Can Machine Generate Traditional Chinese Poetry? A Feigenbaum Test

Qixin Wang[1,4]***, Tianyi Luo[1,3], and Dong Wang[1,2]†

[1]CSLT, RIIT, Tsinghua University, China
[2]Tsinghua National Lab for Information Science and Technology, Beijing, China
[3]Huilan Limited, Beijing, China
[4]CIST, Beijing University of Posts and Telecommunications, China
{wqx,lty}@cslt.riit.tsinghua.edu.cn
wangdong99@mails.tsinghua.edu.cn

Abstract. Recent progress in neural learning demonstrated that machines can do well in regularized tasks, e.g., the game of Go. However, artistic activities such as poem generation are still widely regarded as human's special capability. In this paper, we demonstrate that a simple neural model can imitate human in some tasks of art generation. We particularly focus on traditional Chinese poetry, and show that machines can do as well as many contemporary poets and weakly pass the Feigenbaum Test, a variant of Turing test in professional domains.
Our method is based on an attention-based recurrent neural network, which accepts a set of keywords as the theme and generates poems by looking at each keyword during the generation. A number of techniques are proposed to improve the model, including character vector initialization, attention to input and hybrid-style training. Compared to existing poetry generation methods, our model can generate much more theme-consistent and semantic-rich poems.

1 Introduction

The classical Chinese poetry is a special cultural heritage with over $2,000$ years of history and is still fascinating many contemporary poets. In history, Chinese poetry flourished in different genres at different time, including Tang poetry, Song iambics and Yuan songs. Different genres possess their own specific structure, rhythmical and tonal patterns. The structural pattern regulates how many lines and how many characters per line; the rhythmical pattern requires that the last characters of certain lines hold the same or similar vowels; and the tonal pattern requires characters in particular positions hold particular tones, i.e., 'Ping' (level tone), or 'Ze' (downward tone). A good poem should follow all these pattern regulations (in a descendant order of priority), and has to express a consistent theme as well as a unique emotion. For this reason, it is widely admitted that traditional Chinese poetry generation is highly difficult.

*** The first two authors contributed equally.

† Corresponding author: Dong Wang; RM 1-303, FIT BLDG, Tsinghua University, Beijing (100084), P.R. China. This work was supported by the National Science Foundation of China (NSFC) under the project NO.61371136, and the MESTDC PhD Foundation Project No.20130002120011. It was also supported by Huilan Ltd. and FreeNeb.

Among all the genres of traditional Chinese poetry, perhaps the most popular is the quatrain, a special style with a strict structure (four lines with five or seven characters per line), a regulated rhythmical form (the last characters in the second and fourth lines must follow the same rhythm), and a required tonal pattern (tones of characters in some positions should satisfy some pre-defined regulations). This genre of poems flourished mostly in Tang Dynasty, so often called 'Tang poem'. An example of quatrain written by Lun Lu, a famous poet in Tang Dynasty [16], is shown in Table 1.

塞下曲 Frontier Songs		
月黑雁飞**高**，	(*ZZPP)	The wild goose flew high to the moon shaded by the cloud,
单于夜遁**逃**。	(PPZZP)	With the dark night's cover escaped the invaders crowd,
欲将轻骑逐，	(*PPZZ)	I was about to hunt after them with my cavalry,
大雪满弓刀。	(*ZZPP)	The snow already covered our bows and swords.

Table 1. An example of a quatrain. The rhyming characters are in boldface, and the tonal pattern is shown at the end of each line, where 'P' indicates level tone and 'Z' indicates downward tone, and '*' indicates the tone can be either.

Due to the stringent restriction in rhythm and tone, it is not trivial to create a fully rule-compliant quatrain. More importantly, besides such strict regulations, a good quatrain should also read fluently, hold a consistent theme, and express a unique affection.

We are interested in machine poetry generation, not only because of its practical value in entertainment and education, but also because it demonstrates an important aspect of artificial intelligence: the creativity of machines. We hold the belief that poetry generation (and other artistic activities) is a pragmatic process and can be largely learned from past experience. In this paper, we focus on traditional Chinese poetry generation, and demonstrate that machines can do it as well as many human poets.

There have been some attempts in this direction, e.g., by machine translation models [6] and recurrent neural networks (RNN) [22]. These methods can generate traditional Chinese poems with different levels of quality, and can be used to assist people in poem generation. However, none of them can generate poems that are fluent and consistent enough, not to mention innovation.

In this paper, we propose a simple neural approach to traditional Chinese poetry generation based on the attention-based Gated Recurrent Unit (GRU) model. Specifically, we follow the sequence-to-sequence learning architecture that uses a GRU [3] to encode a set of keywords as the theme, and another GRU to generate quatrains character by character, where the keywords are looked back during the entire generation process. By this approach, the generation is regularized by the keywords so a global theme is assured. By enriching the set of keywords, the generation tends to be more 'innovative', resulting in more diverse poems. Our experiments demonstrated that the new approach can generate traditional Chinese poems pretty well and even pass the Feigenbaum Test.

2 Related work

A multitude of methods have been proposed for poem automatic generation. The first approach is based on rules and templates. For example, [15] and [18] employed a phrase search approach for Japanese poem generation, and [11] proposed an approach based on word association norms. [12] and [13] used semantic and grammar templates for Spanish poem generation.

The second approach is built on various genetic algorithms. For example, [23] proposed to use a stochastic search algorithm to obtain the best matched sentences. The search algorithm is based on four standards proposed by [9]: fluency, meaningfulness, poeticness, and coherence.

The third approach involves various statistical machine translation (SMT) methods. This approach was used by [7] to generate Chinese couplets, a special regulated verses with only two lines. [6] extended this approach to Chinese quatrain generation, where each line of the poem is generated by translating the preceding line.

Another approach to poem generation is based on text summarization. For example, [19] proposed a method that retrieves high-ranking candidates of sentences from a large poem corpus, and then re-arranges them to generate rule-conformed new sentences.

More recently, deep learning methods gain much attention in poem generation. For example, [22] proposed an RNN-based approach that was reported to work well in quatrain generation [22]; however, the structure seems rather complicated (a CNN and two RNN components in total), preventing it from extending to other genres. Our model is a simple sequence-to-sequence structure, which is much simpler than the model proposed by [22] and can be easily extended to more complex genres such as Sonnet and Haiku without modification.

Finally, [17] proposed an attention-based model for Song Iambics generation. However, their model performed rather poor when was applied directly to quatrain generation using keywords input, possibly because quatrains are more condensed and more individually unique than iambics. Our approach follows the attention-based strategy in [17], but introduces several innovations. Firstly, the poems were generated through key words rather than the first sentence to provide more clear themes; Secondly, a single-word attention mechanism was used to improve the sensitivity to key words; Thirdly, a loop generation approach was proposed to improve the fluency and coherence of the attention-based model.

3 Method

In this section, we first present the attention-based Chinese poetry generation framework, and then describe the implementation of the encoder and decoder models that have been tailored for our task.

3.1 Attention-based Chinese Poetry Generation

The attention-based sequence-to-sequence model proposed by [1] is a powerful framework for sequence generation. Specifically, the input sequence is converted by an 'encoder' to a sequence of hidden states to represent the semantic status at each position of the input, and these hidden states are used to regulate a 'decoder' that generates the target sequence. The important mechanism of the attention-based model is that at each generation step, the most relevant input units are discovered by comparing the 'current' status of the decoder with all the hidden states of the encoder, so that the generation is regulated by the fine structure of the input sequence.

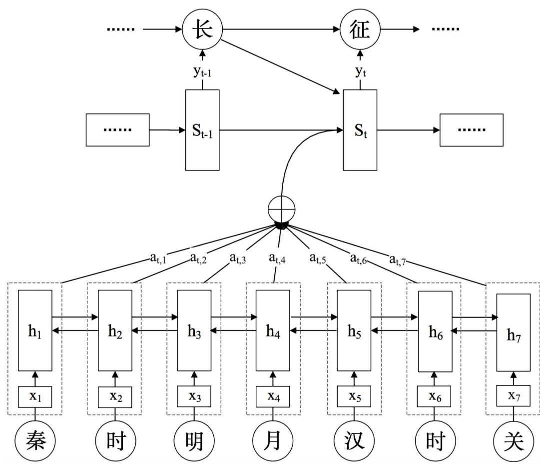


Fig. 1. The attention-based sequence-to-sequence framework for Chinese poetry generation.

The entire framework of the attention-based model applied to Chinese poetry generation is illustrated in Figure 1. The encoder (a bi-directional RNN) converts the input keywords, a character sequence denoted by $(x_1, x_2, ...)$, into a sequence of hidden states $(h_1, h_2, ...)$. The decoder (another RNN) then generates the whole poem character by character, denoted by $(y_1, y_2, ...)$. At each step t, the prediction for the next character y_t is based on the 'current' status s_t of the decoder as well as all the hidden states $(h_1, h_2, ..., h_T)$ of the encoder. Each hidden state h_i contributes to the generation according to a relevant factor $\alpha_{t,i}$ that measures the similarity between s_t and h_i. To alleviate the problem that vanilla RNN tend to forget historical input quickly[21], we used GRU in every RNN in our model to provide a strong memory for theme.

3.2 Model Training

The goal of the training is to let the predicted character sequence match the original poem. We chose the cross entropy between the distributions over Chinese characters given by the decoder and the ground truth (essentially in a one-hot form) as the objective function. To speed up the training, the minibatch stochastic gradient descent (SGD) algorithm was adopted. The gradient was computed in sentences, and the AdaDelta algorithm was used to adjust the learning rate [21]. In the training phase, there are no keyword input, so we use the first line as the input to generate the entire poem.

4 Implementation

The basic attention model does not naturally work well for Chinese poetry generation. A particular problem is that every poem was created to express a special affection of the poet, so it tends to be 'unique'. This means that most valid (and often great) expressions can not find sufficient occurrence in the training data. Another problem is that the theme may become vague towards the end of the generation, even with the attention mechanism. Several techniques are presented to improve the model.

4.1 Character Vector Initialization

Since the uniqueness of each poem, it is not easy to train the attention model, as many expressions are not statistically significant. This is a special form of data sparsity. A possible solution is to train the model in two steps: firstly learn the semantic representation of each character, possibly using a large external corpus, and then train the attention model with these pre-trained representations. By this approach, the model most focuses on possible expressions and hence is easier to train. In practice, we first derive character vectors using the word2vec tool[1], and then use these character vectors to initialize the word embedding matrix in our model. Since part of the model (embedding matrix) has been pre-trained, the problem of data sparsity can be largely alleviated.

4.2 Input Reconstruction

Poets tend to express their feelings following an implicit theme, instead of an explicit reiteration. We found this implicit theme is not easy for machines to understand and learn, leading to possible theme drift at run-time. A simple solution is to force the model to reconstruct the input after it has generated the entire poem. More specifically, in the training phase, we use the first line of a training poem as the input, and based on this input to generate five lines sequentially: line 1-2-3-4-1. The last generation step for line 1 forces the model to keep the input in mind during the entire generation process, so learns how to focus on the theme.

4.3 Input Vector Attention

The popular configuration of the attention model attends on hidden states. Since hidden states represent *accumulated* semantic meaning, this attention is good to form a global theme. However, as the semantic contents of individual keywords have been largely averaged, it is hard to generate diverse poems sensitive to each and different keywords.

We propose a multiple-attention solution that attends on both hidden states and input character vectors, so that both accumulated and individual semantics are considered during the generation. It has been found that this approach is highly effective for generating diverse and novel poems: just given sufficient keywords, new poems can be generated with high quality. Compared to other approaches such as noise injection or

[1] `https://code.google.com/archive/p/word2vec/`

n-best inference, this approach can generate unlimited alternatives without any quality sacrifice. Interestingly, our experiments show that more keywords tend to generate more unexpected but highly impressive poems. Therefore, the multiple-attention approach can be regarded as an interesting way to promote innovation.

4.4 Hybrid-style Training

Traditional Chinese quatrains are categorized into 5-char quatrains and 7-char quatrains that involve five and seven characters per line, respectively. These two categories follow different regulations, but also share the same words and similar semantics. We utilize the hybrid-style training method as [17] that trains the two types of quatrains using the same model, with a 'type indicator' derived from eigen vectors of a 200×200 dimensional random matrix, to notify the model the type of present training sample.

5 Experiments

We describe the experimental settings and results in this section. Firstly the datasets used in the experiments are presented, and then we report the evaluation in two phases: (1) the first phase focuses on searching for optimal configurations for the attention model; (2) the second phase compares the attention model with other methods; (3) the third phase is the Feigenbaum Test.

5.1 Datasets

Two datasets are used to conduct the experiments. Firstly a Chinese quatrain corpus was collected from Internet. This corpus consists of $13,299$ 5-char quatrains and $65,560$ 7-char quatrains. As far as we know, this covers most of the quatrains that are retained today. We filters out some poems which contain 100% low frequency words. Through corpus cleaning, a corpus which contains $9,195$ 5-char quatrains and $49,162$ 7-char quatrains was obtained. $9,000$ 5-char and $49,000$ 7-char quatrains are used to train the GRU model of the attention model and LSTM model of a comparative model based on RNN language models and the rest poems are used as the test datasets.

The second dataset was used to train and derive character vectors for attention model initialization. This dataset contains $284,899$ traditional Chinese poems in various genres, including Tang quatrains, Song iambics, Yuan Songs, Ming and Qing poems. This large amount data ensures a stable learning for semantic content of most characters.

5.2 Model Development

In the first evaluation, we intend to find the best configurations for the proposed attention-based model. The 'Bilingual Evaluation Understudy' (BLEU) [14] is used as the metric to determine which enhancement techniques are effective. BLEU was originally proposed to evaluate machine translation performance [14], and was used by [22] to evaluate quality of poem generation. We used BLEU in the development phase to determine which design option to choose, without the costly human evaluation.

The method proposed by [6] and employed by [22] was adopted to obtain reference poems. A slight difference is that the reference set was constructed for each *input keyword*, instead of each sentence as in [22]. This is because our attention model generates poems as an entire character sequence, while the vanilla RNN approach in [22] does that sentence by sentence. Additionally, we used 1-gram and 2-grams in the BLEU computation, according to the fact that semantic meaning is mostly represented by single characters and some character pairs in traditional Chinese.

Model	BLEU	
	5-char	7-char
Basic model	0.259	0.464
+ All poem training	0.267	0.467
+ Input Reconstruction	0.268	0.500
+ Input Vector Attention	0.290	0.501
+ Hybrid training	**0.330**	**0.630**

Table 2. BLEU scores with various enhancement techniques.

Table 2 presents the results. The baseline model is trained with character initialization where the character vectors are trained using quatrains only. This is mostly the system in [17]. Then we use the large corpus that involves all traditional Chinese poems to enhance the character vectors, and the results demonstrated a noticeable performance improvement in fluency (from our human judgements) and a small improvement in BLEU (2nd row in Table 2). This is understandable since poems in different genres use similar languages, so involving more training data helps infer more reliable semantic content for each character. Additionally, we observe that reconstructing the input during model training improves the model (3rd row). This is probably due to the enhancement in theme consistence. What's more, attention to both input vectors and hidden states leads to additional performance gains (4th row). Finally, the hybrid-style training is employed to train a single model for the 5-char and 7-char quatrains. The BLEUs are tested on 5-char and 7-char quatrains respectively and the results are shown in the 5-th row of Table 2. Note that in the hybrid training, we stop the training before convergence in favor of a good BLEU.

From these results, we obtain the best configuration that involves character vector trained with extern training data, input reconstruction, input vector attention and hybrid training. In the reset of the paper, we will use this configuration to train the attention model (denoted by 'Attention') and compare it with the comparative methods.

5.3 Comparative Evaluation

In the second phase, we compare the attention model (with the best configuration) and three comparative models: the SMT model proposed by [6], the vanilla RNN poem generation (RNNPG) proposed by [22], and an RNN language model (RNNLM) that

Model	Compliance		Fluency		Consistence		Aesthesis		Overall	
	char-5	char-7	char-5	char-7	char-5	char-7	char-5	char-7	char-5	char-7
SMT	3.04	2.83	2.28	1.92	2.15	2.00	1.93	1.67	2.35	2.10
LSTMLM	3.00	3.71	2.39	3.10	2.19	2.88	2.00	2.66	2.39	3.08
RNNPG	2.90	2.60	2.05	1.70	1.97	1.70	1.70	1.45	2.15	1.86
Attention	**3.44**	**3.73**	2.85	3.13	2.77	2.98	2.38	2.87	2.86	3.17
Human	3.33	3.54	**3.37**	**3.33**	**3.45**	**3.26**	**3.05**	**2.96**	**3.30**	**3.27**

Table 3. Averaged ratings for Chinese quatrain generation with different methods. 'char-5' and 'char-7' represent 5-char and 7-char characters quatrains respectively in the evaluation.

can be regarded as a simplified version (One-direction LSTM RNN neural network without attention mechanism) of the attention model [10].

Following the work of [22], we selected 30 subjects (e.g., falling flower, stone bridge, etc.) in the Shixuehanying taxonomy [8] as 30 themes. For each theme, several phrases belonging to the corresponding subject were selected as the input keywords. For the attention model, these keywords were used to generate the first line directly; For the other three models, however, the first line had to be constructed beforehand by an external model. We chose the method provided by [22] to generate the first lines for the SMT, vanilla RNN and LSTMLM approaches. A 5-char quatrain and a 7-char quatrain were generated for each theme by the four methods, and were evaluated by experts.

For reference, some poems written by ancient poets were also involved in the evaluation. To prevent the impact of prior knowledge of the experts, we deliberately chose the poems written by poets that are not very famous. The poems were chosen from [5], [20] and [2]; and a 5-char quatrain and a 7-char quatrain were selected for each theme.

The evaluation was conducted by experts based on the following four metrics, in the scale from 0 to 5:

– Compliance: if the poem satisfies the regulation on tones and rhymes;
– Fluency: if the sentences read fluently and convey reasonable meaning;
– Consistence: if the poem adheres to a single theme;
– Aesthesis: if the poem stimulates any aesthetic feeling.

In the experiments, we invited 26 experts to conduct a series of scoring evaluations[2]. These experts were asked to rate the generation of our model and three comparative approaches: SMT, LSTMLM, and RNNPG. The SMT-based approach is available online[3] and we use this online service to obtain the generation. For RNNPG, we invited the authors to conduct the generation for us. The LSTMLM approach was implemented by ourselves, for which we used the GRU instead of the vanilla RNN to enhance long-distance memory, and used character vector initialization to improve model training.

[2] These experts are professors and their postgraduate students in the field of Chinese poetry research. Most of them are from the Chinese Academy of Social Sciences (CASS).
[3] http://duilian.msra.cn/jueju/

Poems written by ancient poets are also involved in the test. For each method (including human-written), a 5-char quatrain and a 7-char quatrain were generated or selected for each of the 30 themes, amounting to 300 poems in total in the test. For each expert, 80 poems were randomly selected for evaluation.

Table 3 presents the results. It can be seen that our model outperforms all the comparative approaches in terms of all the four metrics. More interestingly, we find that the scores obtained by our model are approaching to those obtained by human poets, especially with 7-char poems. This is highly encouraging and indicates that our model can imitate human beings to a large extent, at least from the eyes of contemporary experts.

The second best approach is the LSTMLM approach. As we mentioned, LSTMLM can be regarded as a simplified version of our attention model, and shares the same strength in LSTM-based long-distance pattern learning and improved training strategy with character vector initialization. This demonstrated that a simple neural model with little engineering effort may learn artistic activities pretty good. Nevertheless, the comparative advantage of the attention model still demonstrated the importance of the attention mechanism.

The RNNPG and the SMT approaches perform equally worse, particularly RNNPG. A possible reason is that RNNPG requires an SMT model to enhance the performance but the SMT model was not used in this test[4]. In fact, even with the SMT model, RNNPG can hardly approach to human as the attention model does, as shown in the original paper [22]. The SMT approach, with a bunch of unknown optimizations by the Microsoft colleagues, can deliver reasonable quality, but the limitation of the model prevents it from approaching a human-level as our model does. The T-test results show that the difference between the attention LSTM model (ours) and the vanilla RNN and SMT are both significant (p ¡ 0.01), though the difference between the attention LSTM model and LSTMLM is weakly significant ($p = 0.03$).

It is noticeable that the human ratings of human-written poems are lower than the ratings reported by [22]. We are not sure the experts that Zhang and Lapata invited, but the experts in our experiments are truly professional and critical: most of them are top-level experts on classical Chinese poetry education and criticism, and some of them are winners of national competitors in classic Chinese poetry writing.

Interestingly, in the metric of compliance, our attention model outperforms human. This is not surprising as computers can simply search vast candidate characters to ensure a rule-obeyed generation. In contrast, human artists put meaning and affection as the top priority, so sometimes break the rule.

Finally, we see that the quality of the 7-char poems generated by our model is very close to that of the human-written poems. This should be interpreted in two perspectives: On one hand, it indicates that our generation is rather successful; On the other hand, we should pay attention that the poems we selected are from infamous poets. Our intention was to avoid biased rating caused by experts' prior knowledge on the poems, but this may have limited the quality of the selected poems, although we have tried our best to choose good ones.

[4] The author of RNNPG [22] could not find the SMT model in the reproduction, unfortunately.

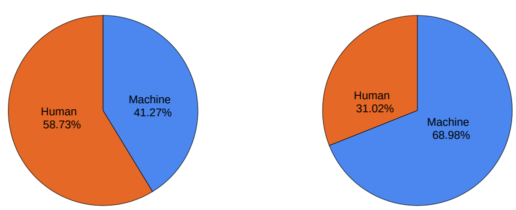
5.4 Feigenbaum Test

We design a Feigenbaum Test [4] to evaluate the quality of poems generated by our models. Feigenbaum test (FT) can be regarded as a generalized Turing test (TT), the most well-known method for evaluating AI systems. A particular shortcoming of TT is that it is only suitable for tasks involving interactive conversions. However, there are many professional domains where no conversations are involved but still require a simple method like TT to evaluate machines' intelligence. Feigenbaum Test follows the core idea of TT, but focuses on professional tasks that can be done only by domain experts. The basic idea of FT is that an intelligent system in a professional domain should behave as a human expert, and the behavior can not be *distinguished from* human experts, when *judged by* human experts in the same domain. We believe that this is highly important when evaluating AI systems on artistic activities, for which mimicking the behavior of human experts is an important indicator of its success.

In this section, we follow this idea and utilize FT to evaluate the poetry generation models. Specifically, we distributed the 30 themes to some experts in traditional Chinese poem generation[5]. We asked these experts to select one theme that they are most favor so that the quality can be ensured.

We received 144 submissions. To ensure the quality of the submission, we generated the same number of poems by our model and then asked a highly-professional expert in traditional Chinese poem criticism to give the first round filtering. After the filtering, 83 human-written poems (57.6%) and 180 computer-generated poems (86.5%) were remained, respectively. This indicates that human-generated poems are in a larger variance in quality, which is not very surprising as the knowledge and skill of people tend to vary significantly.

The remained 263 poems were distributed to 24 experts for evaluation[6]. The experts were asked to answer two questions: (1) if a poem was generated by people; (2) quality of a poem, rated from 0 to 5.



(a) (b)

Fig. 2. Decision option for (a) human-written (b) machine-generated poems.

[5] These experts were nominated by professors in the field of traditional Chinese poetry research.

[6] These experts again are mostly from CASS, part of them attended the previous test.

The results of the human-machine decision are presented in Figure 2. For a clear representation, the minor proportions of zero scores are omitted in the figure. We observe that 41.27% of the human-written poems were identified as machine-generated, and 31% of the machine-generated poems were identified as human-written. This indicates that a large number of poems can not be correctly identified by people. According to the criterion of Turing Test(Actually, Feigenbaum Test can be regarded as domain specific "Turing Test"), our model has weakly passed[7]. Interestingly, among the top-5 high-ranked poems, the machine takes the position 1 and 2, and among the top 10, the machine takes the position 1, 2 and 7. This means that our model can generate better poems even than human poets, although in general it is still beat by human.

5.5 Generation Example

Finally we show a 7-char quatrain generated by the attention model. The theme of this poem is 'crab-apple flower'.

海棠花 Crab-apple Flower	
红霞淡艳媚妆水，	Like the rosy afterglows with light make-up being sexy,
万朵千峰映碧垂。	Among green leaves, thousands of crabapples blossoms make the branch droopy.
一夜东风吹雨过，	After a night of wind and shower,
满城春色在天辉。	With the bright sky, spring is all over the city.

Table 4. A quatrain example generated by the attention model.

6 Conclusion

This paper proposed an attention-based neural model for Chinese poetry generation. Compared to existing methods, the new approach is simple in model structure, strong in theme preservation, flexible to produce innovation, and easy to be extended to other genres. Our experiments show that it can generate traditional Chinese quatrains pretty well. A future work will employ more generative models, e.g. variational generative deep models, to achieve more natural innovation. We also plan to extend the work to other genres of traditional Chinese poetry, e.g., Yuan songs.

7 References

References

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

[7] The criterion is to fool people in more than 30% of the trials. Refer to `https://en.wikipedia.org/wiki/Turing_test`.

2. Chen, Y.: Jin Poetry Chronicle. Shanghai Chinese Classics Publishing House (2013)
3. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., Bengio, Y.: Learning phrase representations using rnn encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078 (2014)
4. Feigenbaum, E.A.: Some challenges and grand challenges for computational intelligence. Journal of the ACM (JACM) 50(1), 32–40 (2003)
5. Han, W.: Study on the Liao, Jin and Yuan Poetry and Music Theory. China Social Sciences Press (2015)
6. He, J., Zhou, M., Jiang, L.: Generating chinese classical poems with statistical machine translation models. In: Twenty-Sixth AAAI Conference on Artificial Intelligence (2012)
7. Jiang, L., Zhou, M.: Generating chinese couplets using a statistical mt approach. In: Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1. pp. 377–384. Association for Computational Linguistics (2008)
8. Liu, W.: ShiXueHanYing (1735)
9. Manurung, R., Ritchie, G., Thompson, H.: Using genetic algorithms to create meaningful poetic text. Journal of Experimental & Theoretical Artificial Intelligence 24(1), 43–64 (2012)
10. Mikolov, T., Karafiát, M., Burget, L., Cernockỳ, J., Khudanpur, S.: Recurrent neural network based language model. In: INTERSPEECH 2010, 11th Annual Conference of the International Speech Communication Association, Makuhari, Chiba, Japan, September 26-30, 2010. pp. 1045–1048 (2010)
11. Netzer, Y., Gabay, D., Goldberg, Y., Elhadad, M.: Gaiku: Generating haiku with word associations norms. In: Proceedings of the Workshop on Computational Approaches to Linguistic Creativity. pp. 32–39. Association for Computational Linguistics (2009)
12. Oliveira, H.: Automatic generation of poetry: an overview. Universidade de Coimbra (2009)
13. Oliveira, H.G.: Poetryme: a versatile platform for poetry generation. In: Proceedings of the ECAI 2012 Workshop on Computational Creativity, Concept Invention, and General Intelligence (2012)
14. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: Bleu: a method for automatic evaluation of machine translation. In: Proceedings of the 40th annual meeting on association for computational linguistics. pp. 311–318. Association for Computational Linguistics (2002)
15. Tosa, N., Obara, H., Minoh, M.: Hitch haiku: An interactive supporting system for composing haiku poem. In: Entertainment Computing-ICEC 2008, pp. 209–216. Springer (2009)
16. Wang, L.: A Summary of Rhyming Constraints of Chinese Poems (Shi Ci Ge Lv Gai Yao). Beijing Press (2002)
17. Wang, Q., Luo, T., Wang, D., Xing, C.: Chinese song iambics generation with neural attention-based model. In: IJCAI'16 (2016)
18. Wu, X., Tosa, N., Nakatsu, R.: New hitch haiku: An interactive renku poem composition supporting tool applied for sightseeing navigation system. In: Entertainment Computing–ICEC 2009, pp. 191–196. Springer (2009)
19. Yan, R., Jiang, H., Lapata, M., Lin, S.D., Lv, X., Li, X.: I, poet: automatic chinese poetry composition through a generative summarization framework under constrained optimization. In: Proceedings of the Twenty-Third international joint conference on Artificial Intelligence. pp. 2197–2203. AAAI Press (2013)
20. Yoshikawa, K.: GEN MINSHI GAISETSU. Iwanami Shoten Publishers (1963)
21. Zeiler, M.D.: Adadelta: an adaptive learning rate method. arXiv preprint arXiv:1212.5701 (2012)
22. Zhang, X., Lapata, M.: Chinese poetry generation with recurrent neural networks. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 670–680 (2014)
23. Zhou, C.L., You, W., Ding, X.: Genetic algorithm and its implementation of automatic generation of chinese songci. Journal of Software 21(3), 427–437 (2010)