The Machine Poetry Generator Imitating Du Fu's Styles

Kai Wang
Institute for DeepCom Research
Beijing, China
e-mail: KaiWang@dichang.com

Jacky Tian
Institute for DeepCom Research
Beijing, China
e-mail: jackypku2013@gmail.com

Ron Gao
Institute for DeepCom Research
Beijing, China
e-mail: rongao1986@gmail.com

Chang Yao
Institute for DeepCom Research
Beijing, China
e-mail: 1033643002@foxmail.com

***Abstract*—Du Fu is one of the greatest poets in Tang Dynasty. His poems are depressed and sonorous, look gloomy and contemplative, but show desires for better life and brighter future. His poems depict the era of 1000 years ago. To understand and inherit his poems better, we propose a novel machine learning approach with LSTM to generate new Chinese poetry imitating Du Fu's styles automatically. Given the first character, the model will generate the rest characters of a poem according to the writing style of the author. In contrast with the previous methods, besides following the formats, satisfying tonal and rhythm requirements and having semantic coherence, our method could inherit Du Fu's styles to some extent. We build a prototype with Du Fu's poems and the evaluations have illustrated the effectiveness of our approach.**

Keywords-poetry; LSTM; Du Fu; depressed and sonorous style

I. INTRODUCTION

The traditional Chinese poetry is considered as one of the most significant heritages of Chinese culture. Chinese poets have written millions of poems to express their feelings and ambitions since about two thousand years. Different types came into being with the development of the traditional Chinese poetry, such as Tang poetry, Song iambics and so on. Nowadays, traditional Chinese poetry still has influence on many aspects of daily life.

Whereas, traditional Chinese poems are quite difficult to write, as each type of poetry has to follow some rules and only few people have a command of them. Take the rules of quatrain as the example. A quatrain consists of four lines that are usually made up with five or seven characters. These characters have particular tones, Ping or Ze. Ping is the level tone and Ze is the downward tone. The last characters of the second and the last line have to belong to the same rhyme category. Chinese poets wrote beautiful poems with the strict principles, among which Du Fu is one of the greatest ones. On account of a disordered society, Du Fu's poems are often depressed and sonorous. Table 1 is a quatrain of Du Fu [1], which can clearly show the feature.

TABLE I. HAPPY RAIN ON A SPRING NIGHT

春夜喜雨 Happy Rain on a Spring Night
好雨知时节, It rains at an appropriate time;
当春乃发生 。 Everything comes back to life.
随风潜入夜， There are breezes in night;
润物细无声。It moistens everything mutely.
野径云俱黑, Small lanes are dark;
江船火独明。In boat a lantern looms.
晓看红湿处， Saturated reds appear in the dawn;
花重锦官城。The town is filled by blooms.

In recent years, more and more people focus on automatic poetry generation. Rule-based templates, genetic algorithms, statistical machine translation and summarization method are widely used to generate poems.

Along with the rapid growth of machine learning techniques [16-18], the Chatbot XiaoIce produced by Microsoft could be able to generate modern poetry with RNN. There are also researches of generating traditional Chinese poems. However, generating poems focusing on some specific authors haven't been taken yet. Using previous methods, we have chances to understand the styles of each author in a deeper level.

In this paper, we propose a method that can generate poems with specific features. Besides, we apply LSTM to the poem-generation model.

Having collected poems of a poet, such as Du Fu, the poem-generation model will perform statistics analysis, learning the probabilities of each character and the character combination. When users put down one certain character after the model's training, the poem-generation model will generate a poem of two lines consisting of five to seven characters, using LSTM. The generated poem has the similar features with what Du Fu writes.

The contributions of this paper are as follows: First, we collect Du Fu's poems and do some data preparation including pre-process and word embedding. Second, we

978-1-5386-6987-7/18/$31.00 ©2018 IEEE

apply the LSTM to poetry generation, improving some defects discovered in the previous method. Thirdly, we find that poems generated by our approach could have similar features with Du Fu's poems.

The rests of the paper are organized as follows. Some related works are listed in the Section 2. Section 3 describes the approach adapted in our experiments. And we illustrate the evaluation in Section 4. The conclusion of this paper and future work are demonstrated in Section 5.

II. RELATED WORKS

Since the poetry generation became a popular topic, researches from various countries have carried out several experiments. Greene et al. [2] use statistical Machine Translation to generate cross-lingual poems. Manunung et al. [3] propose Genetic Algorithm to create poetic text in English. Oliveira et al. [4] uses rule-based Templates to generate poems. Besides, Haiku, a system based on rule-based phrase search, is created to generate Japanese poetry [5-6]. Netzer et al. [7] propose another way of Haiku with World Association Norms. Agirrezabal et al. [8] use patterns that are based on parts of speech and Word Net to compose Basque poems.

In addition of English, Japanese, Portuguese and Basque poetry, there are also methods used to create Chinese poetry. A system called Daoxiang is produced to generate Chinese poetry. Jiang and Zhou [9] propose a LSTM model to generate the 2nd sentence of a couplet when given the 1st sentence. A general algorithm is used by Zhou [10] to generate Chinese poetry. Then a LSTM model is extended by He [11], [14-15], which can generate poems made up with four lines.

III. APPROACHES

A. Overview

Based on the observation that a human poet usually writes the first line of a poem, we propose a poetry generation approach that generates the later character on the basis of the formers characters.

We first divide the poems into N lines ($N_1, N_2 ..., N_N$) of which the each character is regard as a term (($t_1, t_2, \dots, t_N$). Using LSTM, the generator will calculate the probability of each term. Given a character belonging to terms, the generator generates the second term, selecting a term with the largest probability from several terms that meet the demand randomly. In accordance with the first and the second term, the third term is generated same as before. The generator eventually generates two lines made up with five to seven terms.

B. Experiment Workflow

The whole experiment workflow is shown in Figure 1. At the beginning of the experiment, we collect Du Fu's poems, which are split into terms and transformed into word embeddings later. Then the model starts to train to get the parameters of the complex neural network. Given the first character, the model will generate the next characters on the basis of the former characters until encountering the end character.

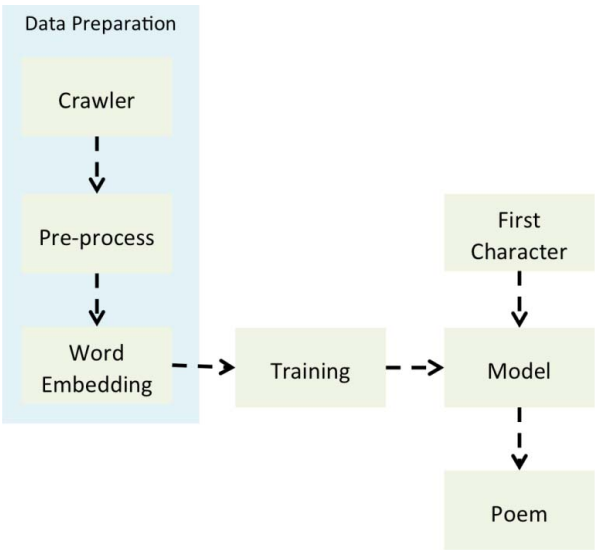

Figure 1. The workflow of the whole experiment.

C. Dataset

Du Fu is one of the greatest Chinese realistic poets in Tang Dynasty. He was born in a rich family but his life suffered from chaos and homeless caused by war. Therefore, his poems could be divided into two styles: romantic and realistic.

In this experiment, we first collect 1,528 poems written by Du Fu from [12]. Each sentence of his poem consists of five to seven characters or was filled with some specific pattern according to rhythm. Based on the characteristics, we divide these poems into single sentences. The length of each sentence is less than 20 characters. Finally, we get 9773 sentences in total, and each of them ends up with a period, a question mark or an exclamation mark. In order to normalize these sentences, we assign each sentence with a beginning character "B" and an ending character "E".

To make the corpus meet the demand of the neural network, we expect to transform each sentence into a vector. In order to complete the task of word embedding, more steps are taken as follows: First, we calculate the frequency of each character. Second, we assign each character with an integer by the ordering of the frequency. Finally, the corpus is transformed into combinations of vectors. Table 2 shows examples of word embedding.

TABLE II. EXAMPLES OF WORD EMBEDDING

Name	Content
Chinese	B 国 破 山 河 在 , 城 春 草 木 深 。 E
Embedding	2,152,504,12,215,50,0,58,28,101,259,91,4,1
English	The city turns spring again, but the country is gone.

Name	Content
Chinese	B 国 破 山 河 在 , 城 春 草 木 深 。E
Chinese	B 感 时 花 溅 泪 , 恨 别 鸟 惊 心 。E
Embedding	2,608,17,41,2436,238,0,379,181,124,346,71,4,1
English	Lamenting the demise of the country, even the flowers and birds alarm my heart.
Chinese	B 烽 火 连 三月 , 家 书 抵 万 金 。E
Embedding	2,1194,522,368,96,61,0,79,73,1798,49,108 4,1
English	The war stretches about three months, and a letter from home is worth thousands of gold.
Chinese	B 白 头 搔 更 短 , 浑 欲 不胜 簪 。E
Embedding	2,15,11,2147,112,475,0,1158,107,5,677,1408,4,1
English	Due to my scratching, my white hair becomes shorter and is not long enough to hold a hairpin.

D. Training

In this experiment, we use the Long Short-Term Memory (LSTM), a special kind of Recurrent Neural Networks (RNN). RNN is a network with loops, allowing information to persist. In Figure 2, the chunk of neural network is reading some inputs x and outputs a value h. When we unfold the chunk, we can see that RNN has relevance to sequences and lists in essence.

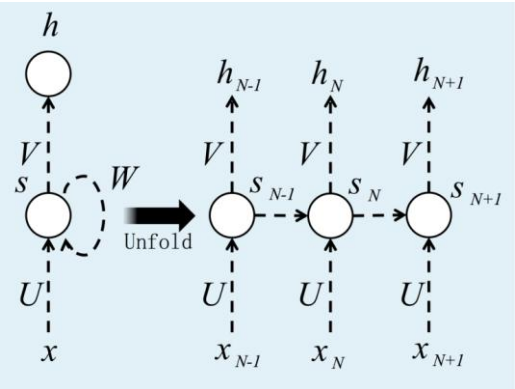

Figure 2. The Working Principles of RNN.

In the past few years, RNN has made great contributions to a variety of problems in the area of speech recognition, language modeling, machine translation and so on. The key to the success is LSTM, a special kind of RNN.

Compared with RNN, LSTM is able to avoid the problem of long-term dependencies, for the repeating module in LSTM has a different structure. There are four neural network layers instead of a single one, which can interact in a special way. In Figure 3, an entire vector is carried by a line, from one node's output to others' inputs. The yellow points denote pointwise operations, and the blue boxes are learned neural network layers. Lines merging represent concatenation, while lines forking represent the content of vector being copied.

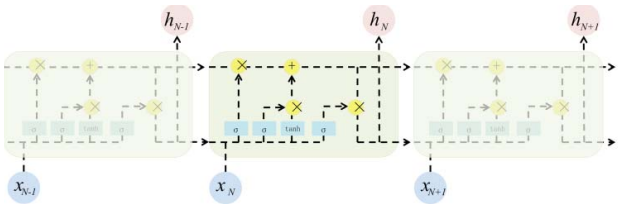

Figure 3. The Working Principles of LSTM.

We apply Stochastic Gradient Descent (SGD) in order to minimize the loss. The loss is computed as follow:

$$L(\theta) = \frac{1}{N}\sum_{n=1}^{N}(h_N - Y_N)^2 + \lambda \parallel \theta \parallel^2 \qquad (1)$$

In the function above, h_N represents predicted values and the real values are denoted by Y_N. θ denotes the complexity of the model. The process of training is described as follows: We first import data into our model. Second, the data is transformed into the format that can be recognized by the LSTM model whose parameters are set in the next step. Then, we begin to train the model. Table 3 demonstrates the process and the pseudo codes used in the implementation.

TABLE III. STEPS OF TRAINING

Algorithm 1

```
#Step1 Data Input
#Step2 Format Conversion
input_data=tf.placeholder (tf.int32, [FLAGS.batch_size, None])
output_targets=tf.placeholder (tf.int32, [FLAGS.batch_size, None])
#Step3 Parameter Setup
end_points = rnn_model(model='lstm', input_data=input_data,
output_data=output_targets, vocab_size=len(
  vocabularies), rnn_size=128, num_layers=2, batch_size=64,
learning_rate=FLAGS.learning_rate)
#Step4 Training
 for batch in range(n_chunk):
   loss, _, _ = sess.run([
     end_points['total_loss'],
     end_points['last_state'],
     end_points['train_op']
   ], feed_dict={input_data: batches_inputs[n], output_targets: batches
outputs[n]})
   n += 1
```

In the implementation, we set the RNN with 128 neurons and 2 neural network layers and keep the learning rate default. The batch size is 64. The trend of loss is illustrated in Section 4.

IV. EVALUATION

A. Statistics of Du Fu Poems

In Du Fu's poems, 4,211 Chinese characters are used out of 5,000 frequent Chinese characters. That's one of the reasons why Du Fu is called "The Poem Saint".

In order to understand the major intent of Du Fu, we selected the most frequent Chinese characters in Du Fu's poems and removed some prepositions, auxiliary words and characters having multiple meanings. The results are illustrated in the Table 4. What we can find are as follows:

- In Chinese, “不” and “无” means negation and “老” means old age. These Chinese characters suggest the frustration of Du Fu. Du Fu lived in a chaos society, hence he was frustrated with officialdom and future.
- In general, Chinese poets often create sorrowful artistic conception through the Chinese character “秋” which actually means autumn. The same is true for Du Fu. He also used “寒” with the meaning of chill to express his sorrow and loneliness.

TABLE IV. TOP FREQUENT CHINESE CHARACTERS IN DU FU’S POEMS

Character	Frequency	Character	Frequency	Character	Frequency
不 (Not)	965	无 (No)	578	水 (Water)	391
人 (Person)	785	山 (Hill)	551	老 (Old)	385
日 (Sun)	604	天 (Sky)	516	秋 (Fall)	321
江 (River)	580	云 (Cloud)	502	客 (Guest)	314
风 (Wind)	579	生 (Occur)	421	寒 (Chill)	255

Besides, we make a cloud tag composed of top frequent Chinese characters used by Du Fu. Figure 4 illustrates the cloud tag [13].

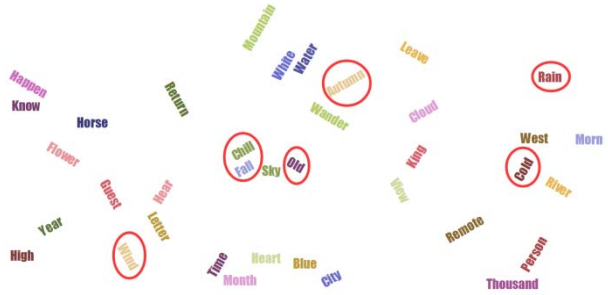

Figure 4. What's in Du Fu's mind.

B. The Trend of Losss

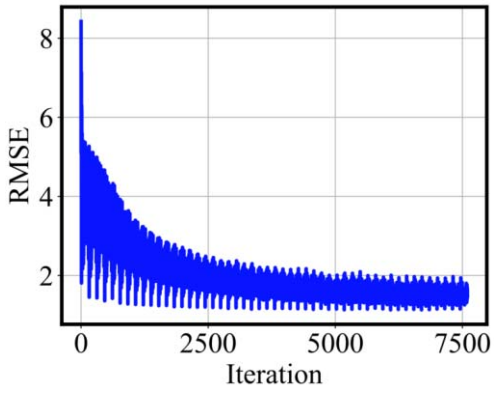

Figure 5. The trend of loss.

Based on SGD, we compute the loss each time we train the model and draw an image of the losses. As is shown in Figure 5, there is a downward trend of losses with times of training increasing. Eventually, the loss remains as low as about 1.4.

C. Poems Generated By LSTM

After training, we apply the LSTM model every time we provide the first Chinese character. The illustration of results is in Table 5. In Table 5, we translate the main idea of each poem into English.

It’s visible that poems generated by the LSTM model have some merits. First, the poems are composed of images used frequently by Du Fu. These images make great contributions to performing Du Fu’s style and expressing a certain emotion. Second, ensuring the semantic coherence, some of the poems are able to rhyme. Then, the generator could generate poems with length of five to seven characters, which have the similar structure with the data we provide. Besides, some of the poems are capable of creating mental pictures.

TABLE V. POEMS GENERATED BY LSTM MODEL

Name	Content
Chinese	天子径意朝
English	The emperor head for palace to handle governmental affairs
Chinese	堂来无人归
English	When I back home, none of my family is at home.
Chinese	江坐日雨雪
English	I sit in the river edge accompanied by rain and snow.
Chinese	满日萧风惊
English	The wind is blowing and I am shocked by its chill.
Chinese	断楹雨与山荒
English	The pillars that support the ceiling are broken on a rainy day. It seems that mountains in the distance become desolate on account of my broken house.

V. DISCUSSION

Whereas, there are some defects in the experiment. Table 6 suggests several examples of poems with defects. The first defect is the chaos of the part of speech (Poem 1). Poem 1 initiates with nouns “日”(sun) and “宅”(house) that are followed by two adjectives “明”(light) and “故”(old), but ends with a adverb “莫” (not) .When people read these poems, they find it almost impossibly to understand what’s the sentence means. Logical error (Poem 2) is the second problem. Generally, the character “百” (hundred) is combined with the character “千” （thousand） to express a large number of things, like “thousands of” or “hundreds of”. But in the example, the generator combines “百” with the character “干”, which usually means dryness. The logical error makes the poem unreadable. Thirdly, the model is

likely to generate the beginning character “B” as a result of the miscalculations (Poem 3). In order to improve the defects, there are still a large number of works to do in the future. Besides, machine learning techniques rely on more training data colleting techniques, such as smartphones; and the collected data can be uploaded to the cloud or shared by device-to-device (D2D) technologies [19-22].

TABLE VI. EXAMPLES OF POEMS WITH DEFECTS

	Content
Poem 1	日宅明故莫
Poem 2	风幕干百
Poem 3	空 B 际一贤

Finally, we hope that some description of our work might be of relevance other generation tasks.

VI. CONCLUSION

It’s a hard task to generate poems, especially traditional Chinese poems with a specific style. We present a novel approach based on LSTM to generate Chinese poems with a specific style. Given the first character, the model will generate two lines of poems by calculating the probabilities. In contrast with previous methods, our approach is able to avoid generating unnatural poems caused by style confusion. According to the experimental results, we can see that our model can provide high quality poems. Some of the poems are even comparable with the great human poets.

REFERENCES

[1] Translation of Du Fu’s Poems [Online]. Available: http://m.kekenet.com/kouyi/201502/357358.shtml

[2] E. Greene, T. Bodrumlu and K. Knight, “Automatic analysis of rhythmic poetry with applications to generation and translation”, in EMNLP, 2010.

[3] R. Manurung, G. Ritchie and H. Thompson, “Using genetic algorithms to create meaningful poetic text. Journal of Experimental & Theoretical Artificial Intelligence”, 24(1):43–64, 2011.

[4] H. Oliveira, “Automatic generation of poetry: an overview”, Universidade de Coimbra, 2009.

[5] N. Tosa, H. Obara and M. Minoh, “Hitch haiku: An interactive supporting system for composing haiku poem”, Entertainment Computing ICEC 2008, pages 209–216, 2008.

[6] X.Wu, N.Tosa and R.Nakatsu, “Newhitch haiku: An interactive renku poem composition supporting tool applied for sightseeing navigation system”, Entertainment Computing–ICEC 2009, pages 191–196, 2009.

[7] Y. Netzer, D. Gabay, Y. Goldberg and M. Elhadad, “Gaiku: generating haiku with word associations norms”, in workshop of CALC, 2009.

[8] M. Agirrezabal, B. Arrieta, A. Astigarraga and M. Hulden, “POS-Tag Based Poetry Generation with WordNet”, in workshop of Natural Language Generation, 2013.

[9] L. Jiang and M. Zhou, “Generating Chinese Couplets using a Statistical MT Approach”, in International Conference on Computational Linguistics, 2008

[10] C. Zhou, W. You and X. Ding, “Genetic algorithm and its implementation of automatic generation of Chinese songci”, Journal of Software, 21(3):427–437, 2010.

[11] J. He, M. Zhou and L. Jiang, “Generating Chinese classical poems with statistical machine translation models”, in AAAI, 2012.

[12] Collection of Du Fu’s Poems [Online]. Available: http://m.shicimingju.com/chaxun/zuozhe/10.html

[13] Word Cloud Generator [Online]. Available: https://www.jasondavies.com/wordcloud/

[14] Z. Wang, W. He, H. Wu, H. Wu, W. Li, H. Wang and E. Chen, “Chinese Poetry Generation with Planning based Neural Network”, in COLING, 2016.

[15] M. Ghazvininejad, X. Shi, Y. Choi and K. Knight, “Generating Topical Poetry”, in EMNLP, 2016

[16] Y. Zhang, Z. Li, C. Gao, K. Bian, L. Song, S. Dong and X. Li, “Mobile Social Big Data: WeChat Moments Dataset”, Network Applications, and Opportunities. IEEE Network, 2018.

[17] Y. Zhang, Y. Guo, C. Gao, K. Bian, X. Jin, Z. Yang, L. Song, J. Cheng, H. Tuo and X. Li, “Proactive Video Push for Optimizing Bandwidth Consumption in Hybrid CDN-P2P VoD Systems”, in IEEE INFOCOM, 2018.

[18] Z. Li, L. Chen, Y. Bai, K. Bian and P. Zhou, “On Diffusion-restricted Social Network: A Measurement Study of WeChat Moments”, in IEEE ICC, 2016.

[19] Y. Yang, Z. Zheng, K. Bian, L. Song and Z. Han, “ Realtime Profiling of Fine-Grained Air Quality Index Distribution using UAV Sensing”, IEEE Internet of Things Journal, Vol. 5(1):186--198, Feb. 2018.

[20] C. Xu, L. Song, Z. Han, Q. Zhao, X. Wang and B. Jiao, “Efficiency Resource Allocation for Device-to-Device Underlay Communication Systems: A Reverse Iterative Combinatorial Auction Based Approach”, IEEE Journal on Selected Areas in Communications, special issue “on Peer-to-Peer Networks”, vol. 31, no. 9, pp. 348-358, Sep. 2013.

[21] L. Song, D. Niyato, Z. Han, and E. Hossain, “Game-theoretic Resource Allocation Methods for Device-to-Device (D2D) Communication”, IEEE Wireless Communications Magazine, vol. 21, no. 3, pp. 136-144, Jun. 2014

[22] Y. Liao, K. Bian, L. Song and Z. Han, “Full-Duplex MAC Protocol Design and Analysis,” IEEE Communications Letters, vol. 19, no. 7, pp. 1185-1188, Jul. 2015