SI630 Project Proposal

Fangzhe Li
University of Michigan / 536 S. Forest Ave
School of Information / Ann Arbor, MI
`fangzhel@umich.edu`

1 Introduction

Classical Chinese poems excel themselves by their conciseness and elegance. It is always an interesting task to create poem. In homework 1, we have done a binary-classification problem in English. In this project, I initiate a task of creating Classical Chinese poems from several keywords. Through this project, I hope we can see the difficulties and challenges of doing nature language processing on a different language as mentioned in the first lecture of our class. Except for the reason that we want our "poet" to surprise us with beautiful rhythms, the other natural language generation tasks can also be benefited from the research on poem generation. (**?**) The problem to solve is that given some keywords, the output is supposed to be a quatrain with 7 Chinese characters in each line which focuses on the topic of keywords and fits the rhythm.

2 Task Definition and Data

2.1 Task Definition

I got some collections of classical Chinese poem as the corpus. I want to get a model that given some keywords from users as input and outputs a quatrain with 7 Chinese characters in each line which focuses on the topic of keywords and fits the rhythm. Let $X = (X_1, X_2, \ldots, X_n)$ be a sequence of keywords, where X_j represents jth word in X, the output will be $Y = (Y_1, Y_2, \ldots, Y_{28})$, where y_j represents jth character in Y.

2.2 Data

I found the data on the public repo in GitHub: https://github.com/chinese-poetry/chinese-poetry/tree/master/json There are 315,000 poems stored in json. Each data contains id, author, paragraphs and title.

3 Related Work

Generally, according to the methodology, works can be divided into three main categories: (1) templated-based method (2) statistical machine translation model (He et al., 2012) (3) deep learning method(Li et al., 2018) and (Zhang et al., 2017). In (He et al., 2012), the authors use the keywords as the input and generate the poem sentence by sentence through a phrase-based SMT model. In this way, each sentence takes all the previous sentences into consideration to ensure coherence between lines. Also, the authors make a comparison between BLEU and human evaluation to show BLEU metric is a good way to evaluate poem generation models. In (Li et al., 2018), the authors use CAVE to generate novelty and discriminator to ensure coherence. They combines CAVE with adversarial training. In (Zhang et al., 2017), a memory-augmented neural model is used to solve the problem that the model only generate poems based on general rule and has very few innovations.

4 Evaluation

Since it is very subjective to judge whether a poem is a good way, it is very hard to find a good criteria to assess the poems generated. After reading the literature in the related works part, we find there are two main ways to evaluate the machine-generated poems: human expert evaluation and BLEU (Papineni et al., 2002), which is a metric used widely in machine translation.

Human expert evaluation: In most of paper related to poem generation, human expert evaluation is used as the main criteria for evaluation because the poem is relatively complicated form of literature and it is very hard to quantitatively evaluate. Human evaluation can be done by setting several group of standard: Conciseness, Elegance, Rhythm.

Another task can be done by human evaluation is that let experts tell if one poem is written by human poets or generated by our model, we can get the accuracy as the criteria. I expect to have the accuracy of distinguishing below 70

BLEU: Another quantitative way to evaluate poems is the BLEU. BLEU is initially used in machine translation to evaluate the similarity between human translation and generated machine translation. Here a slightly different way will be used: first keywords is extracted from a human-written poem, then these keywords are put into model as input, we compare the BLEU between the output the the original poem. Here a low BLEU value is expected because there are more ways to write a poem compared to general translation. So BLEU is better used to compare the result of different poem generation model. In (He et al., 2012), the author made an analysis to show BLEU actually a good criteria because it shows some relationship with the result of human evaluation.

5 Work Plan

I will start the project by embedding the dataset and choosing models. This process is expected to finished before the project update, in another word, from week 6 to 8. During this process, I will discuss with the instructor team and read more related work to figure out the blueprint for the solution. From week 9 to 12, I am expected to finish the implementation of the model. From week 13 to 15, I will finish the presentation and evaluation part as well as the report for project.

References

Jing He, Ming Zhou, and Long Jiang. 2012. Generating chinese classical poems with statistical machine translation models. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, AAAI'12, page 1650–1656. AAAI Press.

Juntao Li, Yan Song, Haisong Zhang, Dongmin Chen, Shuming Shi, Dongyan Zhao, and Rui Yan. 2018. Generating classical chinese poems via conditional variational autoencoder and adversarial training. pages 3890–3900.

Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: A method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, page 311–318, USA. Association for Computational Linguistics.

Jiyuan Zhang, Yang Feng, Dong Wang, Yang Wang, Andrew Abel, Shiyue Zhang, and Andi Zhang. 2017. Flexible and creative Chinese poetry generation using neural memory. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1364–1373, Vancouver, Canada. Association for Computational Linguistics.