

Generating Chinese Classical Poems Based on Images

Xiaoyu Wang, Xian Zhong, Lin Li

Abstract—With the development of the artificial intelligence technology, Chinese classical poems automatic generation has received a fair amount of focus in decades. In this paper, based on technology of natural language processing and image description, we present a Chinese classical poems automatically generative model, which can compose a piece of poetry related to the picture content. At the first stage, we use improved VGG16 model to predict the input image. The output of this part is forecast result in Chinese. Then, the model will generate a piece of poetry according to the prediction result based on RNN (Recurrent Neural Network). Specially, we use acrostic poems to make the generative classical poetry associated with the given picture.

Index Terms—Chinese poems generation, recurrent neural network, natural language processing, artificial intelligence

I. INTRODUCTION

In modern society, Chinese culture is favored by people all around the world. Undeniably, Chinese classical poetry is one of the unique cultural heritage of China. Classical poems have a long history with more than two thousand years. It manifests itself in many aspects of people's life, for example, as a method of recording important events, expressing personal emotions, or communicating messages at special festivals. For two thousand years, the classical poems are brilliant stars in the human civilization.

With the rapid development of technology, Chinese classical poetry automatic generation has received a fair amount of focus in decades, with lots of computational systems written to generate poetry online. Meanwhile, with the boom of artificial intelligence, researches about image description with natural text also have made remarkable progress. It has been achieved that machine can read images and make description about the image's contents. But in the traditional culture of China, people prefer to use classical

poems to express their feelings, which is called 'Lyric Expression Through Scenery'. Thus, in this paper, we propose a novel approach for Chinese classical poems generation given an image.

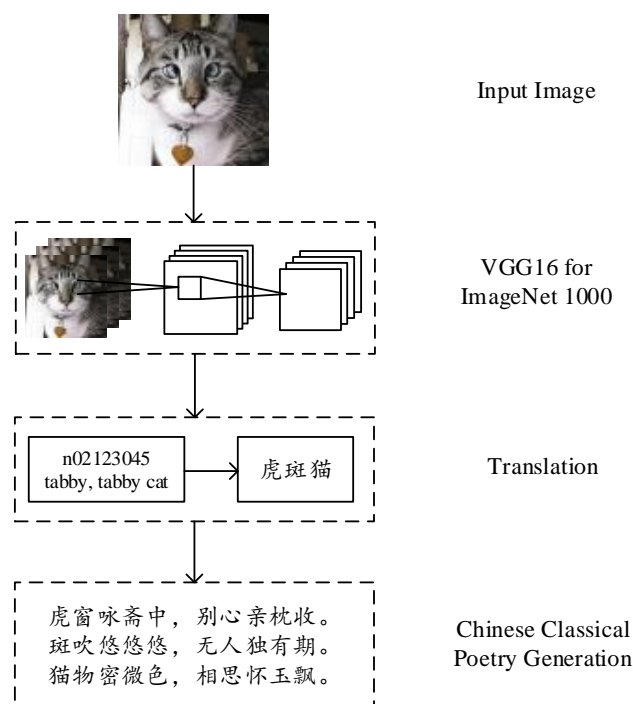


Fig. 1. We propose an approach that can be used to generate Chinese classical poetry given an image. In the first part, we use VGG16 for ImageNet 1000 to produce predicted result about the input image. Secondly, we need to translate ImageNet 1000 data set into Chinese both automatically and manually. In the last part, we use the predicted result as the input to poetry generation model.

As shown in Fig. 1, to fulfil this system, we use VGG16 model for ImageNet 1000 to predict the category about the input image, and use the result as input to poetry generation model. However, when it comes to Chinese, there is only a little Chinese image annotation data set. Thus, before all of this, we translate the label of ImageNet 1000 into Chinese.

The contributions of our work are as follows:

- 1) For the first time, we present a Chinese classical poetry generation approach combining computer vision and machine translation and that can be used to generate Chinese classical poetry given an image.
- 2) To generate Chinese poetry, we translate the label of

Manuscript received November 28, 2017; revised January 15, 2018. This work was supported by the Fundamental Research Funds for the Central Universities, the National Natural Science Foundation of China (NSFC Grant Number 61003130), as well as the Natural Science Foundation of Hubei Province (Grant Number 2015CFB525).

Xiaoyu Wang is with School of Computer Science and Technology, Wuhan University of Technology, Wuhan. (e-mail: xiaoyuwang@whut.edu.cn).

Xian Zhong is with School of Computer Science and Technology, Wuhan University of Technology, Wuhan (e-mail: zhongx@whut.edu.cn).

Lin Li is with School of Computer Science and Technology, Wuhan University of Technology, Wuhan. (e-mail: cathylilin@whut.edu.cn).

ImageNet 1000 into Chinese using Youdao online dictionary. Also, we review the results manually.

- 3) Acrostic poetry is a familiar form of Chinese poetry. In an acrostic poem, make the first word of each line together, and you will get the author's unique ideas. In this paper, we use acrostic poems to make the generative classical poetry associated with the given picture.

II. RELATED WORK

The research about poetry generation started in 1960s, and becomes a hotspot in recent decades. The early methods are based on rules and templates. The system named Daoxiang [1] basically depends on manual pattern selection. The system contains a list of manually created terms associated with pre-defined keywords, and randomly inserts terms into the selected template as a poem. Daoxiang system is simple, and random term option usually results in unnatural sentences. Also, there have been some other poetry automatic generation researches via statistic machine translation. L. Jiang and M. Zhou [2] propose a phrase-based SMT approach to generate Chinese couplets, which can be seen as two lines poems. Based on this algorithm, J. He et al. [3] sequentially translate the current line from the previous line.

With the cross field of Deep Learning and Natural Language Process being focused, Neural Network has been applied on poetry generation. X. Yi et al. [4] take the generation of Chinese classical poems as a sequence-to-sequence learning problem. Based on the RNN Encoder-Decoder structure, they build a novel system to generate Chinese poetry (more specifically, quatrains), with a topic word as input. X. Zhang [5] presents a RNN model for Chinese poem generation based on recurrent neural networks. The model is ideally applied to capturing poetic contents and form. Given the user writing intents as queries, R. Yan et al. [6] utilize the poetry corpus to generate quatrains (Jueju in Chinese), and formally formulate the summarization process based on iterative term substitution. Later, R. Yan [7] proposes a new generative model with a polishing schema, and outputs a refined poem composition. Recently, M. Ghazvininejad [8] gives an automatic poetry generation system called Hafez. Given an arbitrary topic, Hafez will show you a piece of modern poetry in English. The system integrates a Recurrent Neural Network (RNN) with a Finite State Acceptor (FSA). By means of adjusting various style configurations, Hafez enables users to revise and polish generated poems if you are not satisfied with the result.

Neural network is commonly used in the field of languages and images. Being able to automatically describe the contents of an image is also a very challenging task in artificial intelligence. In recent years, researches about image description with natural texts have made remarkable progress. Given an image, machine can describe the contents of the picture using English words or sentences. The first method to using neural networks for caption generation was proposed by R. Kiros et al. [9]. They present a multimodal log-bilinear model that was biased by features from the image. Based on previous work, R. Kiros et al. [10] continue to study. Their method is designed to build a joint multimodal embedding space via a computer vision model and an LSTM (Long Short-Term Memory) that encodes text. In 2014, combining recent advances in computer vision and machine translation, O. Vinyals et al. [11] present a generative model based on a

deep recurrent architecture that can be used to generate natural sentences describing an image. The model is trained to maximize the likelihood of the target description sentence given the training image.

Inspired by the work in machine translation and object detection, K. Xu et al. [12] present an attention based model that can automatically learn to describe the contents of images. Using standard backpropagation techniques, they train the model in a deterministic manner by maximizing a variational lower bound. K. Simonyan and A. Zisserman [13] use a very small (3×3) convolution filter architecture to evaluate the depth of the network in a comprehensive manner, indicating that significant improvements in the prior art configuration can be achieved by pushing the depth up to 16-19 layers. Combining computer vision and natural language processing, Xiaobing [14], developed by Microsoft, can compose a modern Chinese poem given a picture. The system has learned 519 poet's modern poetry since 1920 and been trained more than 10000 times.

III. OVERVIEW

Recent advances in computer vision and natural language processing make artificial intelligence closer to people's daily life. In this paper, we propose a novel generative approach that can be used to generate Chinese classical poetry given an image.

As shown in Fig. 2, our system consists of three parts. In the first part, we use VGG16 model for ImageNet 1000 to predict the category about the input image. VGG16 model has been trained well and published on the internet. And its training results are recognized by the industry. People can download the model easily from the Internet. Given an image, the model will make a prediction about the contents. Undoubtedly, the predicted result is in English. Because of the lack of Chinese database, we need to translate the label of ImageNet 1000 into Chinese, only one output is reserved. Then, we split the result into individual characters, and each character will become the beginning of every line in the generative poetry. In the end, we use Recurrent Neural Network (RNN) to generate Chinese classical poetry based on the keyword. Correspondingly, our work mainly includes three parts: image classification, database translation, and poetry generation.

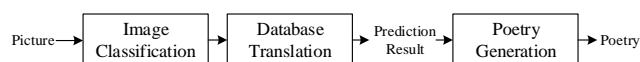


Fig. 2. we propose a novel generative approach that can be used to generate Chinese classical poetry given an image. The system mainly includes three parts: image classification, database translation, and poetry generation.

A. Image Classification

In this part, we use VGG16 [13] model to make a prediction about the image. VGG is a convolutional neural network model proposed by K. Simonyan and A. Zisserman from the University of Oxford in the paper "Very Deep Convolutional Networks for Large-Scale Image Recognition" [13]. The model achieves 92.7% top-5 test accuracy in ImageNet [15]. ImageNet is an image database organized according to the WordNet hierarchy (currently only the nouns), in which each node of the hierarchy is depicted by

hundreds and thousands of images. It is an easily accessible but authoritative image database for researchers around the world. It covers more than 14 million images belonging to 1000 classes.

B. ImageNet 1000 Database Translation

We utilize the online dictionary to complete the translation work. We choose Youdao Online Dictionary, and then crawl down the translation result through the program. There are 1000 categories. Considering accuracy, we select 5 students to review the translation results again and again. Each category reserves one translation result only.

C. Poetry Generation

The research about poetry generation becomes popular in recent decades. Machine learning methods based on statistics mainly contains two shortcomings:

- 1) Traditional machine learning methods are based on statistics, when the relationship between the data can't be described by statistics, the performance of traditional method will be poor.
- 2) Traditional machine learning methods often require expert knowledge to pick characteristics, which determines the outcome performance of learning.

Gradually, using neural network to solve the problem of poetry generation has shown a good effect. The most obvious advantages are as follows:

- 1) The length of poetry is limited, usually not too long. The neural network can easily remember the preceding word.
- 2) The format of classical poetry is fixed, and the location of punctuation is easily remembered by neural network.

IV. THE POEM GENERATOR

In this paper, we use VGG16 model to do predict with the input image. And the prediction result will be translated into Chinese. Finally, the model composes poems based on the keyword. Comparing with traditional method, RNN (Recurrent Neural Networks), particularly, the Encoder-Decoder structure shows a good character in sequence-to-sequence learning tasks. Thus, we use RNN to achieve the automatic generation of Chinese classical poetry. For the sake of making the generative poetry associated with the picture, we use Chinese results generated by VGG16 model as input to the poetry generation part.

A. Word Embedding

The input and output form of the neural network is a vector or matrix representation. For this reason, we need to build a vector representation of the poetry. Word embedding is a standard approach in text processing. The process, called vectorization, is to match a word to a low dimensional, real-valued vector. We select the 5382 common used words in the classical poetry. And each word is mapped to a numeric ID. For example, id '4' means character '不', id '0' stands for character '，'. By this way, we convert the poem into a vector form.

B. Start and Terminator

The length of Chinese classical poem lines is fixed. Usually, each sentence has five or seven characters. To output complete poetry in the neural network, rather than semi-

finished products, we have dealt with the input poems. Specifically, we add start character '[' at the beginning of each poem, and terminator ']' at the end of each poem.

C. Acrostic Poetry

The result, given by improved VGG16 model, will be the input of poetry generation model. Specifically, we define the following formulations:

- 1) Input. The Chinese result, given by improved VGG16 model, can be expressed as $R = \{x_1, x_2, x_3, \dots\}$, $x_i \in V$, where x_i is a character and V is the vocabulary.
- 2) Output. We generate a Chinese classical poem P according to R . We have $P = \{Y_1, Y_2, Y_3, \dots\}$. We have $Y_i = \{x_i, y_{i1}, y_{i2}, \dots, y_{ij}\}$, $y_{i,j} \in V$. Y_i stands for a line of poetry, including twelve or fourteen characters, especially containing two Chinese punctuation marks, '，' and '。'. To be more in detail, each character x_i in R will be the beginning of each line of poetry, which is Y_i . We predict the next character based on the previous one.

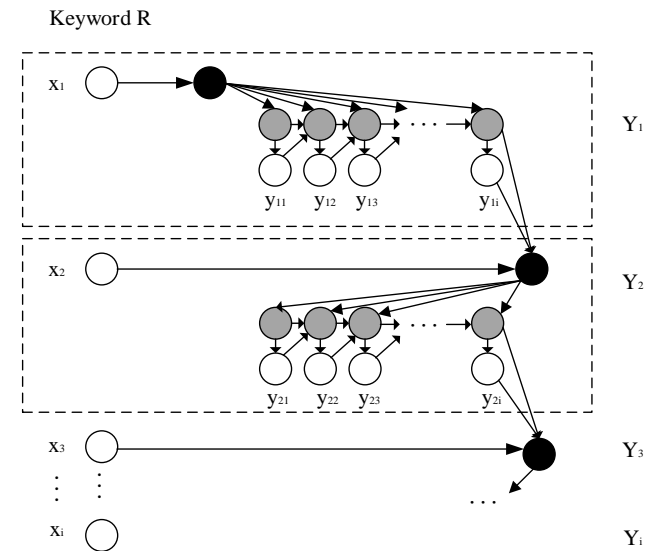


Fig. 3. When we get the Chinese predicted result from improved VGG16 model, we split the keyword into individual character, and each character will become the beginning of every line in the generative poetry. Every character will be influenced by previous characters. Every line is sensitive to all previously generated characters and currently input character. We use Recurrent Neural Network (RNN) to generate Chinese classical poetry given the keyword. Also, it should be noted that the number of character of every line is fixed, usually twelve or fourteen characters, including two Chinese punctuation marks, '，' and '。'.

As shown in Fig.3, for the first line Y_1 , we generate second character y_{11} based on x_1 . And every character will be influenced by previous characters. Later, every line is sensitive to all previously generated characters and currently input character. We compute the probability of line $Y_{i+1} = \{x_{i+1}, y_{i+1,1}, y_{i+1,2}, \dots, y_{i+1,j}\}$, given all previous lines $Y_{1:i} (i \geq 1)$. The equation is as follows:

$$P(Y_{i+1}|Y_{1:i}) = \prod_{n=1}^{j-1} P(y_{n+1}|y_{0:n}, Y_{1:i}) \quad (1)$$

As shown in equation (1), $P(Y_{i+1}|Y_{1:i})$ means the product of the probability of each character y_n in current line given all previous character $y_{0:n-1}$ and lines $Y_{1:i}$. We have $y_0 = x_i$.

V. EXPERIMENTS

A. Data

Our research contains two data sets. One is the image set, another is poems set. As for images, we use ImageNet 1000, which can recognize 1000 kinds of things. Besides, a large Chinese poetry corpus is important to learn the model for poems generation. There is some large Chinese poetry corpus available openly. Thus, we collect 17497 poems from Tang Dynasty to the contemporary, either the five words poem or seven words poem. We randomly choose 2,000 poetry for testing.

B. Training

As the model can divide into two parts, the training also includes two processes. For image recognition, we use VGG16 model, which is available publicly. By the way, once we get the training result, we will do check. As for poems generation, the model is trained with LSTM (Long Short-Term Memory). The aim for training is the cross entropy errors of distribution between the predicted character and the actual one in the corpus. We trained all sets of training parameters using stochastic gradient descent with specified learning rate.

C. Evaluation

Chinese ancient poetry not only pay attention to the structure neat, but more focus on rhythm beauty and artistic conception. without doubt, it is a much challenging task to make evaluation about machine-generated poems, let alone poems generated by the picture. We choose three different evaluation metrics to evaluate the quality of the results.

1) Perplexity

Since people put Perplexity is a sanity check in NLP (Natural Language Processing). In brief, perplexity is the probability. It means average branch factor, which can reflect how many choices we have when predict the next word. In fact, perplexity is an evaluation based on entropy. An ordinary form of perplexity is as follows:

$$P(S) = 2^{-\frac{1}{N} \sum \log(P(w_i))} \quad (2)$$

In equation (2), N stands for the length of sentence S , $P(w_i)$ means the probability of the word w_i . Intuitively, the lower perplexity for poems generated, the better performance for the results, and accordingly, the poems are likely to be good.

2) Human Evaluation

Since people pay more attention to the aesthetic of poetry, it is necessary to do human judgments. We invited 10 graduate students who are majoring in Chinese literature to Do some evaluation with the results. Referring to evaluation standards discussed in [6] [17], we design three criteria: ‘syntactic’, ‘semantic’ and ‘correlation’ satisfaction. Syntactic shows the neatness of the sentence structure. It can determine whether the poetry is well-formed. For a higher level semantic side, it reflects whether the poem is meaningful. Meanwell, evaluators should consider if the poem convey the input image messages. To make the evaluation process easier, each criterion is scored 0-1(‘0’-no, ‘1’-yes).

D. Performance

The research focuses on Chinese classical poems automatic generation given an image. To judge the performance of the poem, we compared our system, PG-image, with SMT, He’s system [3]. As shown in table 1, for perplexity, they have similar performance. As for human evaluation, our system performed better on syntactic satisfaction. Since we use acrostic poems to make the generative classical poetry associated with the given picture, PG-image gets a good grade on correlation satisfaction.

TABLE I
 PERFORMANCE COMPARISON

Models	Perplexity	Human Evaluation		
		Syntactic	Semantic	Correlation
SMT	121	0.802	0.516	0
<i>PG-image</i>	128	0.853	0.435	0.735

VI. CONCLUSION

In this paper, we have present a novel approach for Chinese classical poetry generation given an image based on the technology of computer vision and natural language processing. Also, we have translated the label of ImageNet 1000 into Chinese using Youdao online dictionary. Meanwhile, we review the results manually. For the first time, we use the form of acrostic poems to make sure that the generative poetry is related to the input image. The application makes Chinese classical poetry much closer to people’s daily life.

From the traditional method to deep learning, poetry generation technology has made great development. And even to a certain extent, it can produce poetry that ordinary people can’t distinguish easily. But the existing technology can’t learn thoughts and feelings in the poetry. Therefore, although the result seems to be a poem, but still lack human spirituality.

There are lots to do for our approach in the future. Based on previous work, our approach is extensible. We will improve our approach to generate better poems, even Chinese couplets. We also hope our work could be helpful to other related work.

ACKNOWLEDGMENT

We would like to thank all the reviewers for their valuable and constructive comments. We are grateful to ChenGui, WuBohao for participating in our study.

REFERENCES

- [1] (2017) Daoxiang Computer Poetry Machine website. [Online]. Available: <http://www.poeming.com/web/index.htm>
- [2] L. Jiang, M. Zhou. “Generating Chinese couplets using a statistical MT approach”, in Proc. International Conference on Computational Linguistics, 2008, pp.377-384.
- [3] J. He, M. Zhou, and L. Jiang. “Generating Chinese Classical Poems with Statistical Machine Translation Models”, in Proc. Twenty-Sixth AAAI Conference on Artificial Intelligence, 2012.
- [4] X. Yi, R. Li, M. Sun. “Generating Chinese Classical Poems with RNN Encoder-Decoder”. arXiv: 1604.01537, 2016.
- [5] X. Zhang, M. Lapata. “Chinese Poetry Generation with Recurrent Neural Networks”, in Proc. Conference on Empirical Methods in Natural Language Processing, 2014, pp.670-680.

- [6] R. Yan, H. Jiang, M. Lapata, et al. “i, Poet: Automatic Chinese Poetry Composition through a Generative Summarization Framework under Constrained Optimization”, in Proc. International Joint Conference on Artificial Intelligence, 2013, pp.2197-2203.
- [7] R. Yan. “i, Poet: Automatic Poetry Composition through Recurrent Neural Networks with Iterative Polishing Schema”, in Proc. International Joint Conference on Artificial Intelligence, 2016, pp.2238-2244.
- [8] M. Ghazvininejad, X. Shi, J. Priyadarshi, et al. “Hafez: An Interactive Poetry Generation System”, in Proc. ACL, 2017, pp.43-48.
- [9] R. Kiros, R. Salakhutdinov, R. Zemel. “Multimodal neural language models”, in Proc. International Conference on International Conference on Machine Learning, 2014, pp.595-603.
- [10] R. Kiros, R. Salakhutdinov, R S. Zemel. “Unifying visual-semantic embeddings with multimodal neural language models”. arXiv: 1411.2539, 2014.
- [11] O. Vinyals, A. Toshev, S. Bengio, et al. “Show and tell: A neural image caption generator”, in Proc. Computer Vision and Pattern Recognition. IEEE, 2015, pp.3156-3164.
- [12] K. Xu, J. Ba, R. Kiros, et al. “Show, attend and tell: Neural image caption generation with visual attention”, in Proc. International Conference on Machine Learning, 2015, pp.2048-2057.
- [13] K. Simonyan, A. Zisserman. “Very deep convolutional networks for large-scale image recognition”. arXiv:1409.1556, 2014.
- [14] (2017) Microsoft Xiaobing website. [Online]. Available: <http://www.msxiaoice.com>
- [15] (2016) ImageNet website. [Online]. Available: <http://image-net.org>
- [16] K. Papineni, S. Roukos, T. Ward, et al. “IBM Research Report Bleu: a Method for Automatic Evaluation of Machine Translation”. in Proc. Annual Meeting of the Association for Computational Linguistics, 2002, pp.311--318.
- [17] Li Wang 2002. A summary of rhyming constraints of Chinese poems. Beijing Press.