Learning to be Poetic: Automatic Generation of Chinese Song Ci Using RNN

Nan Du
Michigan State University
East Lansing, MI 48823, USA
dunan@msu.edu

Wei Wang
Michigan State University
East Lansing, MI 48823, USA
wangwe90@msu.edu

Zhuangdi Zhu
Michigan State University
East Lansing, MI 48823, USA
zhuzhuan@msu.edu

ABSTRACT

Judy: To be re-written In this project, we are going to develop a poem generator for Ci poetry. A system based on recurrent neural network (RNN) will be implemented to solve the sequence-to-sequence learning problem. We will also compare the performance of this model with the traditional automatic method that used to generate poem. We hope that our system can learn the complete rule from training dataset without any given constraints and generate poems with meaningful syntax that following the special rules for rhyme and tone in the Ci.

Keywords

Song Ci (poetry); Recurrent Neural Network; sequence-to-sequence learning problem

1. PROBLEM DESCRIPTION

1.1 Motivation

Judy: To be extened In this project, we propose and evaluate different approaches to automatically generate Chinese poems. Especially, we study how to automatically generate Chinese Ci using machine learning skills. Ci are one of the most important genres of Chinese classical poetry. As a precious cultural heritage, not many of them have been passed down onto the current generation. Therefore, the study of automatic generation of Ci is meaningful, not only because it supplements entertainment and education resources to modern society, but also because it demonstrates the feasibility of applying artificial intelligence in Art generation.

1.2 Background

Judy: To be extened Ci is a form of Chinese classical poetry. It arose with the so-called banquet music in Tang dynasty and reached its peak one hundred years later, as a major alternative to Shi poetry [2] .

Derived from the structure used in Tang poetry, Ci follows strict rule determining the number of characters for different lines, the arrangement of rhyme, and the location of tones. There are more than 800 rule sets for Ci, which is called Cipai [8]. The author of Ci needs to fill in the words according to the matrix associated to the Cipai. The uneven lines in Ci follow more continuous syntax than traditional Chinese Tang poetry [2].

1.3 Proposed Approach

Judy: To be extened We propose an AI system which generate Ci in an interactive approach. First, our system will prompt the user to provide a Cipai name. Because Ci belonging to different Cipai may contain different emotions or grammatical rules. Next, the system will receive few of keyword inputs that convey the detailed sentiments of the Song Ci. the first sentence of the iambic will be generated based on the keyword inputs. Further, the system generate following sentences based on previously-generated contexts using both RNN and SMT technique. Finally, we evaluate the quality of the generated Ci using an evaluation tool named BLEU.

1.4 Technical Challenges and Proposed Solutions

Judy: To be extened The first challenge to build a general model for all types of Song Ci. Different from Shi poetry whose structure is strict, Song Ci has more than 800 set of Cipai, and different Cipai follows different structural or rhythmic patterns. Therefore, it is difficult to generalize a model for all the Song Ci from limited training dataset. Our solution is to create a model based on Recurrent Neural Network. For every line generated in the SongCi, its probability is based on the probability of all previously lines.

Another challenge is to maintain consistent and poetic meanings throughout the generated SongCi. Compared with Shi poetry, Song Ci are much longer in length and therefore more complicated in context. It is difficult to keep long-distance memory using conventional RNN. Our solution is to use a Long Short Term Memory (LSTM) model that can track the long-distance information.

2. RELATED WORK

Judy: To be extened Approaches to poetry automatic generation can be divided into the following categories. **Using rules and templates.** This approach adopts templates to generate poems that comply with grammatical rules, such as the rhythms, lines, and word frequencies [6, 9].

Using evolutionary algorithms. This approach is mainly based on natural selection. It generates all possible candidates, and use search and evaluation algorithms to select the optimal one [4, 5].

Using Statistical Machine Translation (SMT) methods. [3]. This approach first receives keywords and extract most relevant constituents to theses keywords. Next, it generates poems by iteratively selecting among these constituents based on phonological, structural, and poetic requiremetns.

Using neural network. This approach adopts an RNN Encoder-Decoder structure [1, 7]. It generates new iambics context using previously-generated contexts. based on the rationale that, in Chinese poems, two consecutive lines have high semantical relevance.

3. DATA DESCRIPTION

Judy: To be filled ...

3.0.1 Tang Poetry Corpus

3.0.2 Song Ci Corpus

4. PROJECT MILESTONES

4.1 Completed Milestones

Judy: To be re-written

4.1.1 Background Survey

For this initial step, we plan to search for related works to computational literary creation to gain the basic knowledge of Song Ci. We are interested in the following questions: what is the criterion of a good Song Ci? How to evaluate the correctness, fluency and style of poems generated? Better understanding of related work and Song Ci composition rules will provide us with great help for the following work, especially algorithm testing and comparison.

4.1.2 Corpus Search and Analysis

Second, we will search and select a proper Song poem corpus for our project. The ideal corpus should be comprehensive on poem styles, and are precisely analyzed for content.

4.1.3 Implementation of Vector Space Model

Test cases will be generated with both poem generator under the same keywords and topics. Poems will be test on aspects of grammar, semantic correctness, style and content. Both computational evaluation and human evaluation are expected to be used in last part of our project.

4.1.4 Implementation of RNN + SLTM Model

Third, implementation poem generator based on both algorithm of RNN and genetic algorithm would be the most important work in our project. So we will assign more time on this step.

4.2 Remaining Milestones

- Implementation of Genetic Algorithms
- Model Testing and Comparison
- Project Summary and Writing

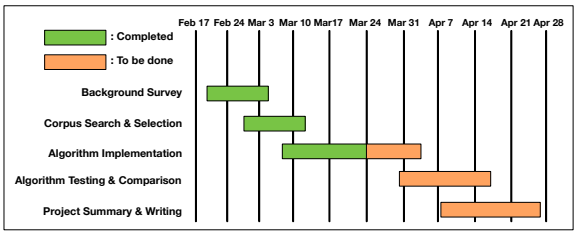

Figure 1: Project Timeline

5. REFERENCES

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

[2] Z.-q. Cai. *How to read Chinese poetry: a guided anthology*. Columbia University Press, 2008.

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

[4] H. Manurung. An evolutionary algorithm approach to poetry generation. 2004.

[5] 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, 2012.

[6] N. Tosa, H. Obara, and M. Minoh. Hitch haiku: An interactive supporting system for composing haiku poem. In *International Conference on Entertainment Computing*, pages 209–216. Springer, 2008.

[7] Q. Wang, T. Luo, D. Wang, and C. Xing. Chinese song iambics generation with neural attention-based model. *arXiv preprint arXiv:1604.06274*, 2016.

[8] Wikipedia. Ci (poetry). https://en.wikipedia.org/wiki/Ci_(poetry)/, last accessed 2017.

[9] X. Wu, N. Tosa, and R. Nakatsu. New hitch haiku: An interactive renku poem composition supporting tool applied for sightseeing navigation system. In *International Conference on Entertainment Computing*, pages 191–196. Springer, 2009.