CSLT TECHNICAL REPORT-20160032 [Sunday 8th January, 2017]

Can Machine Generate Traditional Chinese Poetry? A Turing Test

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

*Correspondence:
wqx@cslt.riit.tsinghua.edu.cn
[1]Center for Speech and Language
Technology, Research Institute of
Information Technology, Tsinghua
University, ROOM 1-303, BLDG
FIT, 100084 Beijing, China
[3]CIST, Beijing University of Posts
and Telecommunications , Beijing,
China
Full list of author information is
available at the end of the article

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 Turing test.

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.

Keywords: RNN; Poem generation

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 structural, 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 and can be only performed by a very few knowledgeable people.

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 [1], 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. This is like dancing in fetters, hence very difficult and can be performed only by knowledged people with long-time training.

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 in art generation. 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 [2] and recurrent neural networks (RNN) [3]. 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 [4] 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 Turing 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, [5] and [6] employed a phrase search approach for Japanese poem generation, and [7] proposed an approach based on word association norms. [8] and [9] used semantic and grammar templates for Spanish poem generation.

The second approach is based on various genetic algorithms. For example, [10] proposed to use a stochastic search algorithm to obtain the best matched sentences. The search algorithm is based on four standards proposed by [11]: fluency, meaningful, poetic, and coherent.

The third approach to poem generation involves various statistical machine translation (SMT) methods. This approach was used by [12] to generate Chinese couplets, a special regulated verses with only two lines. [2] 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, [13] proposed a method that retrieves high-ranking candidates of sentences from a large poem corpus, and then re-arranges the candidates to generate rule-conformed new sentences.

More recently, deep learning methods gain much attention in poem generation. For example, [3] proposed an RNN-based approach that was reported to work well in quatrain generation [3]; however, the structure seems rather complicated (a CNN and two RNN components in total), preventing it from extending to other genres.

Our approach follows the RNN-based approach but introduces several innovations. Firstly, we use the attention-based framework to enforce theme consistency and trigger innovation; Secondly, we use the GRU rather than the vanilla RNN to obtain long-distance memory; Thirdly, our model is a simple sequence-to-sequence structure, which is much simpler than the model proposed by [3] and can be easily extended to more complex genres.
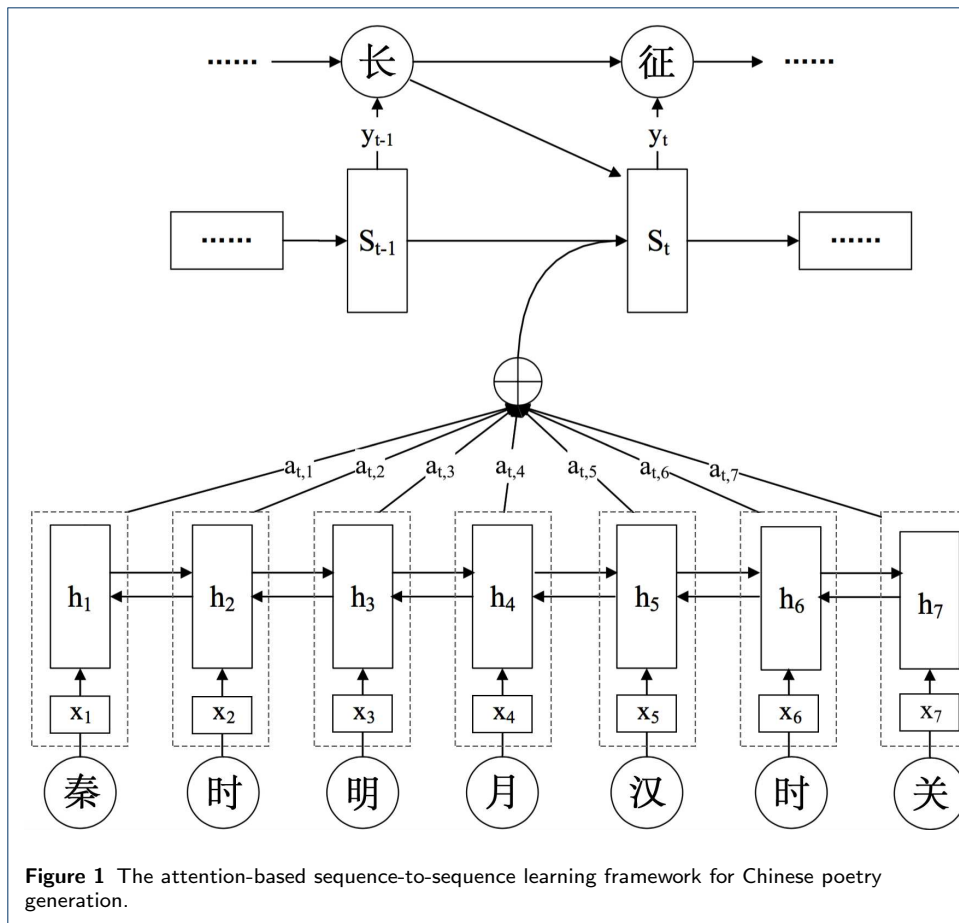
3 Method

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 [14] 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.

The entire framework of the attention-based model applied to Chinese poetry generation is illustrated in Figure 1. The encoder (a bi-directional GRU that will be discussed shortly) 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 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 relevance factor $\alpha_{t,i}$ that measures the similarity between s_t and h_i.

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

3.2 GRU-based Model Structure

A potential problem of the RNN-based generation approach proposed by [3] is that the vanilla RNN used in their model tend to forget historical input quickly, leading to theme shift in generation. To alleviate the problem, [3] designed a composition strategy that generates only one line at each time. This is certainly not satisfactory as it complicates the generation process.

In our model, the quick-forgetting problem is solved by using the GRU model. For encoder, a bi-direction GRU is used to encode the input keywords, and for the decoder, another GRU is used to conduct the generation. The GRU is powerful in remembering input and thus can provide a strong memory for the theme, especially when combined with the attention mechanism.

3.3 Model Training

The goal of the model 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 sentence by sentence, and the AdaDelta algorithm was used to adjust the learning rate [15]. Note that in the training phase, there are no input keywords, 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 simple to train the attention model from scratch, 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 attention model will focus on the transitional properties among characters only, 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 the attention model. Since part of the model (embedding matrix) have 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 leans 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 that are 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 n-best inference, this approach can generate unlimited alternatives without any quality sacrifice. Interestingly, our experiments show that

---

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

more keywords tend to generate more unexpected but highly impressive poems. Therefore, the multiple-attention approach can be regard 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 propose a hybrid-style training that trains the two types of quatrains using the same model, with a 'type indicator' to notify the model which type the present training sample is. In our study, the type indicators are derived from the eigen vectors of a 200×200 dimensional random matrix. Each type of quatrain is assigned a fixed eigenvector as its type indicator. The indicators are provided as part of the input to the first hidden state of the decoder, and keep constant during the training.

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 three 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 Turing test.

5.1 Datasets

Two datasets were 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. This corpus was used to train the GRU component of the attention model and the LSTM component of a comparative method based on RNN language models.

The second dataset was used to train and derive character vectors that were used 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 of data ensures a stable learning for semantic content of most characters.

5.2 Model Development

In the first evaluation, we intended to find the best configurations for the proposed attention-based model. The 'Bilingual Evaluation Understudy' (BLEU) [16] was used as the metric to determine which enhancement techniques are effective. BLEU was originally proposed to evaluate machine translation performance [16], and was used by [3] to evaluate quality of poem generation. We used BLEU as a cheap evaluation metric in the development phase to determine which design option to choose, without costly human evaluation.

The method proposed by [2] and employed by [3] 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 [3]. This is because our attention model generates poems as an entire character sequence, while the vanilla RNN approach in [3] does that sentence by sentence. Additionally, we used 1-grams 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 was trained with character initialization where the character vectors were trained using quatrains only. Then we used the large corpus that involves all traditional Chinese poems to enhance the character vectors, and the results (2nd row in Table 2) demonstrate a noticeable performance improvement. This is understandable since poems in different genres use similar languages, so involving more training data may help 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 was 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 stopped the training before convergence was reached in favor of a good BLEU.

From these results, we obtain the best configuration that involves character vectors trained with extern training data, input reconstruction, input vector attention and hybrid training. In the rest of the paper, the attention model with this configuration will be denoted by 'Attention'.

5.3 Comparative Evaluation

In the second phase, we compared the attention model (with the best configuration) and three comparative models: the SMT model proposed by [2], the vanilla RNN poem generaion (RNNPG) proposed by [3], and an RNN language model (RNNLM) that can be regarded as a simplified version of the attention model [17].

Following the work of [3], we selected 30 subjects (e.g., falling flower, stone bridge, etc.) in the Shixuehanying taxonomy [18] 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 provide by [3] to generate the first lines for the SMT, vanilla RNN and RNNLM approaches. A 5-char quatrain and a 7-char quatrain were generated for each theme by the four methods, and were evaluated by experts.

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
RNNLM	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.

For reference, some poems written by ancient poets were also involved in the evaluation. Note that to prevent the impact of prior knowledge of the experts, we deliberately chose the poems that were written by poets that are not very famous. The poems were chosen from [19], [20] and [21]; 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 the three comparative approaches: SMT, RNNLM, and RNNPG. The SMT-based approach is available online[3] and we used this online service to obtain the generation. For RNNPG, we invited the authors to conduct the generation for us. The RNNLM 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.

Poems written by ancient poets were 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 in the 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 RNNLM approach. As we mentioned, RNNLM 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 with character vector initialization. This demonstrated that a simple neural model with little engineering effort may learn artistic activities pretty good. Nevertheless, the

---

[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/

comparative advantage of the attention model still demonstrated the importance of the attention mechanism.

The RNNPG and the SMT approaches perform equally worse, particularly RN-NPG. 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 [3]. 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 quality as our model does.

Additionally, we note that almost in all the evaluations, the human-written poems beat those generated by machines. On one hand, this indicates that human are still superior in artistic activities, and on the other hand, it demonstrates from another perspective that the participants of the evaluation are truly professional and can tell good and bad poems. 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 from two aspects: 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 unfamous 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.

5.4 Turing Test

We designed a Turing test to evaluate the quality of poems generated by our model and by contemporary poets. 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 favored so that the quality can be ensured.
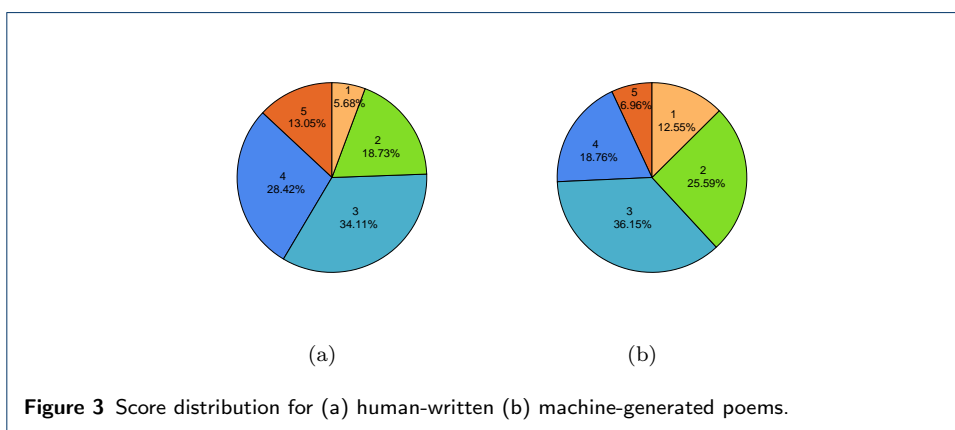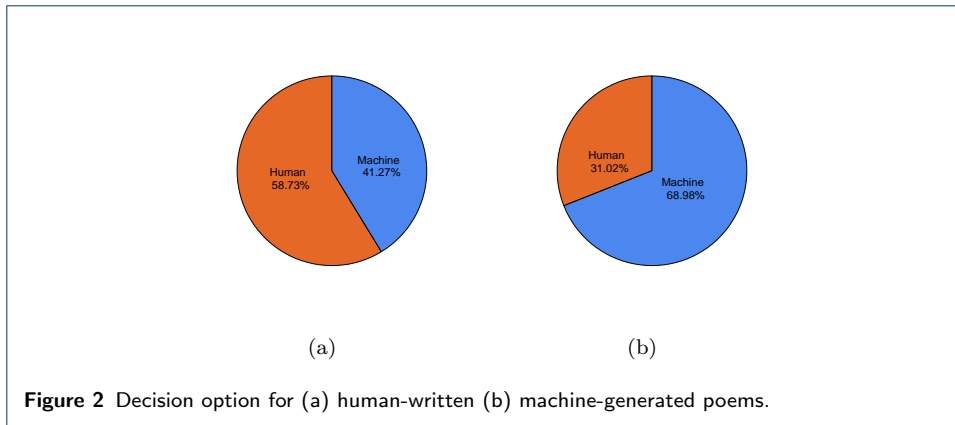
We received 144 submissions. To ensure the quality of the submission, we generated the same number of poems by our model and then asked a highly-professional expert in traditional Chinese poem criticism to give a pre-filtering. After the filtering, 83 human-written poems (57.6%) and 180 computer-generated 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) what is the quality of the poem, rated from 0 to 5.

---

[4]The author of RNNPG [3] could not find the SMT model in the reproduction, unfortunately.
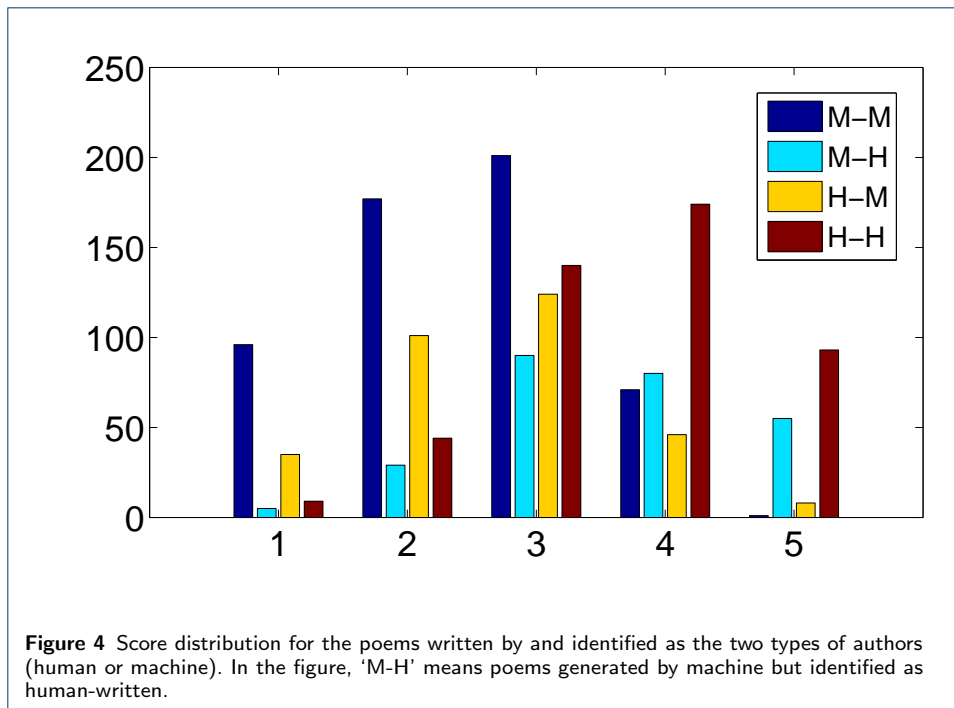[5]These experts were nominated by professors in the field of traditional Chinese poetry research.
[6]These experts again are mostly from CASS, and some of them attended the previous test but not all.

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

Figure 3 Score distribution for (a) human-written (b) machine-generated poems.

The results for the human-machine decision are presented in Figure 2. For a clear presentation, 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, our model has weakly passed[7]. The score distributions for human-written and machine-generated poems are presented in Figure 3. It can be seen that our model is still inferior to human in average. However, a large proportion (61.9%) of the machine-generated poems were scored equal to or more than 3, which means that our model works pretty well, as human poets can only achieves 75.6%. Interestingly, among the top-5 high-ranked poems, the machine takes the position 1 and 2, and among the top-10 high-ranked poems, the machine takes the position 1, 2 and 7. This means that our model can generate very good poems, even better than human poets, although in general it is still beat by human.

A more detailed analysis is presented in Figure 4, where the poems are categorized into four groups according to their 'genuine' and 'identified' authors (human or machine). From the two pairs M-M vs. M-H and H-M vs. H-H, we observe that a poem tends to be rated high if the experts consider them as Human-written. This

---

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

Figure 4 Score distribution for the poems written by and identified as the two types of authors (human or machine). In the figure, 'M-H' means poems generated by machine but identified as human-written.

means that the identification is positively related to the score, and people still tend to recognize human writes better. This is also true anyway at present.

To have a better understanding of the decision process, we invited another 4 experts to specify the metrics by which the human-machine identification was made for each poem. Multiple metrics can be selected. The proportions that each metric was selected are shown in Table 4. It can seen that the experts tend to regard fluency and aesthesis as the most important factors in the decision. When evaluating human-written poems, it shows that a fluent poem tends to be identified correctly, while a poem without any aesthetic feeling tends to be recognized as machine-generated.

	Comp.	Flu.	Cons.	Aes.
M-M	20.8%	**68.0%**	51.2%	60.0%
M-H	22.9%	51.4%	**54.3%**	51.4%
H-M	32.9%	51.8%	45.9%	**76.5%**
H-H	10.7%	**72.0%**	58.7%	61.3%
Overall	21.9%	62.8%	51.9%	**63.8%**

Table 4 Percentage of each metric was chosen in the identification decision. 'M-H' means the category that machine-generated poems are identified as human-written.

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'.

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

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 5 A quatrain example generated by the attention model.

to other genres. Our experiments show that it can generate traditional Chinese quatrains pretty well and weakly pass the Turing test. 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.

Author details
[1]Center for Speech and Language Technology, Research Institute of Information Technology, Tsinghua University, ROOM 1-303, BLDG FIT, 100084 Beijing, China. [2]Center for Speech and Language Technologies, Division of Technical Innovation and Development, Tsinghua National Laboratory for Information Science and Technology, ROOM 1-303, BLDG FIT, 100084 Beijing, China. [3]CIST, Beijing University of Posts and Telecommunications , Beijing, China. [4]Huilan LTD , Beijing, China.

References
1. Li Wang, *A Summary of Rhyming Constraints of Chinese Poems (Shi Ci Ge Lv Gai Yao)*, Beijing Press, 2002.
2. Jing He, Ming Zhou, and Long Jiang, "Generating chinese classical poems with statistical machine translation models," in *Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.
3. 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)*, 2014, pp. 670–680.
4. Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio, "Learning phrase representations using rnn encoder-decoder for statistical machine translation," *arXiv preprint arXiv:1406.1078*, 2014.
5. Naoko Tosa, Hideto Obara, and Michihiko Minoh, "Hitch haiku: An interactive supporting system for composing haiku poem," in *Entertainment Computing-ICEC 2008*, pp. 209–216. Springer, 2009.
6. Xiaofeng Wu, Naoko Tosa, and Ryohei Nakatsu, "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.
7. Yael Netzer, David Gabay, Yoav Goldberg, and Michael Elhadad, "Gaiku: Generating haiku with word associations norms," in *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*. Association for Computational Linguistics, 2009, pp. 32–39.
8. H Oliveira, "Automatic generation of poetry: an overview," *Universidade de Coimbra*, 2009.
9. Hugo Gonçalo Oliveira, "Poetryme: a versatile platform for poetry generation," in *Proceedings of the ECAI 2012 Workshop on Computational Creativity, Concept Invention, and General Intelligence*, 2012.
10. Cheng-Le Zhou, Wei You, and Xiaojun Ding, "Genetic algorithm and its implementation of automatic generation of chinese songci," *Journal of Software*, vol. 21, no. 3, pp. 427–437, 2010.
11. Ruli Manurung, Graeme Ritchie, and Henry Thompson, "Using genetic algorithms to create meaningful poetic text," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 24, no. 1, pp. 43–64, 2012.
12. 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*. Association for Computational Linguistics, 2008, pp. 377–384.
13. 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 *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*. AAAI Press, 2013, pp. 2197–2203.
14. Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio, "Neural machine translation by jointly learning to align and translate," *arXiv preprint arXiv:1409.0473*, 2014.
15. Matthew D Zeiler, "Adadelta: an adaptive learning rate method," *arXiv preprint arXiv:1212.5701*, 2012.
16. Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu, "Bleu: a method for automatic evaluation of machine translation," in *Proceedings of the 40th annual meeting on association for computational linguistics*. Association for Computational Linguistics, 2002, pp. 311–318.
17. Tomas Mikolov, Martin Karafiát, Lukas Burget, Jan Cernockỳ, and Sanjeev Khudanpur, "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*, 2010, pp. 1045–1048.
18. Wenwei Liu, *ShiXueHanYing*, 1735.
19. Wei Han, *Study on the Liao, Jin and Yuan Poetry and Music Theory*, China Social Sciences Press, 2015.
20. Kojiro Yoshikawa, *GEN MINSHI GAISETSU*, Iwanami Shoten Publishers, 1963.
21. Yan Chen, *Jin Poetry Chronicle*, Shanghai Chinese Classics Publishing House, 2013.