.

After acquiring the dataset, we use GPT-3 (Da Vinci and Babbage) and BLOOM-7b1 to train. As discussed earlier, even though we are confident on the quality of our dataset, the financial, dataset quantity and computing limitation to build and train language model means we can only attempt with position of a student: we try to use as much as free resources as possible and the goal is to create a proof of concept which could be later picked up and improved.

B.2. Poem to poem pipeline

After the initial training, we realize there is a way that may expand the models' capability creatively.

We create a new dataset, in which, we turn the poems into pure texts, paraphrase and use the resulted texts as input prompts. This allows for the capturing of all context within the input to be use directly for generation, word by word.

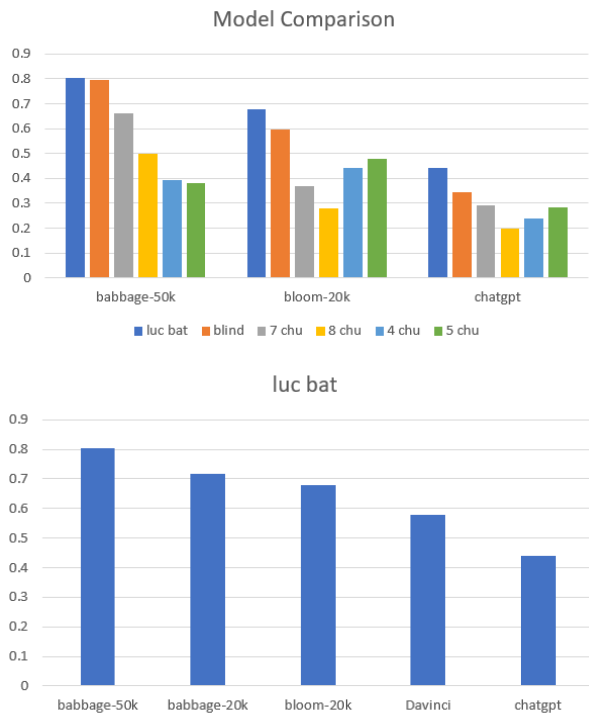
With this method, is possible to use any foreign piece of text, including foreign poems, preprocessed through a pipeline of translating to Vietnamese with online API such as Google Translate and using as input. For this method, we only train on the particular Luc Bat genre as it is the most popular genre of Vietnamese poems.

In the last section, we shall talk about the results of our model and make comparisons to see where we are with our method of approach

4. RESULT AND DISCUSSION

A. Our result

Below are our score using various models and different amount of training samples (for some models we could not evaluate other genre except for Luc bat due to resource limitation)



B. Comparison and discussion

B.1. Text to poem

The best result comes from training Babbage on 50000 samples with the score of 0.805 for Luc Bat and 0.795 for blind generation (meaning no genre specified in prompt). This is an expected result because it is trained on our whole available dataset and has tendency to generate Luc Bat if unspecified.

Since Google Colab can only train BLOOM model upto around 20000 samples, we also train the second Babbage model with 20000 samples for comparison. In this case, Babbage performs slightly better than BLOOM for Luc Bat genre (0.718 versus 0.678 respectively). Moreover, we can see that the reduction in the amount of samples significantly reduce the score.

And one may notice that for the Babbage 20k model, the 4 chu and 5 chu genre score abnormally high despite much fewer sample. This is due to underfitting which causes the model to repeat and duplicate its lines, leading to higher tone and rhyme score.

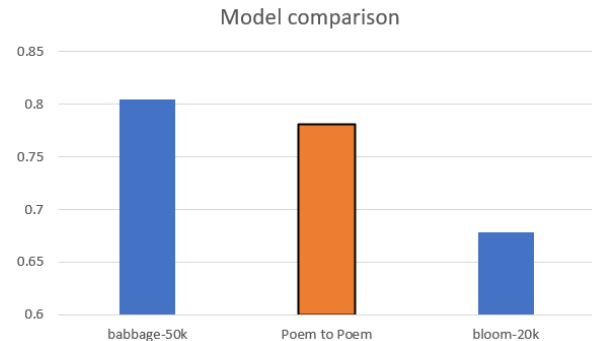
We also train using GPT-3 Davinci, yet, due to limited budget, the maximum afforded sample of 500 on 2 epochs scores very low at 0.58.

However, all of our results are significantly better than using Chat-GPT, which doesn't understand the specified

genre and always generates poems of inconsistent count-per-line.

B.2. Poem to poem

When we add a downstream task of paraphrasing then training using Babbage model, this yields the score of 0.781.



This is slightly lower than the previous Babbage-50k experiment even though the number of Luc Bat samples is the same. It is because the input prompts used are now much more complex, requiring the model to not only fit them into the generation, but also paraphrase them into synonyms that match the tone and rhyme rules.

Nevertheless, the lower score does not necessarily mean this approach is worse. Instead, this gives full control over the generated content, which the previous technique can manage barely with only topics and keywords.

Furthermore, this method opens up the new reality of a poem-to-poem translation pipeline. With an input poem or text from foreign language, it is possible to create a correct corresponded Vietnamese poem.

5. CONCLUSION AND FURTHER IMPROVEMENT

In conclusion, we believe our project results is adequately successful. We set out to build a model to generate Vietnamese poem creatively and accurately based on the current success of recent language models. The result is that only we made a Vietnamese poem generator, we also create poem filtering and scoring method, of which could be customized based on user need.

We would like to compare our project with SP-GPT2 however it is unfortunate that the repository is currently not working, and while we can see the score of the paper and the proposed model architecture, we could not verify how many sample they used to train to achieve such result. Hence, comparing would not yield significant meaning.

To further improve on this project, one possible way is to increase the dataset size, by artificially creating new data using the model and filter for high quality ones. But the most obvious solution is to train all the data on the more powerful davinci model. This would be prohibitively expensive for a small team, but more than achievable by companies, research institutions and the likes.

6. REFERENCES

- [1] Jacob Devlin, Ming-Wei Chang, Kenton Lee and Kristina Toutanova (2018) GPT-3, Language Model are few shot learner (<https://arxiv.org/abs/2005.14165>)
- [2] HG Oliveira (2012) Poetryme: a versatile platform or poetry generation
- [3] R. Yan (2013) I, poet: Automatic Chinese poetry composition through a generative summarization framework under constrained optimization
- [4] L Jiang (2008) Generating Chinese couplets using statistical approach
- [5] Ming-Wei Chang (2014) Chinese poetry generation with recurrent neural network
- [6] Q. Wang (2016) Chinese song iambics generation with neural attention-based model
- [7] Jacob Devlin, et (2018) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding (<https://arxiv.org/abs/1810.04805>)
- [8] Ashish Vaswani, et (2014) Attention is all you need (<https://arxiv.org/abs/1706.03762>)
- [9] Tuan Nguyen, Phong Nguyen (2020) SP-GPT2: Semantics Improvement in Vietnamese Poetry Generation (<https://github.com/fsoft-ailab/Poem-Generator>)
- [10] Big Science Workshop (2023) BLOOM: A 176B-Parameter Open-Access Multilingual Language Model (<https://arxiv.org/pdf/2211.05100.pdf>)