# VERBS PREDICTION AND METAPHOR CONSTRUCTION IN MODERN CHINESE POETRY ENRICHED WITH GLYPH INFORMATION

**BY ZIYANG JIA**

**A thesis submitted to the**

**Graduate School—New Brunswick**

**Rutgers, The State University of New Jersey**

**in partial fulfillment of the requirements**

**for the degree of**

**Master of Science**

**Graduate Program in Computer Science**

**Written under the direction of**

**Matthew Stone**

**and approved by**

**New Brunswick, New Jersey**

**October, 2019**

# ABSTRACT OF THE THESIS

## Verbs Prediction and Metaphor Construction in Modern Chinese Poetry Enriched With Glyph Information

**by Ziyang Jia**

**Thesis Director: Matthew Stone**

Recent research on metaphor processing reflects the importance of employing the properties of human cognition. In this paper, we propose a modern Chinese poetic model to predict the verbs of a line according to the context and compose a metaphor by enriching the verb with glyph properties of the vehicle. Our model consists of three parts: First, a language model which employs multi-layer encoder; Second and innovatively, we employ a glyph model to emphasize the inherent information of Chinese character structure; Third, a rhyme model to apply the patterns of end rhymes, if the verb is at the end of a line.

## Acknowledgements

I'd like to thank my mentor, Prof.Matthew Stone who kept on pushing me forward by all his means. I'd also like to thank Jing Feng who helped me on the expert judgement of our model. Finally, I'd like to thank my dear parents who cheered me up when the experiments went to failures.

# Dedication

My friends Xunjie Zhu, Siyuan Wu and Lu Liu.

# Table of Contents

# Chapter 1

# Introduction

In the past few years, natural language processing is becoming more and more important thanks to the improved computing models. It has made huge improvements on text classification (Zhang et al. [2015]), machine translation (Bahdanau et al. [2014]), Q&A task (Devlin et al. [2018]) and named entity recognition(Lample et al. [2016]). In these studies, most works are based on English corpus which consists of broadly available large language texts. But for other languages, like Chinese, the models may not make full use of the information provided by the words or characters leading to the omission of some interesting linguistic features.

Metaphor is a significant figure of speech for human languages and computational metaphor generation is a replication of human creativity. Researchers craft reasonable metaphors via extracting triples from web (Veale and Hao [2007]) or knowledge-based database (Veale [2016]). The generation of metaphors depends heavily on the knowledge extracted from similes crawled from web (Veale [2016]). Although the outputs of these approaches successfully avoid under-specified inference, the knowledge triples determine the domain of the metaphors. Besides, in poetic languages, the elements in metaphors are not always shown in common similes.

In this work, we propose a heuristic method to transform Chinese poetic lines into metaphorical lines by enhancing the glyph correlations between the nouns and verbs.

Chinese characters are often composed of meaningful components (radicals and radical-like components), as stated via Li et al. [2015]. But different from other languages, like English, the pronunciations of two characters with common components can be totally different which benefits the readability of the verse with similar-looking characters in a short distance, like 卵(egg, pronounce as luan in Pinyin) and 孵(hatch, pronounce as fu in Pinyin).

These evidences inspire us to craft metaphors by enhancing glyph correlation between verbs and nouns in Chinese verses. According to the expert judgement in Sec. 3.4, our model achieved considerable improvement on metaphor composition compared with human verse lines.

# Chapter 2

# Background Knowledge and Previous Works

## 2.1 Metaphor

Metaphors illustrates literal breaking of linguistic preferences of the corresponding tenors, the subject to which new attributes are applied to. And the vehicle in a metaphor is the object where the attributes are borrowed. As stated by Wilks et al. [1996], the metaphor can also be interpreted as the linkage between the nested viewpoint, or propositional attitude, of entity A on entity B and the belief environment of entity B in the knowledge-based system, which is more objective.

Since the weakness of linkage between tenors and vehicles in metaphoric expressions, it's a challenging task to process or generate metaphors. In recent years, metaphor processing is a rapidly growing area in NLP (Klebanov et al. [2016]), the tasks includes metaphor detection and structural mapping, (Lederer [2016], Wilks et al. [1996]), metaphor generation (Veale and Hao [2007]), and crafting metaphor triple resources (Veale [2016]).

Our model transforms a Chinese poetic line into metaphor by ascribing glyph properties of the object to the verb. Then borrowed attributes will be passed to the subject making it possible to compose a tenor-vehicle-verb triplet if the flowing attributes are new and meaningful for the subject.

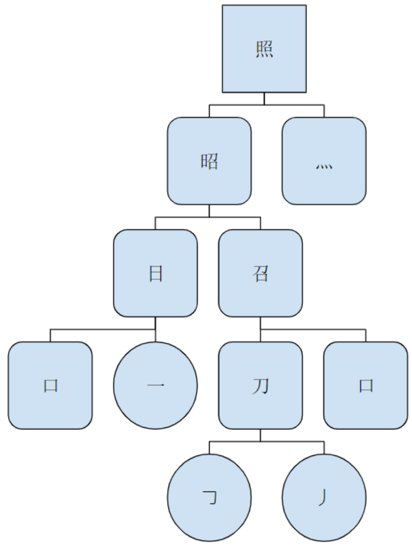



Figure 2.1: Hierarchical composition structure of 照(shine) . Squares represents origin characters; Rounded squares represents radicals; Circles are meaningless components.

## 2.2 Glyph, Character Components and Chinese Modern Poetry

As stated by Su and Lee [2017], Chinese characters contain rich semantics via glyph. In Chinese sentences, the verbs and nouns commonly have relationships in glyph, which is constituted by the components. Like the sentence in Table 2.1, where 月(moon) is the opposite of 日(sun) which is one of the components of 照(bask); 光(light) is also part of the glyph of 耀(shine). However, not all Chinese characters are composed of the simplest radicals. Like the character 照mentioned above, it has a hierarchical structure of components composition including both radicals and non-radicals, see Figure 2.1. The sophisticated structures of Chinese characters make it a heavy workload to capture the useful glyph information by weighting and representing each components directly. Inspired by Li et al. [2015], we employ the radicals and some selected components to express the inherent information of the character which will be illustrated in detail in Section 3.3.2.

This characteristic can also be used in poems. Modern Chinese poetry is a vernacular style poetry which is more free as opposed to traditional Chinese poems that strictly follow the formats, rhyming and complicated tonal rules, as stated byYan et al. [2013]. And also, modern Chinese poetry breaks the serious impression of classic poems in various ways of expression including metaphor. Like the sentence in Table 2.2, where 亻(human) is the common component of 你(you) and 停(stand, stay, stop) through which the sentence employs personification to illustrate a gentle and warm scene for the 夕照(evening glow).

Besides, other kinds of metaphors are also easy to find. In the verse shown in Table 2.3 composed by Gu-Cheng, a famous Chinese modern poet, the appearance of both 孵化(hatch) and 卵石(cobbles) leads to an interesting imagination. Since in Chinese, 卵石means *a stone like an egg* and *hatch an egg* is a rational cue inspired by the sentence. And this relationship, with the glyph characteristic of Chinese characters, is easy to get. And also, 卵石(cobbles) is reasonable to be followed by the verb 垒(pile). According the the components, 石(stone), and 土(dust), the bottom half of 垒, have high cosine similarity in the embedding space and then another verb-tenor pair is elaborated: verb 垒(pile) and pronoun 我(I).

For the metaphor tasks, the characteristics of Chinese grammar are particularly in favor of the verbs in the metaphorical sentences. There are two most frequent syntactic constructions for metaphoric and literal uses: subject-verb-object (SVO) and adjective-noun (AN) tuples. According to Tsvetkov et al. [2014] and Shutova and Teufel [2010], SVOs account for a considerable large proportion, approximately 60%, of all metaphoric expressions while AN takes 24% and others take the remaining 16%. In English sentences,

| | | | | | | |
|---|---|---|---|---|---|---|
| Sentence | 明晃晃 | 的 | **月光** | **照耀** | 着 | 池塘 |
| Pinyin | ming huang huang | de | yue guang | zhao yao | zhe | chi tang |
| Word in English | bright | 's, of | moon light | shine | -ing | pool |
| Sentence in English | The bright moon light is shining into the pool. | | | | | |

Table 2.1: Sentence example with verb 照耀and subject 月光.

| Poetic Line | 你 | 肩上 | 停 | 着 | 夕照 |
|---|---|---|---|---|---|
| Pinyin | ni | jian shang | ting | zhe | xi zhao |
| Word in English | you | shoulder | stand, stay, stop | -ing | evening glow |
| Sentence in English | The evening glow is staying on your shoulder. | | | | |

Table 2.2: A metaphor consists of tenor 夕照(evening glow) , vehicle 人(people) (not mentioned in the verse) and verb 停(stop) . From 风起的时候*(when the wind is coming)* via poet Yang-Mu.

the adjectives composed of verbs and suffixes commonly behave as verbs in Chinese contexts, since the Chinese verbs don't have any tense or voice. Like in the phrase the broken promise (破碎/break 的/'s 誓言/promise ), obviously, this is a sample of adjective-noun (AN) in English, whereas, the verb break (破碎) doesn't transform in the Chinese sentence. Accordingly, in this work, we're focusing only on the linkages between noun and verbs (SVOs). From the examples of Table 2.2, 2.3, we can tell the importance of sharing same properties between verb and vehicle in composing a metaphor. So, specifically, we build our model focusing on the shared properties between the verb and object (vehicle).

## 2.3 Previous Works

### 2.3.1 Knowledge-Based Metaphor Modeling

The previous works on metaphor modeling are mostly based on additional background knowledge constructed with attributes and nested perspectives. Wilks et al. [1996] proposed

| Poetic line | 我 | 和 | 无数 | 不 | 能 | 孵化 | 的 | 卵石 | 垒 | 在 | 一起 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Pinyin | wo | he | wu shu | bu | neng | fu hua | de | luan shi | lei | zai | yi qi |
| Word in English | I | and | countless | not | can | hatch | 's, of | cobble | pile | at | together |
| Sentence in English | I'm piled together with countless cobbles which is not capable of being hatched. | | | | | | | | | | |

Table 2.3: Their are two pairs of metaphors in the verse: 1. tenor: 卵石(cobble) , vehicle: 鸡蛋/卵(egg) (not mentioned directly) and verb: 孵化(hatch) ; 2. tenor: 我(I) , vehicle: 卵石(cobble) and verb: 垒(pile) . From 微微的希望*(My puny hopes)* via poet Gu-Cheng.

a metaphor model based on the nested viewpoints. They believe metaphor is binded up with the propositional attitudes since the metaphorical view on a topic treats it as something else. Leaded by this idea, they buid ViewGen, a knowledge based belief engine, to ascribe belief from one agent to another by the notion of default reasoning, the process of drawing conclusions via statistics and induction. As a result, the metaphors for one are modeled as the states of mind of other people. The default reasoning plays an very significant role in metaphor collection, implied by Veale [2016]. According to this work, the most common way of building the knowledge base is gathering XYZ comparisons of the form "X is the Y of Z" with additional manually, suggested by the authors, or automatically polulated information extraction. The idea of this work is to construct a database of Y, a proper-named individual that is used to represent a whole cluster. By analysing the attributes of Y, they generates metaphors with a linkage between literal and the eventual interpretaion in the mind of a listener.

### 2.3.2 Transformer Encoder & Attention Mechanism

Transformer has a typical structure of encoder-decoder for the task of sequence transduction. The encoder maps the input $(x_1, ..., x_n)$ to a sequency of continuous representations $(z_1, ...z_n)$ for the hidden states. After that, the decoder generates an output sequence one element at a time. Transformer works well in the field of sequence-to-sequence tasks. However, considering our goal of word prediction out of a sequence, we can only employ the encoder to predict the verb according to the output of self-attention and feed forward. The Figure 2.2 shows the outline of the model but the utilized ones is just the encoder on the left side. The flow of the encoder is as the followings:

– Take a sentence;

– Embed each word in the sentence with pre-trained embeddings;

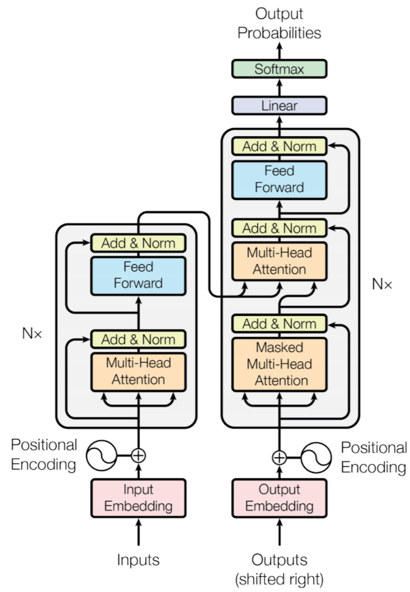



Figure 2.2: The Transformer - model architecture. In our work, we only empoy the encoder (left half) for the language model.

- Calculate the positional embeddings relying only on the positions, see 2.3.2;
- Pass the sum of word embedding and position embedding to the first layer of encoder;
- Each layer of encoder consists of two parts:
  - Multi-head self-attention to concatenate the attention from each other among the same sequence, see 2.3.2;
  - Feed-forward network separately and identically applied to each outputs of attention mode from different position, see ¡¡¡;
- Take the output of last encoder layer as input and pass it to the next layer;
- Repeat the process of last two steps for $N$ times, then the new expression for the input is generated.

**Positional Encoding**

Unlike models like LSTM which process data in a sequential way, Transformer runs in a parallel way in Encoder leading the absence of relative position. To add the position information to the model, Transformer encodes position with the sine and cosine functions of different frequencies, see Equation 2.1, allowing the linear transformation from $PE_{pos}$ to $PE_{pos+k}$ which means it expresses the relative position among the words in a line. In that way, Transformer encoder processes the words in a line simultaneously without lossing position information.

$$PE_{(pos,2i)} = sin(pos/10000^{2i/d_model}) \tag{2.1}$$

$$PE_{(pos,2i+1)} = cos(pos/10000^{2i/d_model}) \tag{2.2}$$

**Multi-head Self-attention**

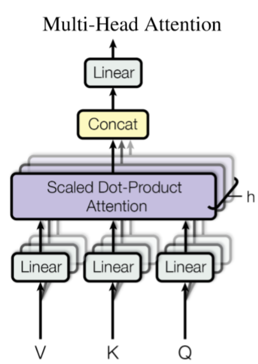



Figure 2.3: The Transformer Multi-Head Attention consists of several attention layers running in parallel.

The attention mechanism acts as shown in Equation 2.3, the attention of word $w_i$ in the context is just the weighted sum of sequence $w_1, ...w_n$ where the weight is positively related with the dot product of $w_i$ and $w_j$. To have a look in detail for the dot product procudure, see Figure 2.4.

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_K}}\right)V \tag{2.3}$$

$$MultiHead(Q,K,V) = Concat(head_1, head_2, ..., head_h)W^O \tag{2.4}$$

$$\text{where } head_i = Attention(QW_i^Q, KW_i^K, VW_i^V) \tag{2.5}$$

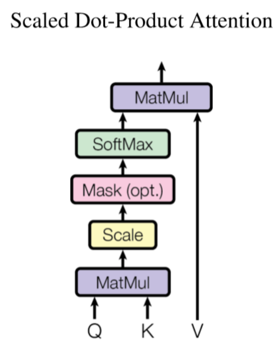



Figure 2.4: Scaled Dot-Product Attention.

The researchers of Transformer found it beneficial to concatenate outputs from different, learned linear projections in the dot product attention. So they apply several attention simultaneously to the same input pair with different linear projections. The number of parallel attention procedures is called *head*.

### Position-wise Feed-forward Network

In the encoder layers, Transformer also employs a fully-connected network which is applied to each position with the Equation 2.6.

$$FFN(x) = max(0, xW_1 + b_1)W_2 + b_2 \tag{2.6}$$

#### Residual Connections & Layer Normalization

To keep loss not increasing as the layer number is increased, Transformer uses residual connection between the input of sublayers (multi-head attention and position-wise feed forward) and the output. After that, to normalize the distribution of data, a layer normalization is applied for each sublayer making the final output of a sublayer is:

$$LayerNorm(x + Sublayer(x)) \tag{2.7}$$

### 2.3.3 Deep-Speare

In the languages of English and Chinese, the most successful poetic sentence processing models are usually, if not mostly, based on classic poems which follow strict formats on both structure and meter. Yan et al. [2013] proposed a joint model consists of language model, pentameter and rhyme model for the generation of quatrains (4-lines). By jointly learning, their model does well in meter and rhyme but not as satisfied as human sonnets in aspects of readability and emotion. They employed crowdworker evaluation and expert judgement for the evaluation. They evaluate the generated lines by letting crowdworkers distinguish them from the human-written ones and make quantitative evaluations in meter, rhyme, readability and emotion by the expert. The result reveals the advantage of computational modeling and also the common problems faced by poetry generation.

# Chapter 3

# Methods and Experiments

## 3.1 Outline

We illustrate our model, which is denoted as GEM (glyph enriched model), in Figure 3.1 to give a quick perception on the structure. The following is the flow of our model from a verse line to the glyph enriched prediction of verb:

- Take a line from data;
- Parse the line into split words and fetch the subject-verb-object triplets using the parser of He [2014];
- Padding the lines to the lenght of 25 by "[PAD]";
- The verb is the target for the language model and the position of it is masked with "[MASK]", see Section 3.2.1 and 3.3.4;
- Pass the verse to the language model to genertate 300 candicates, with scores $score_l$, according to the semantic context, see Section 3.2.1;

∗ Run glyph model for each of the candidates and score them with $score_g$, see Section 3.2.2;

- Rerank the 300 candidates by the weighted average of language score $score_l$ and glyph score $score_g$, see Section 3.3.6;

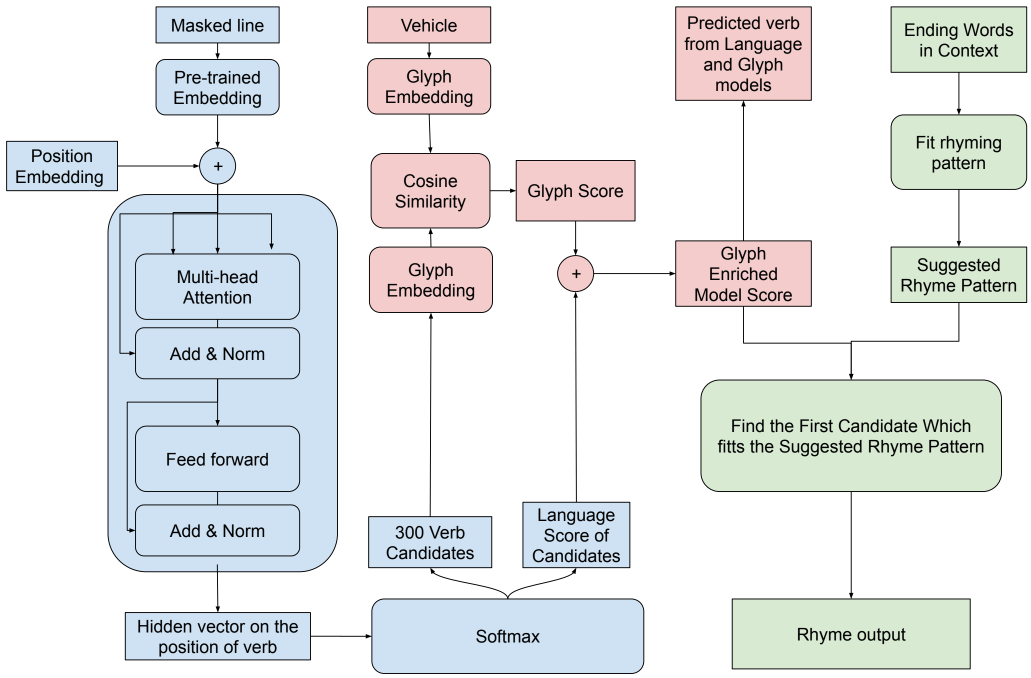



Figure 3.1: The structure of Glyph Enriched Model. The blue nodes compose the language model, the red ones build the rhyme model and the green nodes belong to rhyme model.

– If the verb is at the end of the verse, the ending words of nearby lines will be passed to the rhyme model which decides which candidates will be the prediction. If it's not the ending word, then we just output the top candidate from the reranked candidates, see Section 3.2.3.

## 3.2 Methods

The goal of our work is to generate verbs for modern Chinese verse lines to elaborate metaphors by passing the glyph information of vehicle to the verb. To achieve it, we composed our model with the following 3 components:

- Language model with a multi-layer Transformer encoder structure. The encoder is based on the works of Vaswani et al. [2017] with input of pre-trained word embedding via Li et al. [2018] and position embedding;

- Glyph model designed to capture glyph information of Chinese characters in the poem lines and weight each sample given by the language model;
- Rhyme model to model the patterns of end-rhymes of the poetry. In the generation procedure, if the predicted word is at the end of the line, the candidates from the two models above will be filtered by the rhyme pattern of the context.

### 3.2.1 Language Model

Our language model borrows the idea of Devlin et al. [2018] for training. Instead of masking words randomly, we only mask the verbs of the subject-verb-object triplets in the verses. The structure of our language model is the same with the multi-layer encoder of Vaswani et al. [2017]. We apply KL-divergence for the following loss function:

$$Loss = D_{KL}(P||T) = -\sum_{x \in X} P(x) \log\left(\frac{T(x)}{P(x)}\right) \tag{3.1}$$

where $P$ $(size = size_{vocabulary})$ is the output, normalized by softmax, of the encoder layers on the masked position, $T$ is the target distribution, which is the one-hot vector with "1" on the position of target by default. As suggested by Vaswani et al. [2017], to improve the accuracy, we employed label smoothing by set the target of $T$ to 0.9 and other positions $\frac{0.1}{size_{vocabulary}-1}$.

To accelerate the process of convergence, we accumulated the loss of mini-batches to achieve a larger batch size according to the memory of our GPU.

### 3.2.2 Glyph Model

The basic idea is to enrich the verb with more attributes from the vehicle to elaborate a metaphor. To approach this, we build up a glypy-emphasized embedding for the radicals.

A typical access to the embeddings is using Continuous Bag of Words Model (CBOW) or Skip-gram Model (SkipGram). But if we only want to emphasize the information of character glyph, we can simply get the embeddings of radicals and components via averaging the embeddings of words which contain the radical or component weighted by the frequency in our data.

To get a score based on the glyph information contained in the verb candidates, we apply the cosine similarity between the candidates and the vehicle through which, the attributes of the vehicle will be ascribed to the verb.

### 3.2.3 Rhyme Model

Similar with Lau et al. [2018], we only apply this model when the verb is at the end of the line. We developed the rhyme table from *Newly Edited Poetry Rhymes*[1], which provides the rhyme patterns of final vowels in three tones: even tone (like ī, í), oblique tone (like ǐ, ì) and entering tone (characters ending in voiceless stops) each of which consists of a few vowel patterns. Since the rules on modern Chinese poetry are not as strict as the traditional ones, we merge the three tone categories into 18 vowel patterns, see Table 3.1.

[1] http://www.ccview.net/theory/syxb.htm

| Rhyme Pattern | Example | Rhyme Pattern | Example |
|---|---|---|---|
| a, ia, ua | 巴, 夏, 瓜 | u | 顾 |
| o, uo | 菠, 过 | ü | 鱼 |
| e | 歌 | ou, iu | 周, 求 |
| ie, (y)e | 贴, 也 | ao, iao | 号, 表 |
| zi, ci, si | 子, 词, 思 | an, ian, uan | 站, 天, 团 |
| er | 儿 | en, in, un | 肯, 拼, 困 |
| i | 七 | ang, iang, uang | 唐, 量, 黄 |
| ei, ui | 为, 会 | eng, ing | 坑, 请 |
| ai, uai | 外, 坏 | ong, iong | 鸿, 琼 |

Table 3.1: 18 rhyme patterns concatenated from *Newly Edited Poetry Rhymes*.

## 3.3 Experiments

### 3.3.1 Word Embedding

We use the pre-trained Chinese n-gram embedding of Li et al. [2018] which trained the embedding independently on different kinds of corpus. The embedding dictionary we select is trained on solely on Chinese literature.

### 3.3.2 Glyph Embedding

Semantic radicals are the functional components, compared with the phonetic ones, for compiling the meaning of Chinese characters. The first try of glyph information embedding was based on the radical-character pairs crawled from *online Xinhua Dictionary*[2]. It provides 218 radicals which commonly contributes primarily to the semantic meaning of the characters. However, along the experiments, we found that some of the radical meaning are not used nowadays and also, according to the sophisticated hierarchical structure of Chinese characters, radicals, some of which are reltively small, like 一, 丨 and 丿, compared with the whole word, like the radicals in Figure 2.1, the expression of those radicals are

[2] http://xh.5156edu.com

generalized.

Besides, there's intersection between semantic radicals and components, like the same component 马(horse) of 妈(mother) and 驹(pony) in Table 3.2.

| Character | 妈(mother) | 驹(foal) |
|---|---|---|
| Pinyin | ma | ju |
| Common Component | 马(meaning:horse, Pinyin:ma) | |
| Radical Function | phonetic | semantic |

Table 3.2: The two characters 妈and 驹share the same component 马which acts as phonetic radical and semantic radical respectively.

To improve the expression of glyph embedding for single words, we crawled the $character-component_1-component_2$ triplets from *Wikimedia*[3] where the Chinese characters are separated into at most two large components. As stated by Williams [2010], the upper components of the character are more critical than lower components, and the left components are more so than right ones. Since the $components_1$ is the left or upper part of the character, we built a selective component list, containing 352 components, which are selected from the $component_1$s if:

$$\{component_1, component_2\} \cap \{radical\} = \emptyset \tag{3.2}$$

The way we select the semantic components not only results from the distribution of meaningful particles according to the statements of Williams [2010], but also from the fact that some of the radicals appears on the lower or right side of the character, including 鸟(bird), 灬(fire) and 皿(vessel).

After that we get two embeddings coming from radicals and selective components both of which are derived from the origin pre-trained word embeddings. We get the glyph

[3]https://commons.wikimedia.org/wiki/User:Artsakenos/CCD-TSV

embeddings by averaging those two embeddings for each word.

### 3.3.3 Data

Since the modern poems are easy to access and almost no barrier for composing, only part of the sources have high quality poems.

Our data is collected from two resources: 1.Gallery of Modern Chinese Poetry sgd (1920s-2000s). It is the largest online corpus of modern Chinese poetry which consists of 5174 poems from 519 outstanding poets. 2.Daily Poems of Chinese Poetry zgs (mostly 2010-2018). It is a collection of selected poems organized by a few famous poets and it consists of 4388 poems from 1137 poets and the contents are updating frequently now. After cleaning and formating data, we got 53,765 valid verse lines totally. The corpus is splited into training set (80%, 43,012 lines) and evaluation set (20%, 10,753) randomly.

According to the basic idea that borrow the glyph properties of vehicle to the verb, a valid verse line should have the object (vehicle) for the verb. But to make full use of data and address the situations like 孵化(hatch) and 卵石(cobble) in Table 2.3, we attempted to parsed the subjects in these verse lines into two nouns. The strategy of composing Chinese nouns is similar with the one for English nouns, if a phrase consists of two or more nouns, the main noun will be the last one while others act as adjective. Just like 卵石(cobble) whose meaning is composed by concatenate 卵(egg) and 石(stone) leading to the meaning of *the stone like an egg*. So in those lines without objects, during experiments, we employ the first noun separated from the subject as the vehicle.

### 3.3.4 Preprocessing

Our data is segmented and POS tagged using HanLP developed via He [2014], and also, to get the subject-verb-object triplet, the dependency parser of this work is employed. The verbs are the root of metaphors we're focusing on in the model. There are ten types of verbs among which 5 of them are applicable to our model, see Table 3.3, while others are not, see Table 3.4.

After that, we clear the data and format them by padding them with "[PAD]" to the same length (25). Since the verb is what we want to predict, we masked the verb with "[MASK]" during training and evaluation.

| Tag | Description | Example |
|---|---|---|
| vshi | 是(is) working as verb | 是 |
| | | 小明 **是** 一 只 鸟 。<br>Xiaoming (a name) **is** a qualifier bird .<br>Xiaoming is a bird. |
| vyou | 有(have) working as verb | 有 |
| | | 我 **有** 一 本 书 。<br>I **have** one qualifier book .<br>I have a book. |
| v | Other verbs | 飘, 笑, 保护, 喜欢, 看上去 |
| | | 他 **喜欢** 那个 女孩 。<br>he **like** that girl .<br>He likes that girl. |
| vx | Dummy verbs | 进行(+verb) |
| | | 警方 **进行** 调查 。<br>police **process** investigate .<br>Police are investigating. |
| vd | Verb-making adverbial modifier | 执意(+verb), 讽刺(+verb), |
| | | 他 不顾 劝告 **执意** 闹事 。<br>he not care about advise **insist on** make trouble .<br>He insisted on making trouble without hearing about the advice. |

Table 3.3: Types of verb in HanLP that are applicable to our model.

| Tag | Description | Example |
|---|---|---|
| vl | Verb idioms | 相依为命, 远走高飞 |
| | | 他们 几 人 **相依为命** 。<br>they several people **bound to each other** .<br>They are bound to each other. |
| vn | Verb-making nouns | (verb+)劝告 |
| | | 他 没 考虑 我 的 **劝告** 。<br>he not consider I 's, of **advice** .<br>He didn't consider my advice. |
| vf | Directional verbs | 上, 下, 进, 出, 回, 过, 起, 开, 来, 去 |
| | | 蜻蜓 飞 **进来** 。<br>dragonfly fly inside .<br>The dragonfly is flying inside. |
| vg | Verb-making morphemes | 庆(+verb) |
| | | 球队 举行 了 **庆** 凯旋 活动 。<br>team hold auxiliary word **celebrate** victory activity .<br>The team held a celebration. |
| vi | Intransitive verbs | 发言, 作响 |
| | | 他 昨晚 **发言** 了 。<br>he last night **make a statement** modal particle .<br>He made a statement last night. |

Table 3.4: Types of verb in HanLP that are not applicable to our model.

A simple way to show the reality of data is illustrating the length distribution of verse lines, which is shown in Figure 3.2. The distribution fits $N(\mu = 7.4984, \sigma = 2.3852)$ well showing that the collected data is comprehensive and representative.

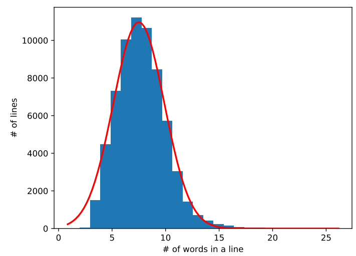



Figure 3.2: Distribution of the length of 65,599 lines (including the lines which is not valid for our model) with $\mu = 7.4984$, $\sigma = 2.3852$ out of totally 254,530 verse lines.

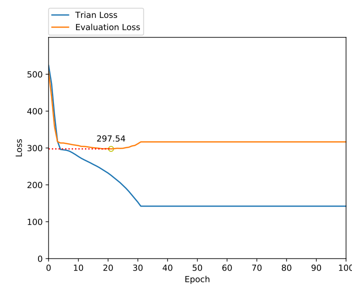



Figure 3.3: $batch\_size = 50 \times 10$, $hidden\_size = 300$

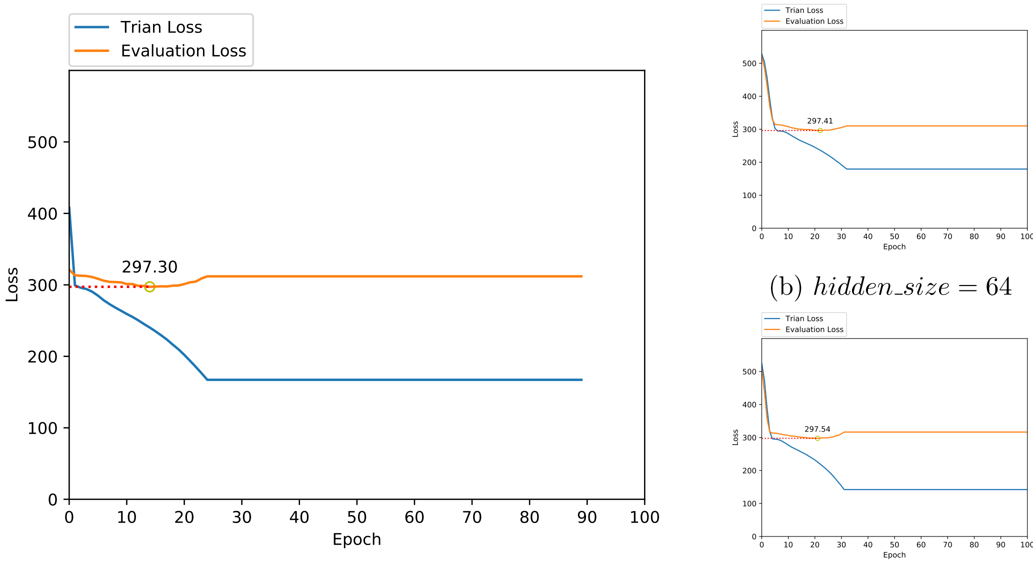
### 3.3.5 Training Details

According to the low memory of our GPU, which can only process 50 lines of data in a single time, we increased the total batch size by employing accumulated gradients.

Because of the small size of data, our language model, see Figure 3.3, got over-fitting by using a batch size of 500 while hidden size is 300. Reducing the number of parameters is a good way to compare and improve the performance of language model. To approach a considerable good language model, We designed several comparisons on hidden size, layer number of the encoder and batch size.

We selected 3 hidden size for the comparison, see Figure 3.4. According to the results, we can find that we do reduced the loss of evaluation by decreasing hidden size from 300 to 128, but when we select hidden size as 64, it increases again.

For the comparison on layer number, we reduced the number from 6 to 3 but only find the increase of evaluation loss, see Figure 3.5.




(a) $hidden\_size = 128$

(b) $hidden\_size = 64$

(c) $hidden\_size = 300$

Figure 3.4: Comparison on hidden size.

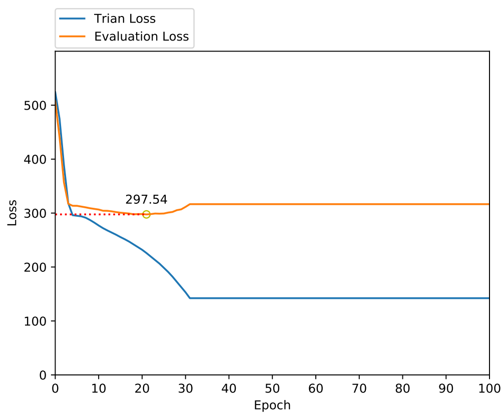
To increase the converging speed we also increased the total batch size by accumulating the gradients. Considering the expected running time to a similar accuracy, we set the accumulating times by doing a comparison in Figure 3.6. Simultaneously, in this step, we set our language model with the parameters in Table 3.5.

| Mini-batch | Gradient Accumulating Times | Hidden Size | Layer Number |
|---|---|---|---|
| 50 | 50 | 128 | 6 |

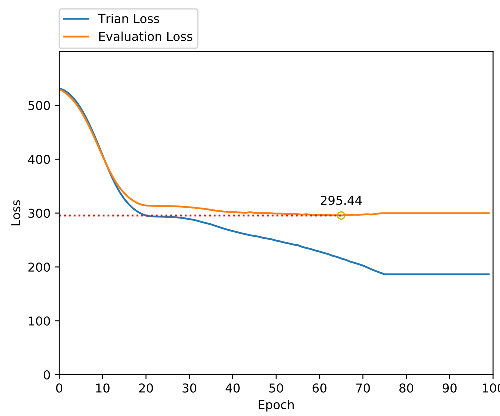
Table 3.5: Language model parameters.

### 3.3.6 Generation

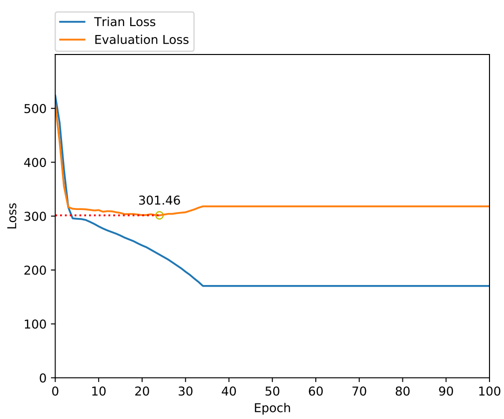
In this step, the language model passes 300 candidates with scores, $score_l$, normalized by softmax to the glyph model. To transform the glyph information from the vehicle to the verb, we calculated the cosine similarity between the vehicle and the candidates of the



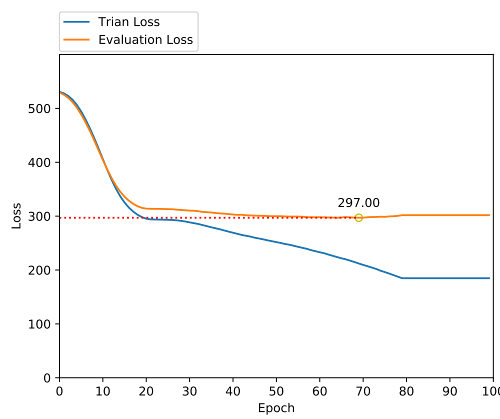

(a) $layer\_num = 6, batch\_size = 50 \times 10$




(c) $layer\_num = 6, batch\_size = 50 \times 50$

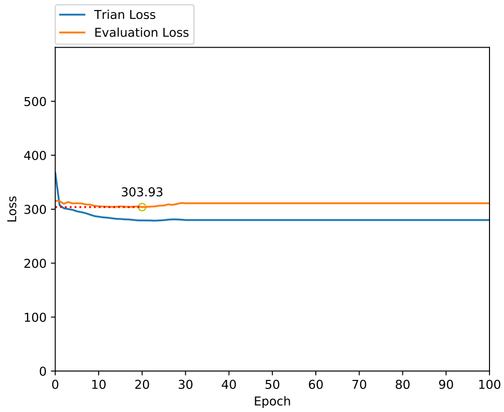



(b) $layer\_num = 3, batch\_size = 50 \times 10$

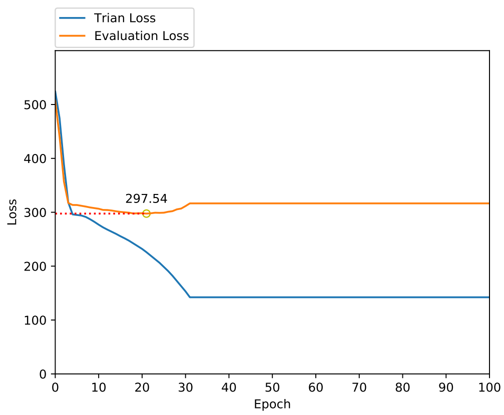



(d) $layer\_num = 3, batch\_size = 50 \times 50$

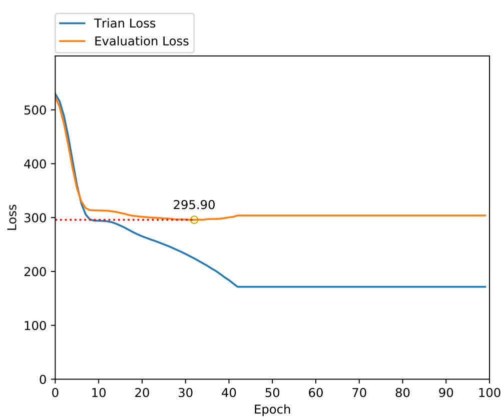
Figure 3.5: Comparison on layer numbers of the encoder.



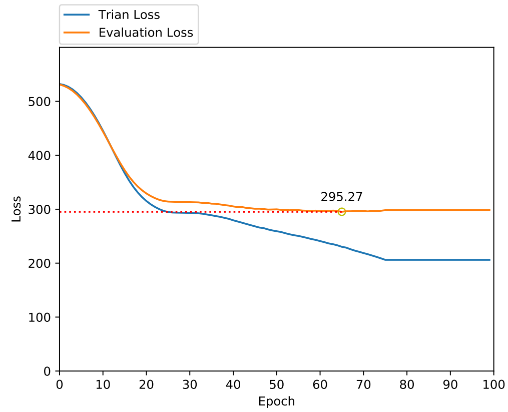

(a) hidden size: 300, batch size: $50 \times 1$



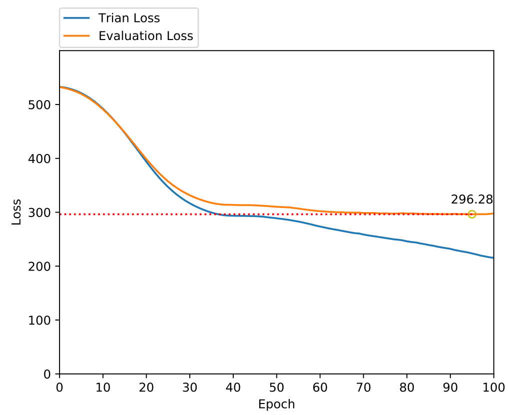

(b) hidden size: 300, batch size: $50 \times 10$



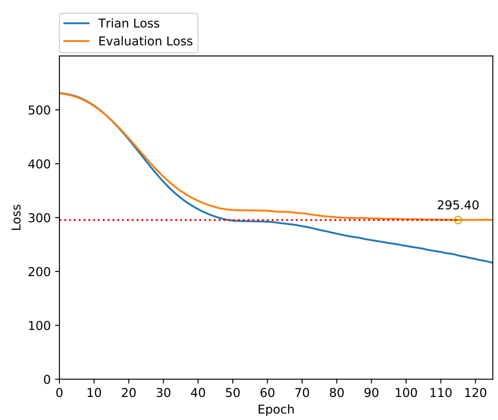

(c) hidden size: 300, batch size: $50 \times 20$




(d) hidden size: 128, batch size: $50 \times 50$




(e) hidden size: 128, batch size: $50 \times 75$




(f) hidden size: 128, batch size: $50 \times 100$

Figure 3.6: Comparison on batch size with (a), (b), (c) with hidden size 300 while (e), (f), (g) with hidden size 128. Batch size here is the accumulated batch size, for instance, $50 \times 100$ means accumulating 100 folds of gradients from the training on 50-size batches.

verb. After that, we apply softmax to it to get a normalized glyph score, $score_g$. We've tried four ways to calculate the score, $score_{GEM}$ where GEM represents Glyph Enriched Model, employed directly by the generator:

$$score_{GEM} = score_l \tag{3.3}$$

$$score_{GEM} = score_g \tag{3.4}$$

$$score_{GEM} = \alpha score_l + \beta score_g \tag{3.5}$$

$$score_{GEM} = score_l * score_g \tag{3.6}$$

Through experiments, we found that Equation 3.3 fits the context well semantically and the readability is considering good, but obviously the lack of glyph information makes it hard to compose a metaphor. Since the glyh embeddings derive from the pre-trained word embeddings, the predictions only applying the glyph model on the filtered candidates from language model make sense to some extent but the readability is pretty low. By comparing Equation 3.5, 3.6, we found that the latter one performs very similarly with the predictions only using the score from language model. This phenomenon is explainable: the glyph embeddings derive from the pre-trained word embeddings by a weighted average procedure from a large number of different words making the glyph embeddings generalized and by reviewing the scores coming from the glyph model proved that.

Then the generator will output the prediction according to the $score_{GEM}$. However, if the verb is at the end of the verse line, the candidates will be passed to the rhyme model sorted by $score_{GEM}$. The rhyme model will try to fit the rhyming patterning appearing in the ending words in the window of 5 lines.

**Demos**

We've illustrated 6 outputs, see the other 5 in Section 5.1, from evaluation dataset as demos to show the validity of our model.

| Chinese | 人间 | **有/ 怜悯** | 悲情 | 降生 |
|---|---|---|---|---|
| English | the world | **there is / show mercy to** | Sadness | be born |
| Origin | There is the birth of sadness in the world. | | | |
| GEM | The world shows its mercy to the sadness. | | | |

Table 3.6: tenor: 人间(the world), vehicle:悲情(sadness), verb: 有(there is)/怜悯(show mercy to). 怜悯(show mercy to) is the prediction.

For the instance in Table 3.6, it is a typical case in the corpus whose original verb is 有(have, there is) which only presents the meaning of *possess* or *existence*. However, this common but pale word could be replaced with a more expressive one. In the instance above, the model employs the properties of 悲情(sadness) by emphasize the radical and meaningful components of it and scores candidates accordingly. The selected verb 怜悯(show mercy to) shares two 忄(heart, emotion) with the vehicle 悲情(sadness) which contains the glyph information of 忄(heart) and 心(heart). In this way, the properties of vehicle, specifically the glyph information of 忄(heart, emotion) is passed to the tenor 人间(the world) which doesn't have the borrowed attributes leading the composition of the metaphor.

## 3.4 Evaluation

To understand the quality of our generated verb in the verse line, we asked a modern Chinese expert (an editor of China Literature Limited, with a master of arts degree majoring in modern Chinese literature) to rate in 5 aspects: metaphor, readability, emotion, rhyme and preference. The first four are rated in integer from 1 to 5 (1=worst, 5=best) and the preference is rated 0 or 1. To give a specific impression on different ratings，we illustrates the ratings in the Table 3.8 for the first four aspects. The preference shows the overall perception for the verbs with 1 meaning the verb performs better than the other and 0 is the opposite. However, since there are circumstances that the predicion is the same with the original verb or both machine-generated and human written verbs perform well, we'll score both of them with 1.

| **Model** | **Metaphor** | **Readability** | **Emotion** | **Rhyme** | **Preference** |
|---|---|---|---|---|---|
| GEM | $\mathbf{3.33 \pm 1.88}$ | $3.07 \pm 1.52$ | $\mathbf{1.92 \pm 1.26}$ | $\mathbf{4.60 \pm 1.20}$ | $\mathbf{0.30 \pm 0.46}$ |
| Human | $2.23 \pm 1.72$ | $\mathbf{4.67 \pm 0.78}$ | $1.72 \pm 1.07$ | $2.20 \pm 1.83$ | $0.87 \pm 0.34$ |

Table 3.7: Expert mean and standard deviation ratings on several aspects of the generated verbs. GEM represents the proposed Glyph Enriched Model. The scores in table follow the format of $mean_{scores} \pm std_{scores}$.

<table>
<tr><th>Score</th><th>Metaphor</th><th>Readability</th><th>Emotion</th><th>Rhyme</th></tr>
<tr><td>5</td><td>a typical metaphor</td><td>rational, smooth and no obstruct for understanding</td><td>direct and strong emotions, like love, miss and hate</td><td>the pronouciation of the predicted verb rhyme with one or more ending words of context lines</td></tr>
<tr><td>4</td><td>a dead metaphor</td><td rowspan="3">not perfect but make sense according to the three aspects</td><td>with strong emotion but not shown directly, like revenge</td><td rowspan="3">not applicative</td></tr>
<tr><td>3</td><td>simile</td><td rowspan="2">verbs with potential intention or negative attitude</td></tr>
<tr><td>2</td><td>others</td></tr>
<tr><td>1</td><td>not a metaphor</td><td>dosen't make sense</td><td>is, have or daily movements, like walk, see</td><td>the verb dosen't rhyme with other ending words of the context</td></tr>
</table>

Table 3.8: Rating rules for evaluation on metaphor, readability, emotion and rhyme.

The evaluation is based on the comparison of 300 pairs of the original verse line and the predicted one with a random order.

The results are presented in Table 3.7. In the aspect of metaphor, our model *GEM* (Glyph Enriched Model) achieves a mean score of 3.33 while human verses get 2.23. Since our model balances between semantic meaning (language model) and the glyph information (glyph model), there are some verse lines where the mearsure of both language model and glyph model $score_{GEM}$ in Equation 3.5 decreases if we increase the $score_g$ while $score_l$ drops. There is another situation affects the result on metaphor: since we ascribe the attribute of objects to the subject via passing the glyph properties to the verb, if the glyph information already in the intersection of the embeddings of subject and object, the metaphor will not occur. By reviewing the evaluation data, we find that the enriched glyph information tend to contribute more on the emotion expression leading a little more score in the evaluation on emotion. Our model also get better score on rhyming but fails on readability and preference which shows the fact that machine-generated poems still have shortage on the readibility.

# Chapter 4
# Conclusion

## 4.1 Conclusion

We proposed a glyph-enriched verb prediction model composed of languge, glyph and rhyme model. We explore a way of generating metaphor whithout building a knowledge base concists of attributes. Our model dose better than human verse lines on metaphor, rhyme and slightly in the term of emotion. However, the evaluation score on metaphor (3.33/5) shows that the passing of glyph information of object not always result in a metaphor. Machine-generated entities still underperform in the aspect of readability. Although the goal of ascribing glyph information to the verb is to get a creative generation, there are many cases in the output that the elaborated verb changes the original meaning of the verse line. In those cases, obviously, there is no verb near the position we calculated with the balanced scores of language and glyph models in embedding space.

## 4.2 Future Work

The glyph attributes in the characters are limited, the mainstream of metaphor generation is still the ways using knowledge base. But we can enrich the existing knowledge base with the semantic information contained in radicals and components. Besides, in the process of embedding radicals and components, we found that the values in the embedding vector are generalized by the weighted average from the origin embeddings which reduces the expression on the difference of words. That leads us to think about selecting more representative radicals or components to avoid the offsets from meaningless radicals or components.

# Chapter 5

# Appendix

## 5.1 Demos

| Chinese | 如果 | 镜子 | **出现/ 装满** | 裂痕 |
|---|---|---|---|---|
| English | if | mirror | **appear / be filled with** | crack |
| Origin | If the cracks appear in the mirror. | | | |
| GEM | If the mirror is filled with cracks. | | | |

Table 5.1: tenor: 镜子(mirror), vehicle:裂痕(cracks), verb: 出现(appears)/装满(be filled with). 装满(be filled with) is the prediction. In the glyph-enriched verse line, the vehicle and verb share the same radical 衣(cloth) which enrich the tenor with the property of holding things.

| Chinese | **藏/ 灌醉** | 了 | 一瓶 | 酒 |
|---|---|---|---|---|
| English | **hide / make sb. drunk** | done | a bottle of | wine |
| Origin | (I) hided a bottle of wine. | | | |
| GEM | (I) made the wine drunk. | | | |

Table 5.2: tenor: not shown, vehicle:酒(wine), verb: 藏(hide)/灌醉(make sb. drunk). 灌醉(make sb. drunk) is the the prediction.

| Chinese | 我 | **遇到/ 点燃** | 了 | 灿烂 | 姹紫 | 和 | 嫣红 |
|---|---|---|---|---|---|---|---|
| English | I | **meet / light** | done | bright, glorious | colorful | and | colorful |
| Origin | I've met the glorious things and the colorful world. | | | | | | |
| GEM | I've lighted them and made a glorious and colorful world. | | | | | | |

Table 5.3: tenor: 我(I), vehicle: 灿烂(bright and glorious), verb: 遇到(meet) / 点燃(light). 点燃(light) is the the prediction.

| Chinese | 犀利 | 的 | 目光 | **鸟瞰/ 堆满** | 大地 |
| --- | --- | --- | --- | --- | --- |
| English | sharp | particle | sight | **bird's-eye view / stack** | ground |
| Origin | The sharp sight look down to the ground from above. | | | | |
| GEM | The sharp sight is piled up on the ground. | | | | |

Table 5.4: tenor: 目光(sight), vehicle: 大地(ground), verb: 鸟瞰(bird's-eye view)/ 堆满(stack). 堆满(stack) is the the prediction.

| Chinese | 绵厚 | 的 | 水 | **是/ 惊醒** | 岁月 | 酿 | 的 | 酒 |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| English | mellow | particle | water | **is / awaken** | time and tide | brew | particle | wine |
| Origin | The mellow water is the wine brewed by time. | | | | | | | |
| GEM | The mellow water awakens the wine brewed by time. | | | | | | | |

Table 5.5: tenor: 水(water), vehicle: 酒(wine), verb: 是(is) / 惊醒(awaken). 惊醒(awaken) is the the prediction.

## 5.2 Partial Radicals

| Radical | Description | Examples |
| --- | --- | --- |
| 页 | related to head | 颊(cheek), 顺(be in the same direction, obey) |
| 石 | stone | 砚(inkstone), 研(grind) |
| 山 | mountain | 岳(high mountain), 屿(mound, island) |
| 老 | old | 耄(80-90 years old) |
| 八 | hand, reverse | 兵(soldier), 公(public) |
| 彡 | figure | 影(shadow), 形(shape) |
| 目 | eye | 盯(stare), 眩(have dim eyesight, dizzy) |
| 鱼 | fish | 鲑(salmon), 鲜(fresh) |
| 亠 | top, head | 京(capital), 亢(high) |
| 皮 | skin, scarfskin | 皱(wrinkle), 皲(skin chaps) |
| 支/攵 | hold up, hit | 教(teach) |
| 大 | big | 夸(boast), 套(cover) |

Table 5.6: Radicals Table Extracted from *online Xinhua Dictionary* via Institute of Linguistics [2019], Part 1.

| Radical | Description | Examples |
|---|---|---|
| 戈 | weapon | 战(fighting), 戳(poke) |
| 人/亻 | human | 令(order), 你(you) |
| 又 | plural | 双(double), 叠(fold) |
| 肉/月 | meat, body | 腐(spoil), 肚(tummy) |
| 日 | time | 旦(daybreak), 时(time) |
| 白 | dawn, white, shining | 皂(black, soap), 皇(emperor) |
| 走/辶 | walk | 迈(step forward) |
| 心/忄 | heart | 念(miss), 情(emotion) |
| 子 | child, juniors | 孕(pregnancy), 孝(filial) |
| 一 | one, line | 上(up), 专(concentrate) |
| 口 | mouth | 叨(gabby), 可(agree) |
| 风 | wind | 飒(cool, the sound of wind), 飘(flutter) |
| 竹 | bamboo | 竿(pole), 笑(laugh) |
| 士 | admirable people | 壮(strong), 声(sound) |
| 丿 | line, wall | 乏(tired), 丹(elixir) |
| 阝 | human | 降(down, surrender), 郭(town) |
| 丶 | possess, self | 主(important, lord), 义(principle, meaning) |
| 母 | mother, production | 每(each, grass growing), (mother) |
| 言/讠 | speak | 誉(reputation, fame), 认(recognize) |
| 巳 | one of Chinese era, snake | 巷(tortuous road), 巴(tail, hope) |
| 火/灬 | fire | 燃(firing), 照(shine, illuminate) |
| 手/扌 | hand | 承(bear, inherit), 提(lift) |
| 穴 | cave, hollow, empty | 穷(poor), 窥(peep through peepholes) |
| 刀/刂 | knife | 争(fight for), 剥(peel) |
| 厂 | shelter, building, cliff | 厄(misfortune), 厌(dislike, press) |
| 木 | wood, tree | 棒(stick), 杨(aspen) |
| 衣/衤 | cloth | 裹(wrap), 补(make up) |
|  | food, eat | 飧(serve with food and drink), 饮(drink) |
| 牛 | cattle | 牵(lead), 牧(herd) |
| 乙 | one of Chinese era, curve | 乞(beg), 乱(in a mass) |
| 土 | soil | 城(city, town), 坍(collapse of the building) |
| 夂 | stop | 冬(winter), 处(at) |
| 鼠 | mouse | 鼬(weasel), 鼢(sokhor) |

Table 5.7: Radicals Extracted from *online Xinhua Dictionary* via Institute of Linguistics [2019]. Part 2.

| Radical | Description | Examples |
|---|---|---|
| 欠 | yawn, owe, lack, mouth open | 欺(bully), 欢(happy) |
| 凵 | mouth open, concave | 凹(concave) |
| 文 | manners, ceremony | 斋(abstain from meat, wine) |
| 广 | wallless house, wide | 庆(celebrate), 庄(village) |
| 彳 | walk | 徜(stroll), 径(pass) |
| 十 | ten, cross, all directions | 千(thousand), 博(erudite) |
| 弓 | bow, bend | 弧(arc), 强(tough, forceful) |
| 宀 | cover, house | 守(protect) |
| 丨 | stick, passthrough | 中(center), 丫(tree twig) |
| 玉/王 | jade | 玺(imperial jade seal) |
| 米 | grain | 粮(food), 粟(millet) |
| 青 | blue, green, black, young | 靓(colorful, buautiful) |
| 干 | shield | 并(parallel), 平(flat) |
| 艹 | grass | 苞(bud) |
| 示/礻 | memory, ceremony | 祭(obit), 礼(manners) |
| 水/氵 | water | 汞(mercury), 池(pool) |
| 方 | tetragonum, square | 旗(flag), 旁(side) |
| 雨 | rain | 雪(snow), 雷(thunder) |
| 田 | field | 畴(farmland), 疆(territory) |
| 彐 | snow | 归(return), 彗(comet) |
| 几 | small table, frame | 凭(lean on), (residence) |
| 鸟 | bird | 鸡(chicken), 鸣(tweet) |
| 斗 | measure for measuring volume of grain | 料(count, estimate) |
| 血 | blood | 衅(sacrifice, conflict) |
| 面 | face | 靥(dimple), (wound on face) |
| 女 | female | 妈(mother), 妩(lovely) |
| 冂 | suburb, frame | 网(net), 冈(ridge of hill) |
| 儿 | child, human | 兄(elder brother) |
| 金/钅 | gold，metal | 鉴(authenticate), 铲(shovel) |
| 聿 | writing brush | 肄(study), 肃(quiet, strict) |
| 犬/犭 | dog, animal | 獒(mastiff), 犯(invade) |
| 生 | grow | 甥(nephew wife's sibling's son) |
| 车 | vehicle | 轧(roll), 辅(assist) |
| 夕 | sunset | 夜(night), 舞(dance) |
| 色 | color | 艳(colorful), (nattierblue) |
| 巾 | towel | 帆(sail), 帖(stick on) |

Table 5.8: Radicals Table Extracted from *online Xinhua Dictionary* via Institute of Linguistics [2019], Part 3.

# Bibliography

Xiang Zhang, Junbo Zhao, and Yann LeCun. Character-level convolutional networks for text classification. In *Advances in neural information processing systems*, pages 649–657, 2015.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*, 2014.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*, 2016.

Tony Veale and Yanfen Hao. Comprehending and generating apt metaphors: a web-driven, case-based approach to figurative language. In *AAAI*, volume 2007, pages 1471–1476, 2007.

Tony Veale. Round up the usual suspects: Knowledge-based metaphor generation. In *Proceedings of the Fourth Workshop on Metaphor in NLP*, pages 34–41, 2016.

Yanran Li, Wenjie Li, Fei Sun, and Sujian Li. Component-enhanced chinese character embeddings. *arXiv preprint arXiv:1508.06669*, 2015.

Yorick Wilks, John Barnden, and Jin Wang. Your metaphor or mine: belief ascription and metaphor interpretation. *Discourse and Meaning: Papers in Honor of Eva Hajicová, gives upturned circumflex*, pages 141–161, 1996.

Beata Beigman Klebanov, Ekaterina Shutova, and Patricia Lichtenstein. Proceedings of the fourth workshop on metaphor in nlp. In *Proceedings of the Fourth Workshop on Metaphor in NLP*, 2016.

Jenny Lederer. Finding metaphorical triggers through source (not target) domain lexicalization patterns. In *Proceedings of the Fourth Workshop on Metaphor in NLP*, pages 1–9, 2016.

Tzu-Ray Su and Hung-Yi Lee. Learning chinese word representations from glyphs of characters. *arXiv preprint arXiv:1708.04755*, 2017.

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 *IJCAI*, pages 2197–2203, 2013.

Yulia Tsvetkov, Leonid Boytsov, Anatole Gershman, Eric Nyberg, and Chris Dyer. Metaphor detection with cross-lingual model transfer. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 248–258, 2014.

Ekaterina Shutova and Simone Teufel. Metaphor corpus annotated for source-target domain mappings. In *LREC*, volume 2, pages 2–2, 2010.

Han He. HanLP: Han Language Processing, 2014. URL `https://github.com/hankcs/HanLP`.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017.

Shen Li, Zhe Zhao, Renfen Hu, Wensi Li, Tao Liu, and Xiaoyong Du. Analogical reasoning on chinese morphological and semantic relations. *arXiv preprint arXiv:1805.06504*, 2018.

Jey Han Lau, Trevor Cohn, Timothy Baldwin, Julian Brooke, and Adam Hammond. Deep-speare: A joint neural model of poetic language, meter and rhyme. *arXiv preprint arXiv:1807.03491*, 2018.

Clay Hunter Williams. Semantic vs. phonetic decoding strategies in non-native readers of chinese. 2010.

Gallery of modern chinese poetry. `http://www.shigeku.org/xlib/xd/sgdq/`. Accessed: 2018-09-28.

Daily poems of chinese poetry. `http://www.zgshige.com/mrhs/`. Accessed: 2018-09-28.

CASS Institute of Linguistics. Online xinhua dictionary. `http://xh.5156edu.com`, 2019. Accessed August 1, 2019.