Generating Chinese Ci with Designated Metrical Structure

Richong Zhang[1], **Xinyu Liu**[2], **Xinwei Chen**[2]
Zhiyuan Hu[1], Zhaoqing Xu[1], Yongyi Mao[2]

[1]Beihang University, China
[2]University of Ottawa, Canada

November 15, 2018

- Deep neural nets are reshaping the landscape of NLP and have been applied successfully in natural language generation:
  - dialogue generation
  - document summarization
  - machine translation
  - ...
- This work: automatic generation of Ci, an ancient Chinese poetry

Fact

Ci is an ancient Chines poetry form with highly restricted metrical rules. The metrical rules that a Ci must follow are specified by a Cipai. A Cipai specifies (at least) three kinds of rules:

- *a rhythmic rule*
- *a tonal rule*
- *a rhyming rule*

Cipai：忆江南（Yi Jiang Nan）

Rhythmic Rule:	3 characters，5 characters。 7 characters， 7 characters 。 5 characters?
Ci:	江南好，风景旧曾谙。日出江花红胜火，春来江水绿如蓝。能不忆江南?
Tonal Rule:	0 + - ， 0 - - + + 。 0 - 0 + + - - ， 0 + + - - + + 。 0 - - + + ?
Rhyming Rule:	- - - ， - - - - x 。 - - - - - - - ， - - - - - - x 。 - - - - x ?
Translation:	Fair Southern shore, with scenes I adore. At sunrise riverside flowers redder than fire. In spring green waves grow as blue as sapphire. Which I cannot but admire.

Fact (Rhythmic Rule)

A rhythmic rule specifies the number of lines (i.e., sentences/clauses) in a Ci and the number of characters in each line.

Cipai：忆江南（Yi Jiang Nan）

Rhythmic Rule:	3 characters，5 characters。 7 characters， 7 characters。 5 characters?
Ci:	江南好，风景旧曾谙。日出江花红胜火，春来江水绿如蓝。能不忆江南?
Tonal Rule:	0 + - ， 0 - - + + 。 0 - 0 + + - - ， 0 + + - - + + 。 0 - - + + ?
Rhyming Rule:	- - - ， - - - - x 。 - - - - - - - ， - - - - - - x 。 - - - - x ?
Translation:	Fair Southern shore, with scenes I adore. At sunrise riverside flowers redder than fire. In spring green waves grow as blue as sapphire. Which I cannot but admire.

Fact (Tonal Rule)

A tonal rule specifies the tone of the character in each location of the Ci.

- *The tones are classified into two categories: Ping (+) and Ze (-).*
- *A tonal rule specifies if a character in each location should take +, −, or 0, where 0 refers to no tonal requirement.*

Cipai：忆江南（Yi Jiang Nan）

Rhythmic Rule:	3 characters，5 characters。 7 characters， 7 characters 。 5 characters?
Ci:	江南好，风景旧曾谙。日出江花红胜火，春来江水绿如蓝。能不忆江南?
Tonal Rule:	0 + - ， 0 - - + + 。 0 - 0 + + - - ， 0 + + - - + + 。 0 - - + + ?
Rhyming Rule:	- - - ， - - - - x 。 - - - - - - - ， - - - - - - x 。 - - - - x ?
Translation:	Fair Southern shore, with scenes I adore. At sunrise riverside flowers redder than fire. In spring green waves grow as blue as sapphire. Which I cannot but admire.

Fact (Rhyming Rule)

A rhyming rule specifies characters at which locations should rhyme.

- *Fundamental sounding units of Chinese characters: initials (≈ consonants) and finals (≈ vowels)*
- *Rhyming are a final, called a "rhyming foot" or Yunshe, which may take 16 different sounds.*
- *Lines may form groups; lines in one group rhyme on the same foot; different groups use different rhyming feet.*

- There have been various works on poetry generation.
  - template-based, e.g. (Oliveira 2012; Oliveira et al. 2014; Rashel and Manurung 2014)
    - requiring human expert's design of templates
  - learning-based, e.g., (Yi, Li, and Sun 2017; Zhang et al. 2017; Wang et al. 2016; Yan 2016; Zhang and Lapata 2014; Yang et al. 2017)
    - learning completely from a poetry corpus

Highlight (Problem Statement)

Given a Cipai (namely, the metrical rules specified therein), how can we use a deep neural network model to generate a Ci complying with the Cipai?

- Previous deep models for Ci generation do not exploit the given Cipai information, while hoping the model is sufficiently trained to generate Ci complying with some Cipai.

Will the symbolist meets the connectionist?

- Rule-based learning traditionally belongs to the symbolist paradigm of AI.
- Neural networks and distributed representations belong to the connectionist paradigm of AI.
- Can the two disconnected paradigms be combined?

Highlight

Generating Ci with designated metrical structure is a perfect example for investigating whether one can bridge the great divide between the connectionist and the symbolist.

Contribution of This Work

- We adopt the CVAE (conditional variational auto-encoder) framework for generating Ci for a given Cipai.
- We explicitly encode the metrical rules in the given Cipai into distributed representations and feed them to the neural networks.
- Experiments show that our model (MRCG) is capable of generating Ci satisfying the metrical requirements nearly perfectly without degrading in semantics.

Outline

1. VAE Framework
2. MRCG Model
3. Experimental Results

Outline

1. VAE Framework
2. MRCG Model
3. Experimental Results

- Let $c = (s_1, s_2, \ldots, s_L)$ be a Ci with L lines.
- Each s_l is a sequence of $N(l)$ words $(w_{l,1}, w_{l,2}, \ldots, w_{l,N(l)})$.
- Let r denote the Cipai regulating the metrical structure of c. We will specify the form of r later.
- A natural probability model for c with a designated Cipai r

$$p_{C|R}(c|r) := \int p_{C|RZ}(c|r,z)p_Z(z)dz, \tag{1}$$

  where Z is a *latent* semantic representation of Ci C.
- Here we have assumed that the Ci C depends on its latent semantics Z and its metrical structure R jointly and that Z does not depend on the Cipai R.
- But the parameters of (1) is intractable to learn using log-likelihood.

- Any arbitrary conditional distribution $q_{Z|CR}$ gives a variational lower bound of $\log p_{C|R}(c|r)$:

  $$\log p_{C|R}(c|r) \geq -\mathrm{KL}(q_{Z|CR}(\cdot|c,r)||p_Z) + \mathbb{E}_{z\sim q_{Z|CR}}\{\log p_{C|RZ}(c|r,z)\}$$

  where the equality is achieved when $q_{Z|CR} = p_{Z|CR}$.
- We will restrict $q_{Z|CR}$ to the form of $q_{Z|C}$.
- $p_{C|RZ}$: "decoder", parametrized by Θ
- $q_{Z|C}$: "encoder", parametrized by Φ
- p_Z: a fixed distribution
- Instead of maximizing the log-likelihood $\log p_{C|R}(c|r)$, we will maximize its variational lower bound.

- Define the "KL loss" by

$$\ell_{\mathrm{KL}}(c, r; \Phi) \quad := \quad \mathrm{KL}(q_{Z|CR}(\cdot|c, r)||p_Z)$$

- Define the "reconstruction loss" by

$$\ell_{\mathrm{rec}}(c, r; \Theta, \Phi) \quad := -\mathbb{E}_{z \sim q_{Z|CR}} \left\{ \log p_{C|RZ}(c|r, z) \right\}.$$

- Let $\mathcal{D}$ denote a set of training examples, each being a pair (c, r).
- Then learning of model (1) can be solved by maximizing the variational lower bound or minimizing the loss function

$$\mathcal{L}(\Theta, \Phi) := \sum_{(c,r) \in \mathcal{D}} \{\ell_{\mathrm{KL}}(c, r; \Phi) + \ell_{\mathrm{rec}}(c, r; \Theta, \Phi)\} \tag{2}$$

- This is CVAE.

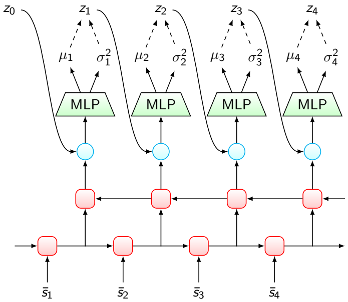
Outline

Encoder

- Character Embedding: $\overline{w}_{l,j} := \mathbf{D}w_{l,j}$
- Line Embedding (GRU):

$$\begin{aligned} h^{\mathrm{s}}_{l,j} &:= \mathbf{GRU}(h^{\mathrm{s}}_{l,j-1}, \overline{w}_{l,j}) \\ \overline{s}_l &:= h^{\mathrm{s}}_{l,N(l)} \end{aligned}$$

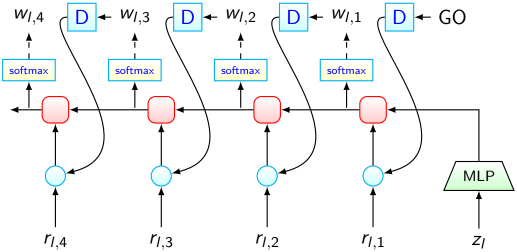
- Context Embedding (2-layer GRU)
- Latent Semantics Encoding (MLP)



From line embedding to latent semantics

- blue circle: concatenation
- red box: GRU

Decoder



Generate a line of characters

- blue circle: concatenation
- red box: GRU
- $r_{l,j}$: rule encoding for the j^{th} character in the l^{th} line.

Metrical Rule Encoding

- $\mathcal{T} := \{+, -, *\}$ denotes the set of possible tones, where $+$ and $-$ denote Ping and Ze respectively and $*$ denotes "NO SOUND".
  - "NO SOUND": the tone for the punctuation marks, the unknown or rare ancient characters whose tone can not be determined.
- Each tone in $\mathcal{T}$ is identified with a one-hot vector in $\mathbb{R}^3$, referred to as a tone vectors.
- $a(w)$ denotes its tone vector of character w.
- $\mathcal{Y} := \{1, 2, \ldots, 16, *\}$ denote the set of all possible rhyming feet (recalling there are 16) and the "NO SOUND"
- A rhyming foot in $\mathcal{Y}$ is identified with a vector in $\mathbb{R}^{17}$, referred to as a Yunshe vector.
- $b(w)$ denotes the Yunshe vector of character w.
- $a(w)$ and $b(w)$ may be seen as probability distributions over $\mathcal{T}$ and $\mathcal{Y}$.

- Construct a tone-rule vector $\widetilde{a}_{l,j}$ for the j^{th} character in the l^{th} output line, which is vector in $\mathbb{R}^3$ representing a distribution over $\mathcal{T}$.
  - If $r_{l,j}$ specifies that $w_{l,j}$ must take tone "$+$" (resp. "$-$"), then $\widetilde{a}_{l,j}$ is the one-hot vector for "$+$" (resp. for "$-$").
  - If $r_{l,j}$ specifies that $w_{l,j}$ must be a punctuation mark, then $\widetilde{a}_{l,j}$ is the one-hot vector for "$*$".
  - If $r_{l,j}$ does not specify the tone of $w_{l.j}$, then $\widetilde{a}_{l,j}$ is the vector $[.5, .5, 0]^T$, which puts probability 0.5 on "$+$" and "$-$", and puts probability 0 on "$*$".

- Construct a rhyme-rule vector $\widetilde{b}_{l,j}$, which is a vector in $\mathbb{R}^{17}$ representing a distribution over the $\mathcal{Y}$.
  - If $r_{l,j}$ suggests that $w_{l,j}$ is a character (i.e., not a punctuation mark) and not a rhyming foot, $\widetilde{b}_{l,j}$ is taken as the empirical distribution, say, $\tilde{p}$, over the 16 Yunshe's in the dataset. [1]
  - If $r_{l,j}$ suggests that $w_{l,j}$ is a punctuation mark, then $\widetilde{b}_{l,j}$ is taken as the one-hot vector for "$*$".
  - if $r_{l,j}$ dictates that $w_{l,j}$ is a rhyming foot in a rhyming group, then $\widetilde{b}_{l,j}$ need to be specified with the rhyme-rule vectors $\widetilde{b}_{l',j'}$'s for all other rhyming feet $w_{l',j'}$'s in the same rhyming group. In this case, for this rhyming group, a random Yunshe Y is drawn from $\widetilde{p}$, and the rhyme-rule vector is set to the one-hot vector for Y.

[1] When using the trained MRCG to generate a Ci, we find it is beneficial to make the $\widetilde{b}_{l,j}$ "peakier". Then we actually draw M samples from $\tilde{p}$ and use the empirical distribution of the M samples as $\widetilde{b}_{l,j}$.

Highlight

The encoding of the metrical rule for the j^{th} character in the l^{th} line is:

$$r_{l,j} := \mathbf{concat}(\widetilde{a}_{l,j}, \widetilde{b}_{l,j}). \tag{3}$$

Training of MRCG

- Mini-Batched SGD
- Reparametrization Trick (Kingma/Welling, 2014) to allow Backprop go through sampler
- KL clamping to avoid KL vanishing

Outline

Dataset

- We crawl a dataset from a Chinese poetry website: `https://sou-yun.com/`
- The dataset contains 82,724 Ci's, written for 818 Cipai's; on average, there are 102 Ci' per Cipai.
- 95% of the corpus are contributed by the most popular 314 Cipai's; the most popular 20 Cipai's contribute to about 45% of the corpus.
- The average number of lines per Ci is about16.
- The number of characters per line is about 7 on average.
- For each Cipai, we randomly select 5% of its Ci's to assemble the testing set. For the Cipai's having less than 20 Ci's, they are not included in the testing set.
- In total, 3,797 Ci's in the testing set and 78,927 in the training set

Evaluated Models

- MRCG
- MRCG$^-$: MRCG with metrical-rule encoding removed.
- Seq2Seq model
- Attention

Note

- *In both Seq2Seq and attention model, bi-directional GRU layers are used in the encoder, and single-directional GRU layers is used as the decoder.*
- *To compensate for the structural disadvantages of the baselines relative to MRCG, we used multiple such layers in the encoder and decoder so that the resulting model has a similar number of parameters as MRCG.*

Two Modes of Ci Generation

Highlight (Blind Generating)

- *In this mode, a random Cipai is passed to the model as input, and a latent semantic representation is drawn from the prior p_Z to generate an output Ci.*
- *This mode is only for MRCG and MRCG$^-$.*

Highlight (Guided Generating)

- *This modeapplies to all compared models.*
- *For the baselines, they take a random input Ci from the testing set as input and attempt to reproduce it at the output.*
- *For MRCG and MRCG$^-$, they take as input the same random Ci and its Cipai; the latent semantic representation is drawn from the approximating posterior $q_{Z|C}$.*

Metrical Performance: Metrics

Highlight (Length Accuracy)

The length correctness of the Ci is defined as the percentage of lines in the Ci that have the correct length. The length accuracy is the average of length correctness over all generated Ci's.

Highlight (Tone Accuracy)

The tone correctness of the Ci is defined as the percentage of the characters in the Ci that have the correct tone. The tone accuracy is the average of tone correctness over all generated Ci's.

Highlight (Rhyme Accuracy)

Given a generated Ci and each rhyming group within, the rhyme correctness of the rhyming group is defined as the fraction of rhyming foot locations that have the correct Yunshe. Specifically, the correct Yunshe is taken as the majority Yunshe in the rhyming group. The rhyme accuracy is the average of rhyme correctness over all rhyming groups in a Ci and then further averaged over all generated Ci's.

Metrical Performance: Results

Table: Metrical Performances

Model	Blind Generation			Guided Generation		
	Length	Tone	Rhyme	Length	Tone	Rhyme
Seq2seq	-	-	-	27.1	55.0	29.8
Attn	-	-	-	28.2	56.3	26.7
MRCG^{-}	35.49	59.43	33.34	41.96	66.21	39.77
MRCG	**99.21**	**92.03**	**96.87**	**99.37**	**93.71**	**98.28**

Semantics Performance: Test Protocols

Highlight (Absolute Semantics Test (AST))

Following (He, Zhou, and Jiang 2012; Jiang and Zhou 2008), this test is conducted as follows. A random Ci is generated from a model. For every two consecutive lines therein, say $(\widehat{s}_1, \widehat{s}_2)$, we find from a Ci database Shixuehanying 10 lines that are the most similar to $\widehat{s}_1$. For each of these 10 lines, we collect its next line in the database into a set S_2, then the BLEU score of $\widehat{s}_2$ against S_2 is computed. After repeating this experiments, the average BLEU score is used to evaluate the model.

Highlight (Relative Semantics Test (RST))

We devise this test to evaluate, under guided generation, whether the outputs of a model preserve the similarities and differences between its inputs. Specifically, a pair of Ci's (c_1, c_2) are drawn at random from the testing set, both of which are used as an input to the model. Let $(\widehat{c}_1, \widehat{c}_2)$ be their respective outputs from the model. We then compute the BLEU score S of c_1 against c_2 and the BLEU score $\widehat{S}$ of $\widehat{c}_1$ against $\widehat{c}_2$. After this experiment is repeated many times, the Pearson Correlation Coefficient between S and $\widehat{S}$ is used to evaluate the model.

Semantics Performance: Human Evaluation

Highlight (Human Evaluation)

A total of 12 Human examiners are invited to evaluate 25 Ci's randomly generated by each model (without revealing the identity of the model). Specifically, fluency (Flu), theme consistency (Thm), aesthetics (Aes) and overall (All) performance are evaluated using scores $\{1, 2, 3, 4\}$*, where the examiners are told to calibrate the score of* 1 *to "poor" and the score of* 4 *to "expert level".*

Semantics Performance: Results

Model	Objective Tests		Human Evaluation			
	AST	RST	Flu	The	Aes	All
Seq2Seq	**0.242**	0.358	**2.65**	**2.8**	2.26	2.33
Attn	0.221	0.323	2.53	2.72	2.25	2.21
MRCG-	0.235	**0.541**	2.38	2.41	2.34	2.26
MRCG	0.229	0.529	2.59	2.68	**2.54**	**2.45**

Examples of Generated Ci by MRCG

Cipai：长相思（Chang Xiang Si）	
钿钗头	The pretty hairpin I wear
等闲愁	With loneliness and boredom
泪滴红蕤衬玉钩	Tears drop on the moon-lit pillow
夜来微湿秋	Lightly drenching this night of autumn
淡墨淋	Let my pen and ink
东晋游	Run through the Dynasty of Jin
百啭莺吭春梦稠	In the dream of romance, let thousands of doves chirp
秋千闲倚楼	Where on a leisureful balcony, I play on a swing

Cipai：清平乐（Qing Ping Yue）	
消魂销尽	Have you lost your mind
昨梦沈郎不	Dreaming of Mr. Shen last night?
莫惜飘茵随逝日	Don’t pity the green grass withering in the wind of time
泪雨潇湘月明	Tears like rain, you linger in the moonlight
只愁听玉钩花	Sadness when you listen to an old song
白苹桥外啼鸦	Resonates with the raven, crying on the bridge
别离恨愁无迹	Leave behind the pain of parting
吟怀谁把年华	For an old memory, who would sacrifice youth, beautiful and shining?

Limitations of MRCG

- We believe that MRCG represents the state of the art in the neural models for Ci generation.
- But MRCG still exhibits a visible gap from human Ci writers (won't pass the Turing Test).
    - Certain implicit rhythmic convention may not be obeyed. For example, a 5-character line supposedly following the $(AB)(CD)(E)$ parsing may be generated to follow the $(A)(BC)(DE)$ parsing. Such a problem may be resolved by more delicate design of rule decoding.
    - Although MRCG is capable of expressing moods, sentiments and emotions very well by creating imageries containing the essential related elements, the model appears to be weak in generating a rich story with logical coherence.

Concluding Remarks

- We present the first neural model, MRCG, that explicitly encodes metrical structure in Ci generation.
- We demonstrate that MRCG generates nearly perfect metrics without sacrificing semantics.
- This exercise suggests that it is possible to integrate the symbolist paradigm in the connectionist learning framework.
- After this work, the real challenge in poetry generation, and more broadly in natural language generation, remains on the semantics side.
- We believe that generating "human-level" semantics relies on the development of new natural language understanding models, particularly those capable of representing reasoning at a fundamental level.