1 Introduction

Classical Chinese poetry is a fundamentally important section of Chinese literature. With over 2000 years of history, it is a precious and special heritage of Chinese culture. Classical Chinese poetry covers a wide range of topics, including traveling, farewell, lovesickness, homesickness, ambition, life insights, and even comments on historical events and figures. The diversity of themes and richness of sentiments make it attractive to people since the ancient time, and now it is sparking much interest worldwide.

One of the most noticeable feature of classical Chinese is the usage of condensed language to construct imagery and convey sentiments. Most words are unigram characters, which makes classical Chinese a highly concise language. The pronunciation of classical Chinese characters falls into two categories, the level tone (Ping 平) and the oblique tone (Ze 仄). Each tone can be further divided into more subcategories, and the combination of different tones often leads to a sense of rhythm.

In classical Chinese poetry, a couplet consists of two sentences. According to the number of couplets in a poem, it can be divided into regulated verses (4 couplets or 8 sentences) and quatrains (2 couplets or 4 sentences). In addition to the semantic coherence, there are three additonal constraining patterns which need to be statified when composing the poems. For most poems, the structural pattern requires that all the sentences have the same length, with each consisting of either 5 or 7 characters. The rhythmical pattern requires that the ending characters of certain sentences bear similar vowels. For regulated verse, the second, fourth, sixth and eighth sentences should rhyme, whereas for quatrains are the first, second and fourth sentences. The tonal pattern requires characters in particular positions hold particular tones, so that the alternation of tones make the poems acoustically beautiful.

Another important rhetorical feature of classical Chinese poetry is the syntactically parallel structure. The weak parallelism (对偶) requires the corresponding characters of both sentences within a couplet have the same part of speech tag, fall into the same semantic category and bear the similar or opposite meaning; the strong parallelism (对仗) addionally requires different tones of corresponding characters. The use of syntactically parallel structure is ubiquitous, and since the middle of Tang dynasty it became a requirement for regulated verses that the second and third couplet should adopt the strong parallel structure.

江雪
Snow on the River
千山鸟飞**绝**，万径人踪**灭**。
From hill to hill no bird in flight; From path to path no man in sight.
孤舟蓑笠翁，独钓寒江**雪**。
A lonely fisherman afloat, Is fishing snow on lonely boat.

题李凝幽居
Li Ning's Seclusion
闲居少邻并，草径入荒**园**。
You may keep your seclusion if you please; A grassy path leads to your garden poor.
鸟宿池边树，僧敲月下**门**。
Only birds perch upon the poolside trees; Rarely a monk knocks at your moonlit door.
过桥分野色，移石动云**根**。
A bridge divides the fileds from rural plain; Clouds float as if rocks were moving on high.
暂去还来此，幽期不负**言**。
Leaving your place now, I will come again; What I have promised, I will not belie.

Table 1: An example of a quatrain and a regulated verse with rhyming characters in boldface. English translation by Xu et al. [2014]. The first couplet of the quatrain confirms to the weak parallel structure; the second and third couplets of the regulated verse comply to the strong one.

2 Related Work

Poem generation is an interesting yet challenging task in natural language generation. In recent years, there have been tremendous interest in computational analysis and generation of classical Chinese poetry. Hou and Frank [2015] adapted Weighted Personalized PageRank (WPR) to build a sentiment lexicon for classical Chinese poetry and extracted topics associated with positive and negative sentiments. Jiang and Zhou [2008] employed statistical machine translation (SMT) models to generate Chinese couplets confirming to the parallel structure, and this method was later adopted by He et al. [2012] to generate quatrains. Another approach is based on summarization [Yan et al., 2013]. In their framework, the poems relevant to input keywords are retrieved and segmented to terms, which are then clustered and re-organized iteratively to generate sentences that comply with the regulation patterns.

Recently, deep neural network has gained much interest in natural language generation, and they have been widely applied to composing Classical Chinese poems. Zhang and Lapata [2014] proposed a framework in which convolutional neural networks are used to model sentences and recurrent neural networks are used to model the entire context and generate current sentence. Wang et al. [2016b] proposed an planning based approach, where each line is generated based on a keyword and the context of previous lines. The model was later adapted by Wang et al. [2016a] to generate Song Iambics, where the length of each sentence varies according to the tunes. Zhang et al. [2017] further improved this model by augmenting a memory component to the neural network. However, all these methods failed to explicitly capture the tones and rhymes of Chinese characters. Some assumed the rhythmical and tonal pattern can be learned in the training process, which is often not the case; other models adopted post-precessing to fit their poems to the regulation patterns. In addition, those models are poor at reconstructing the syntactically parallel structure. Yan et al. [2016] employed a polishing scheme to refine the generation of Chinese couplet, but their model are designed exclusively for generating the second sentence given the first one.

[The following is to be finished after implementation done]

The goals of this work are

- to better capture the tonal and rhythmical restriction with neural network

- to reconstruct the parallel structure of certain couplets in the poems

- to improve the semantic coherence of the poems

References

Jing He, Ming Zhou, and Long Jiang. Generating chinese classical poems with statistical machine translation models. In *AAAI*, 2012.

Yufang Hou and Anette Frank. Analyzing sentiment in classical chinese poetry. In *Proceedings of the 9th SIGHUM Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH)*, pages 15–24, 2015.

Long Jiang and Ming Zhou. Generating chinese couplets using a statistical mt approach. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 377–384. Association for Computational Linguistics, 2008.

Qixin Wang, Tianyi Luo, Dong Wang, and Chao Xing. Chinese song iambics generation with neural attention-based model. *arXiv preprint arXiv:1604.06274*, 2016a.

Zhe Wang, Wei He, Hua Wu, Haiyang Wu, Wei Li, Haifeng Wang, and Enhong Chen. Chinese poetry generation with planning based neural network. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 1051–1060, 2016b.

Yuanchong Xu, Hongmei Li, Haiyun Guo, Min Gao, and Yigao Wang. *100 Classical Chinese Poems with Chinese-English Interpretations*. Dalian Publishing House, Dalian, China, 7 2014. ISBN 978-7-5505-0622-0.

Rui Yan, Han Jiang, Mirella Lapata, Shou-De Lin, Xueqiang Lv, and Xiaoming Li. i, poet: Automatic chinese poetry composition through a generative summarization framework under constrained optimization. In *IJCAI*, pages 2197–2203, 2013.

Rui Yan, Cheng-Te Li, Xiaohua Hu, and Ming Zhang. Chinese couplet generation with neural network structures. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 2347–2357, 2016.

Jiyuan Zhang, Yang Feng, Dong Wang, Yang Wang, Andrew Abel, Shiyue Zhang, and Andi Zhang. 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)***, volume 1, pages 1364–1373, 2017.

Xingxing Zhang and Mirella Lapata. Chinese poetry generation with recurrent neural networks. In ***Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)***, pages 670–680, 2014.