國立雲林科技大學資訊管理系

碩士論文

Department of Information Management

National Yunlin University of Science & Technology

Master Thesis

中國唐宋五言絕句生成研究

Research on Quatrain Generation of 5-Character Quatrain of the Tang and Song Dynasties

鄭詣穎

Yi-Ying Cheng

指導教授：黃純敏　博士

Advisor: Chuen-Min Huang, Ph.D.

中華民國 110 年 6 月

June 2021

國立雲林科技大學
碩士班學位論文考試委員會審定書

本論文係 鄭詣穎 君在本校 資訊管理系 所提論文
「中國唐宋五言絕句生成研究」

合於碩士資格水準，業經本委員會評審認可，特此證明。

口　試　委　員　：　黃純敏　黃純敏
蔡榮婷
許中川

指　導　教　授　：　黃純敏　黃純敏

所　　　　長　：　莊煥銘

中　華　民　國　110　年　8　月　20　日

摘要

中國古典詩詞乃是重要文物遺產之一，其中以唐宋兩代蓬勃發展且嚴謹格律要求之律詩和絕句作為近體詩的代表。有鑑於(1)詩詞生成的研究多未詳細闡述平仄和押韻的評估標準，且多以人工檢驗平仄和詩韻，以致於影響詩詞生成之正確性，(2)現有研究將五言絕句和七言絕句合併訓練和生成，以致於在BLEU-2 評估中，五言絕句的表現明顯較七言絕句差，(3)現代人並非全然具備平仄格式及韻種的完整知識，以致於提升誤判絕句之可能。因此，本研究基於王力先生和陳淵泉先生的研究，歸納並列舉五言絕句之最為常見的三種平仄格式(含基本四式、變格和特殊格式，共 345 個)，並將此平仄格式做為資料集篩選以及生成型絕句之評估標準。本研究透過 CCPF 斷詞法將五言絕句中的各個句子切分成長度為一至三的詞組，再利用 TextRank 演算法找出各句之關鍵字，此關鍵字視為該詩句之含意代表，以訓練模型了解句意。此外，本研究為絕句的四個句子設計個別的專屬字典，負責儲存經由 CCPF 處理和 LDA 分群之詞組。於生成階段，絕句生成模型將根據四個由 LDA 歸屬於同一主題的字詞以生成五言絕句。本研究最終產出 660 首符合常見平仄格式之五言絕句，相較於先前研究提高近 30%的生成型絕句平仄正確率，且在 BLEU-2 和人工評估(不包含平仄和詩韻)上也有較佳表現。

關鍵字：自然語言處理、CCPF 斷詞法、平仄格式、詩詞生成

ABSTRACT

Chinese classical poetry is one of the important cultural heritage. The regulated verses and quatrains, which flourished in the Tang and Song dynasties, were representatives of Recent-style poetry which requires highly strict metrical rules. In view of (1) most researches of poetry generation did not elaborate on the evaluation criteria of tonal pattern and rhyme, and they normally tested them in human evaluation, so it may affect the correctness of the generated poetry, (2) previous researches used 5-character and 7-character quatrains together for training and poetry generation, it may result to lower performance of 5-character quatrains in BLEU-2 evaluation, and (3) modern people do not have the complete knowledge of tonal patterns and rhyme, so it may increase the possibility of misjudgment of quatrains. Therefore, this study organizes and lists the most common three-category tonal patterns (including the four basic tonal patterns, the variations based on the four basic tonal patterns, and the particular tonal patterns, totally 345) for five-character quatrain referring to the studies of Wang Li and Matthew Chen, and uses these tonal patterns as a standard for the selection of dataset and evaluation of generated quatrains. This research uses CCPF segmentation to separate each sentence in the five-character quatrains of the Tang and Song dynasties into words with one to three characters in length, and then uses the TextRank algorithm to find the keyword, which is regarded as the representative of the sentence, in order to train the model to understand the semantic meaning of sentences. In addition, we designed four separate dictionaries for the four consecutive sentences within a quatrain, which store the words processed by CCPF segmentation and clustered by LDA. In poetry generation phase, the generative model will generate five-character quatrains based on four words attributed to the same topic by LDA. This study eventually generated 660 five-character quatrains conforming to the tonal patterns. Compared with the previous researches, the accuracy of tonal patterns in the generated quatrains is improved approximately 30%, and the results in BLEU-2 and human evaluation (excluding tonal patterns and rhyme) also have better performances.

Keywords：Natural language processing, Classical Chinese Poetry Forms (CCPF Segmentation), Tonal patterns, Automatic poetry generation

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Research Background

Chinese classical poetry is an absolute beauty of Chinese classical literature. It recorded the prosperity and decline of a dynasty as well as the sorrow and joy of the populace. In other words, the feelings, experience, and opinions were conveyed by the poetries via ancient poets. It has been placed in the mouths of ancient people and on the lips of contemporary in the twenty-first century. The great influence of this cultural heritage on Chinese people has remained thousands of years.

The Tang Dynasty (A.D. 618~907) was a watershed in Chinese classical poetry – Chinese classical poetry had been categorized into Ancient-style poetry (古體詩 in Chinese) and Recent-style poetry (近體詩 in Chinese) (L. Wang, 1979). Ancient-style poetry has normally four to seven (rarely six) characters to each line regardless of total number of lines within a poetry. In contrast, Recent-style poetry only has either four or eight lines within a poetry, i.e. quatrain (絕句 in Chinese) and regulated verse (律詩 in Chinese), respectively, and the characters in each line are rigid with five or seven characters. Last but not least, Recent-style poetry must follow strict metrical rules, to wit: tonal patterns, which is the core of metrical rules, and rhyme schemes (L. Wang, 2014).

Quatrain is the most popular one among Recent-style poetry, which shares the same values with regulated verse (He, Zhou, & Jiang, 2012). A quatrain consists of four consecutive sentences, and the line length is fixed at 5 or 7 characters, i.e. 5-character quatrain (五言絕句 in Chinese) or 7-character quatrain (七言絕句 in Chinese). It must follow the aforementioned strict structures and complicated linguistic rules including tonal patterns, namely Level tone (平聲 in Chinese) and Oblique tone (仄聲 in Chinese), and rhyme scheme, which the ending characters in the even-numbered lines must be in rhyme, in addition to the basic requirements on grammars and semantics. "The quatrains of the Tang and Song dynasties" will hereafter be cited within the text as Chinese Quatrain.

Chinese Quatrain, which is a cultural heritage, is the bonds to the past, to the present, and to the future. For cultural inheritance, there are numerous researchers making the effort to analyze the metrical rules of it by reviewing information and literature manually and keeping completing metrical structures (Chen, 1979; Downer & Graham, 1963; L. Wang, 1989), and also devoting lots of energies to its simulation by refining various generative models (Huang, Lu, Cheng, & Peng, 2020; L. Liu, Wan, & Guo, 2018; Yi et al., 2017). Although recent researchers in poetry generation research field indicated that the research results of automatic poetry generation are getting better and follow the specific writing format of the Chinese Quatrain, but their samples were labeled with the wrong tones. Moreover, previous studies usually evaluated tonal pattern and rhyme scheme in human evaluation as one of 3 criteria of poetry generation proposed by H. Manurung (2004), i.e. grammaticality (the generated sentences must be readable and grammatically correct), meaningfulness (the sentences should convey something related to the theme), and poeticness (generated poems must have poetic features, e.g. harmonious rhythm and tonal patterns, etc.). However, modern people might equip insufficient know-hows towards rhythm categories and tonal patterns of Chinese classical poetry, which are totally different from the recent language expressions. These subjective opinions of the testers from various backgrounds would influence the credibility of evaluation result. Therefore, the populace might not be knowledgeable to do metrical rules evaluation, and it is unsuitable to test tonal patterns and rhymes in human evaluation.

Chinese is a tonal language, and each word only has one syllable. Because of enormous changes of Chinese tones, modern people have totally different language expressions and insufficient knowledge towards traditional Chinese, including Chinese tones, tonal patterns, and rhymes. The great differences of Chinese tones between traditional and Modern Chinese, e.g. Mandarin, would influence modern people to judge Chinese tonal patterns and rhyme scheme since the entering-tone was combined into the other tones. We are going to address some problems when evaluating Chinese tonal patterns and rhyme scheme with the example of famous 5-character quatrain – Yearning (〈相思〉 in Chinese, aka Xiang Si) written by Wang Wei (王維 in Chinese) of the Tang Dynasty, who is praised as “Poem Buddha” (詩佛 in Chinese). The example is demonstrated with Mandarin Phonetic Symbol (注音符號 in Chinese, aka Bopomofo, or Zhuyin), Hànyǔ Pīnyīn (漢語拼音 in Chinese), traditional Chinese tones (The level tone and oblique tone are symbolized here as “E” and “O”, respectively), Mandarin tones, rhyme, and translation as shown in **Fig. 1**.

Title	相思 王維(唐·五言絕句) Xiang Si by Wang Wei (5 –Character quatrain of the Tang Dynasty)				
1st Line	紅	豆	生	南	國
Mandarin Phonetic Symbol	ㄏㄨㄥˊ	ㄉㄡˋ	ㄕㄥ	ㄋㄢˊ	ㄍㄨㄛˊ
Hànyǔ Pīnyīn	hóng	dòu	shēng	nán	guó
Traditional Chinese Tone	E	O	E	E	O
Mandarin Tone	Tone 2	Tone 4	Tone 1	Tone 2	Tone 2
Rhyme	一東 (Shang Ping sheng)	二十六宥 (Departing tone)	八庚 (Xia Ping sheng)	十三覃 (Xia Ping sheng)	十三職 (Entering tone)
Translation	Red beans grow in the southern lands;				
2nd Line	春	來	發	幾	枝
Mandarin Phonetic Symbol	ㄔㄨㄣ	ㄌㄞˊ	ㄈㄚ	ㄐㄧˇ	ㄓ
Hànyǔ Pīnyīn	chūn	lái	fā	jǐ	zhī
Traditional Chinese Tone	E	E	O	O	E
Mandarin Tone	Tone 1	Tone 2	Tone 1	Tone 3	Tone 1
Rhyme	十一眞 (Shang Ping sheng)	十灰 (Shang Ping sheng)	六月 (Entering tone)	五尾 (Rising tone)	四支 (Shang Ping sheng)
Translation	when spring comes they grow a few branches;				
3rd Line	願	君	多	采	擷
Mandarin Phonetic Symbol	ㄩㄢˋ	ㄐㄩㄣ	ㄉㄨㄛ	ㄘㄞˇ	ㄒㄧㄝˊ
Hànyǔ Pīnyīn	yuàn	jūn	duō	cǎi	xié
Traditional Chinese Tone	O	E	E	O	O
Mandarin Tone	Tone 4	Tone 1	Tone 1	Tone 3	Tone 2
Rhyme	十四願 (Departing tone)	十二文 (Shang Ping sheng)	五歌 (Xia Ping sheng)	十賄 (Rising tone)	九屑 (Entering tone)
Translation	I hope you will gather a few more;				
4th Line	此	物	最	相	思
Mandarin Phonetic Symbol	ㄘˇ	ㄨˋ	ㄗㄨㄟˋ	ㄒㄧㄤ	ㄙ
Hànyǔ Pīnyīn	cǐ	wù	zuì	xiāng	sī
Traditional Chinese Tone	O	O	O	E	E
Mandarin Tone	Tone 3	Tone 4	Tone 4	Tone 1	Tone 1
Rhyme	四紙 (Rising tone)	五物 (Entering tone)	九泰 (Departing tone)	七陽 (Xia Ping sheng)	四支 (Shang Ping sheng)
Translation	these things are what one yearns for the most.				

Fig. 1 Example of 5-Character Quatrain of the Tang Dynasty –Xiang Si (Yearning) by Wang Wei

In traditional Chinese language, Chinese tones include level tone, and oblique tone. Level tone is subdivided into Shang Ping sheng (上平聲 in Chinese), Xia Ping sheng (下平聲 in Chinese), whereas oblique tone includes Rising tone (上聲 in Chinese), Departing tone (去聲 in Chinese) and Entering tone (入聲 in Chinese) (L. Wang, 2014). In contrast, Mandarin tones have mainly four tones and a light tone, which the words are pronounced lightly with a symbol as “ ˙ ”, e.g. “嗎” (ma •). Tone 1 (一聲 in Chinese, aka high level tone, which is merely marked with symbol as “ - “) and Tone 2 (二聲 in Chinese, aka rising tone, symbolized as “ ˊ “) represent level tone, Tone 3 (三聲 in Chinese, aka falling-rising tone, symbolized as “ ˇ “) covers part of Rising tone, and Tone 4 (四聲 in Chinese, aka falling tone, symbolized as “ ˋ “) covers the other part of Rising tone and entire Departing tone. In particular, Entering-tone was specially separated into the aforementioned 4 Mandarin tones (L. Wang, 2014). Table 1 shows the illustration of difference between traditional Chinese tones and Mandarin tones. However, it would result in modern people’s uncertain judgments on Chinese tone toward character, which is changed from entering tone to other four Mandarin tones. The words originally belong to entering-tone in traditional Chinese may be incorrectly identified as one of Mandarin Tones. For example, in **Fig. 1**, although the word” 國” is in the group of Tone 2, it actually belongs to entering-tone. Moreover, the word “發” is a level tone in Mandarin Tones, but in actual fact it is an oblique tone in traditional Chinese. Table 2 shows the illustration of aforementioned example. Therefore, modern people evaluate traditional Chinese tones and tonal patterns with difficulties. The solutions given by L. Wang (2014) are checking the dictionary or rhyme book (韻書 in Chinese).

Table 1

Illustration of Mandarin Tone and Traditional Chinese Tone

<table>
<tr><td>Mandarin Tone</td><td>Tone 1
(High Level Tone)</td><td>Tone 2
(Rising Tone)</td><td>Tone 3
(Falling-rising Tone)</td><td colspan="2">Tone 4
(Falling Tone)</td></tr>
<tr><td rowspan="2">Traditional Chinese Tone</td><td colspan="2">Level Tone</td><td colspan="2">Rising Tone</td><td>Departing Tone</td></tr>
<tr><td colspan="5">Entering Tone</td></tr>
</table>

Table 2

Examples of Entering Tone

Chinese Character	國	發
Mandarin Tone	Tone2	Tone1
Traditional Chinese Tone	Entering Tone	Entering Tone

The other specific feature of Chinese Quatrain is rhyme scheme. The ending characters in the second and fourth lines must rhyme with the same rhyme group (韻目 in Chinese, aka rhyme heading), and they are normally confined to the level tone. Extended Rhymes of the Song Dynasty embody 26,194 words, and these words were groups into 3,874 homophone groups (小韻 in Chinese), and 206 rhyme groups (BYVoid, 2011; "Chinese Text Project (中國哲學書電子化計劃 in Chinese)," 2006; "漢字古今音資料庫," 2011). Each word has at least one homophone group, and each homophone group has its corresponding rhyme groups. However, Taiwanese people judge rhymes in Chinese Quatrains according to the last Mandarin phonetic symbols. For instance, in **Fig. 1**, people are not able to make sure that if the word" 枝" and the word "思" belong to the same rhyme or not. In this case, people are used to using Taiwanese, which is a dialect in Taiwan, to translate Chinese Quatrain, and then get a result that these two words are actually in rhyme since the last letters of them are pronounced "i" in Taiwanese. Table 3 shows the example of rhyme judgement. In conclusion, the populace has confusion of rhymes. The solutions given by L. Wang (2014) are also checking the dictionary or rhyme book (韻書 in Chinese).

Table 3

Examples of Rhyme Judgement

Chinese Character	枝	思	In rhyme
Mandarin phonetic symbols	ㄓ	ㄙ	Not Sure
Taiwanese(dialect)	ki	si	Yes
Rhyme Group	四支	四支	Yes

Besides aforementioned problems, the extant quatrains' compliances with Chinese tonal patterns is a serious problem. In series of Wang, he recorded and indicated few irregular poetries that carry characters with incorrect tones in particular positions (L. Wang, 1979). Moreover, it is possible that the ending characters in even-numbered lines are oblique tones. These abnormal poetries would be one part of the samples for the training phase and automatic poetry generation, and afterwards probably affecting the generated quatrains to deviate from the metrical rules. Therefore, we are curious about the correct tonal patterns, and also the number of widespread quatrains which were compliant with the tonal patterns.

As we mention above, we conclude the existing problems of this study as below:

(1) The evaluations of tonal patterns and rhyme are inappropriate to be one of the criteria in human evaluation.

(2) The difference between traditional and modern Chinese Tones results in modern people's uncertain analysis on Chinese tonal patterns towards Chinese Quatrain.

(3) The populace is unable to analyze rhyme in Mandarin.

(4) Unknown number of irregular quatrains which do not follow the metrical rules might influence the quality of generated quatrains.

1.2 Research Purpose

The tonal patterns are seen as the core of metrical rules (L. Wang, 2014). Based on the tonal pattern, the eight-line regulated verse consists of a double quatrain, and the 5-character quatrains are treated as the 7-character quatrains with 2 prefix characters excised (Downer & Graham, 1963). In other words, a 5-character quatrain is the basic unit of Recent-style poetry, and the other genres of Recent-style poetry vary and extend their tonal patterns according to it.

The common tonal patterns of 5-character quatrains include four writing formats, i.e. the four basic tonal patterns, the variations based on the four basic tonal patterns, the particular tonal patterns, and counterbalance (拗救 in Chinese), which is a remedy for irregular misplaced tones (Downer & Graham, 1963; L. Wang, 2014). Among the four writing formats, the first three types of tonal patterns are the most common ones, i.e. the four basic tonal patterns, their variations, and the particular tonal patterns (L. Wang, 2014; Zhang & Lapata, 2014). Therefore, this study is going to employ the 5-character quatrains in the Tang and Song Dynasties, which acts in compliance with these most common tonal patterns, to generate quatrains automatically.

The major purposes of this study are as follows:

(1) Realizing the fitness of existing quatrains and the most common tonal patterns, i.e. the four basic tonal patterns, their variations, and the particular tonal patterns.

(2) Generating Chinese quatrains automatically to meet the requirements of the most common tonal patterns.

The study is organized as follows. We present review of literature in Chapter 2. In Chapter 3, we describe research model and discuss the research design and methodology. We show the result of this study and the evaluations in Chapter 4. Finally, our conclusion and future work are drawn in Chapter 5.

Chapter 2 Review of Literature

This research is going to review and briefly introduce Chinese Quatrain, Chinese Tones in Section 2.1 and 2.2. The literature of rhyme scheme and tonal patterns will be reviewed in section 2.3 and 2.4. In section 2.5, we are going to explore automatic poetry generation. The applied algorithm and models is described in section 2.6 to section 2.10.

2.1 Introduction of Chinese Quatrain

Chinese classical poetry had been categorized into Ancient-style poetry (古體詩 in Chinese) and Recent-style poetry (近體詩 in Chinese) because of Chinese imperial examinations (科舉考試 in Chinese) of the Tang Dynasty. Since then, the poetic style was highly formalized (L. Wang, 1989). Ancient-style poetry has normally four, five, and seven characters to each line regardless of total number of lines within a poetry. In contrast, Recent-style poetry only has four or eight lines within a poetry and the characters in each line are rigid with five or seven characters, i.e. quatrain (絕句 in Chinese) and regulated verse (律詩 in Chinese), respectively. Last but not least, it has to follow strict metrical rules, to wit: tonal patterns and rhyme schemes. (L. Wang, 2014).

Chinese Quatrain has a fixed number of characters in each line (usually either five or seven) throughout the poem and is obligatory for rhyme scheme in even-numbered lines. According to the rhyme scheme, it is subdivided into "ancient-style quatrain" (古絕 in Chinese) and "regulated quatrain" (律絕 in Chinese) (Lin, 2014; L. Wang, 1979). Ancient-style quatrain is an unlimited poetic style which allows the ending characters in the 2nd and 4th lines to be either level tone or oblique tone freely without the constraint of tonal patterns. In contrast, regulated quatrain must follow the tonal patterns, and the ending characters in even-numbered lines are definitely confined to the level tone and also belong to the same rhyme group (韻目 in Chinese).

2.2 Chinese Tones

Chinese is a tonal language. Traditional Chinese Tones include level tone (平聲 in Chinese), Rising tone (上聲 in Chinese), Departing tone (去聲 in Chinese), and Entering tone (入聲 in Chinese). The first one includes Shang Ping sheng and Xia Ping sheng (上平聲和下平聲 in Chinese), and the others are collectively referred to as oblique tone (仄聲 in Chinese) (Pulleyblank, 1999).

Chinese tones are significantly different from Mandarin Tones. Mandarin is said to have mainly four tones and a light tone. Tone 1 (一聲 in Chinese) is high and level, whereas Tone 2 (二聲 in Chinese) rises moderately. Compared the characters in the group of Tone 3 (三聲 in Chinese), which falls firstly and rises again later, the characters of Tone 4 (四聲 in Chinese) start out high but drops sharply. Furthermore, Mandarin tones represents different proportion of traditional Chinese tones. Tone 1 and Tone 2 represent level tone. Tone 3 is for part of characters in Rising tone, and Tone 4 covers Departing tone and the other part of Rising tone. Entering tone is particularly categorized into the aforementioned 4 Mandarin tones. Among them, Tone 4 has the most Entering tone, second is Tone 1 and Tone 2, and the one contains the least Entering tone is Tone3. (L. Wang, 2014) The illustration of tones in traditional Chinese and Mandarin is shown in Table 1.

2.3 Chinese Rhymes

Rhyme is one of the fundamental elements of poetry metrical rules. Every poet must follow not only rhyme scheme but also tonal patterns to compose a Recent-style poetry with rhyme book (韻書 in Chinese), which is a dictionary organized by rhymes. The earliest rhyme book is Spelling Rhymes (切韻 in Chinese), which arranged character according to pronunciation and originally has 193 rhyme groups. Spelling Rhymes is the predecessor of Rhyme dictionary from the Tang dynasty (唐韻 in Chinese). After repeated revisions and enlargements during the succeeding Tang Dynasty, it finally culminated with the Extended Rhymes (廣韻 in Chinese), which is the earliest extant rhyme dictionary containing 206 rhyme groups (Pulleyblank, 1999). However, these fine rhymes made a lot of constraints to compose a Recent-style poetry. In the early Tang Dynasty, the officials including Hsu Jing Zong (許敬宗 in Chinese) proposed to combine some adjacent rhyme groups for efficient poetic composition, so some rhyme groups became general-purpose and exchangeable (L. Wang, 1989, 2015). Although the official total number of rhyme groups was 206, but it was actually 112. In the Northern Song Dynasty, these 206 rhyme groups were condensed formally into 106 by Liu Yuan (劉淵 in Chinese). The rhyme book of this new version of condensed rhymes was *Newly Compiled Concise Rhymes of the Ministry of Rites in Renzi Year* (《壬子新刊禮部韻略》 in Chinese, aka Renzi xinkan Libu yunlüe). It is a dictionary with characters arranged phonetically according to a rhyme system, and it was renamed as "Pingshui Rhymes" (平水韻 or 佩文詩韻 in Chinese) in the Qing Dynasty (清朝 in Chinese). Therefore, Chinese poets in the Tang and Song Dynasty in fact applied Pingshui Rhymes to compose Recent-style poetry (L. Wang, 1979).

Chinese rhyme is highly related to Chinese tones. The rhymes of characters were arranged according to the traditional Chinese tones. The characters in Pingshui rhyme were categorized into 106 rhyme groups, and the numbers of rhyme groups in Shang Ping sheng, Xia Ping sheng, Rising tone, Departing tone, and Entering tone are 15, 15, 29, 30, and 17, respectively. Table 4 indicated the rhymes in the Pingshui Rhymes. In addition, a word would depend on its meaning or part of speech (詞性 in Chinese) to vary its tone and rhyme. For example, the word "思" (Si) is a Level tone with rhyme group "四支" when it is a verb to describe "miss someone or something (思念 in Chinese)", but it will becomes an Departing tone with rhyme group "四寘" when it is used as a noun to describe "thought or feelings" (思想或情懷 in Chinese). Therefore, the characters with different tones are not in the same rhyme, and likewise the ending characters are not in rhyme (L. Wang, 2014).

Table 4

Pingshui Rhyme

<table>
<tr><td rowspan="8">Level Tone</td><td rowspan="4">Shang Ping Sheng</td><td colspan="5">15 Rhyme Groups</td></tr>
<tr><td>一東</td><td>二冬</td><td>三江</td><td>四支</td><td>五微</td></tr>
<tr><td>六魚</td><td>七虞</td><td>八齊</td><td>九佳</td><td>十灰</td></tr>
<tr><td>十一眞</td><td>十二文</td><td>十三元</td><td>十四寒</td><td>十五刪</td></tr>
<tr><td rowspan="4">Xia Ping Sheng</td><td colspan="5">15 Rhyme Groups</td></tr>
<tr><td>一先</td><td>二蕭</td><td>三肴</td><td>四豪</td><td>五歌</td></tr>
<tr><td>六麻</td><td>七陽</td><td>八庚</td><td>九青</td><td>十蒸</td></tr>
<tr><td>十一尤</td><td>十二侵</td><td>十三覃</td><td>十四鹽</td><td>十五咸</td></tr>
<tr><td rowspan="19">Oblique Tone</td><td rowspan="7">Rising Tone</td><td colspan="5">29 Rhyme Groups</td></tr>
<tr><td>一董</td><td>二腫</td><td>三講</td><td>四紙</td><td>五尾</td></tr>
<tr><td>六語</td><td>七麌</td><td>八薺</td><td>九蟹</td><td>十賄</td></tr>
<tr><td>十一軫</td><td>十二吻</td><td>十三阮</td><td>十四旱</td><td>十五潸</td></tr>
<tr><td>十六銑</td><td>十七筱</td><td>十八巧</td><td>十九皓</td><td>二十哿</td></tr>
<tr><td>二十一馬</td><td>二十二養</td><td>二十三梗</td><td>二十四迥</td><td>二十五有</td></tr>
<tr><td>二十六寢</td><td>二十七感</td><td>二十八儉</td><td>二十九豏</td><td></td></tr>
<tr><td rowspan="7">Departing Tone</td><td colspan="5">30 Rhyme Groups</td></tr>
<tr><td>一送</td><td>二宋</td><td>三絳</td><td>四寘</td><td>五未</td></tr>
<tr><td>六御</td><td>七遇</td><td>八霽</td><td>九泰</td><td>十卦</td></tr>
<tr><td>十一隊</td><td>十二震</td><td>十三問</td><td>十四願</td><td>十五翰</td></tr>
<tr><td>十六諫</td><td>十七霰</td><td>十八嘯</td><td>十九效</td><td>二十號</td></tr>
<tr><td>二十一箇</td><td>二十二禡</td><td>二十三漾</td><td>二十四敬</td><td>二十五徑</td></tr>
<tr><td>二十六宥</td><td>二十七沁</td><td>二十八勘</td><td>二十九艷</td><td>三十陷</td></tr>
<tr><td rowspan="5">Entering Tone</td><td colspan="5">17 Rhyme Groups</td></tr>
<tr><td>一屋</td><td>二沃</td><td>三覺</td><td>四質</td><td>五物</td></tr>
<tr><td>六月</td><td>七曷</td><td>八黠</td><td>九屑</td><td>十藥</td></tr>
<tr><td>十一陌</td><td>十二錫</td><td>十三職</td><td>十四緝</td><td>十五合</td></tr>
<tr><td>十六葉</td><td>十七洽</td><td colspan="3"></td></tr>
</table>

2.4 Rhyme Scheme and Tonal Pattern

Chinese Quatrain (絕句 in Chinese) and regulated verse (a double quatrain, 律詩 in Chinese) were highly formalized during the Tang and Song dynasties and consist of 8 and 4 lines, respectively. Each of them has either 5 characters or 7 characters in each sentence and is obligatory following tonal patterns (a regular alternation of tone sequences not only within the line, but also from one line to the next) and rhyme scheme (rhyming of even-numbered lines necessarily, and the first line occasionally) (Chen, 1979; L. Wang, 2014).

Chinese tonal patterns consist of level tone for a steady-state of pitch and oblique tone for a rising and falling pitch. Since the 20th century, Chinese tonal patterns of Recent-style poetry have been analyzed in various way. Based on the tonal patterns, Regulated verse has been treated as a double quatrain. Moreover, the 7-character format has been seen 5-character format with two additional prefix characters excised, and this rule is a basis for the further analysis of tonal patterns.

L. Wang (1979) listed 4 basic sentence patterns, which alternately changes the tone of the first position within each sentence to create tonal pattern variety, as listed below:

(1) 仄仄平**平仄** (OOE**EO**)　　(2) 平平仄**仄平** (EEO**OE**)

(3) 平平平**仄仄** (EEE**OO**)　　(4) 仄仄仄**平平** (OOO**EE**)

The 4 basic tonal patterns, the variations (變格 in Chinese) of them, and the particular tonal patterns as shown in Table 5 and Table 6. The left column (Form-1A to Form-4A) showed the 4 basic tonal patterns with the symbol E and O to represent level tone and oblique tone respectively. E/*O* and O/*E* means that either E or O is allowed in this position, and the previous one is the original syllable, whereas the latter one is the alternative. Wang mentioned that the traditional rule of thumb "license for 1, 3 and 5, strictness for 2, 4 and 6" was not comprehensive. He took the combinations of two ending characters (EO, OO, EE. and OE) in each sentence to explain the interrelationship among the 1st, 3rd, and 5th characters, which produced the variations of aforementioned 4 basic tonal patterns, as shown from Form-1B to Form-4B in Table 5 and Table 6. Wang's explanation of variations is listed below in detail. The tonal patterns usually vary the tones in the third character in 5-character quatrain or the fifth character in 7-character quatrain according to the combination of two ending characters.

(1) Sentence ends with EO: The 3rd character in 5-character sentence or the 5th character in 7-sentence is allowed to change to an oblique tone, but the character at the same position in next consecutive sentence has to be changed into a level

tone.

(2) Sentence ends with OO: The 3rd character in 5-character sentence or the 5th character in 7-sentence is allowed to change to an oblique tone, but the 1st character in 5-character sentence or the 3rd character in 7-sentence has to be changed into a level tone.

(3) Sentence ends with EE: The 3rd character in 5-character sentence or the 5th character in 7-sentence is allowed to change to a level tone.

(4) Sentence ends with OE: The 3rd character in 5-character sentence or the 5th character in 7-sentence is allowed to change to an oblique tone.

In addition, Form-1C, Form-3C, and Form-4C are the particular tonal patterns for a sentence ending with OO that a level tone in the first position leads a sentence that the tones of the character at the 3rd and 4th position in 5-character sentence or at the 5th and 6th position in 7-character sentence are interchangeable.

Table 5

Tonal Pattern of 5-Character Quatrain by Wang Li

WANG LI 5-CHARACTER QUATRAIN - LEVEL START																	
Form-1A						*Form-1B*						*Form-1C*					
Character / Sentence	1	2	3	4	5	Character / Sentence	1	2	3	4	5	Character / Sentence	1	2	3	4	5
1	E/O	E	E	O	O	1	**E**	E	**O**	O	O	1	**E**	E	**O**	**E**	O
2	O/E	O	O	E	E	2	O/E	O	E	E	E	2	O/E	O	O	E	E
3	O/E	O	E	E	O	3	O/E	O	**O**	E	O	3	O/E	O	E	E	O
4	**E**	E	O	**O**	**E**	4	**E**	E	**E**	**O**	**E**	4	**E**	E	O	**O**	**E**
Form-2A						*Form-2B*											
Character / Sentence	1	2	3	4	5	Character / Sentence	1	2	3	4	5						
1	**E**	E	O	**O**	**E**	1	**E**	E	**E**	**O**	**E**						
2	O/E	O	O	E	E	2	O/E	O	E	E	E						
3	O/E	O	E	E	O	3	O/E	O	**O**	E	O						
4	**E**	E	O	**O**	**E**	4	**E**	E	**E**	**O**	**E**						
WANG LI 5-CHARACTER QUATRAIN - OBLIQUE START																	
Form-3A						*Form-3B*						*Form-3C*					
Character / Sentence	1	2	3	4	5	Character / Sentence	1	2	3	4	5	Character / Sentence	1	2	3	4	5
1	O/E	O	E	E	O	1	O/E	O	**O**	E	O	1	O/E	O	E	E	O
2	**E**	E	O	**O**	**E**	2	**E**	E	**E**	**O**	**E**	2	**E**	E	O	**O**	**E**
3	E/O	E	E	O	O	3	**E**	E	**O**	O	O	3	**E**	E	**O**	**E**	O
4	O/E	O	O	E	E	4	O/E	O	E	E	E	4	O/E	O	O	E	E
Form-4A						*Form-4B*						*Form-4C*					
Character / Sentence	1	2	3	4	5	Character / Sentence	1	2	3	4	5	Character / Sentence	1	2	3	4	5
1	O/E	O	O	E	E	1	O/E	O	**E**	E	E	1	O/E	O	O	E	E
2	**E**	E	O	**O**	**E**	2	**E**	E	O	**O**	**E**	2	**E**	E	O	**O**	**E**
3	E/O	E	E	O	O	3	**E**	E	**O**	O	O	3	**E**	E	**O**	**E**	O
4	O/E	O	O	E	E	4	O/E	O	E	E	E	4	O/E	O	O	E	E

Table 6

Tonal Pattern of 7-Character Quatrain by Wang Li

WANG LI 7-CHARACTER QUATRAIN - LEVEL START																							
Form-1A								*Form-1B*								*Form-1C*							
Character / Sentence	1	2	3	4	5	6	7	Character / Sentence	1	2	3	4	5	6	7	Character / Sentence	1	2	3	4	5	6	7
1	E/O	E	O/E	O	E	E	O	1	E/O	E	O/E	O	**O**	E	O	1	E/O	E	O/E	O	E	E	O
2	O/E	O	**E**	E	O	**O**	**E**	2	O/E	O	**E**	E	**E**	**O**	**E**	2	O/E	O	**E**	E	O	**O**	**E**
3	O/E	O	E/O	E	E	O	O	3	O/E	O	**E**	E	**O**	O	O	3	O/E	O	**E**	E	**O**	**E**	O
4	E/O	E	O/E	O	O	E	E	4	E/O	E	O/E	O	**E**	**E**	**E**	4	E/O	E	O/E	O	O	E	E
Form-2A								*Form-2B*															
Character / Sentence	1	2	3	4	5	6	7	Character / Sentence	1	2	3	4	5	6	7								
1	E/O	E	O/E	O	O	E	E	1	E/O	E	O/E	O	**E**	**E**	**E**								
2	O/E	O	**E**	E	O	**O**	**E**	2	O/E	O	**E**	E	O	**O**	**E**								
3	O/E	O	E/O	E	E	O	O	3	O/E	O	**E**	E	**O**	O	O								
4	E/O	E	O/E	O	O	E	E	4	E/O	E	O/E	O	**E**	**E**	**E**								
WANG LI 7-CHARACTER QUATRAIN - OBLIQUE START																							
Form-3A								*Form-3B*								*Form-3C*							
Character / Sentence	1	2	3	4	5	6	7	Character / Sentence	1	2	3	4	5	6	7	Character / Sentence	1	2	3	4	5	6	7
1	O/E	O	E/O	E	E	O	O	1	O/E	O	**E**	E	**O**	O	O	1	O/E	O	**E**	E	**O**	**E**	O
2	E/O	E/O	O	O	O	E	E	2	E/O	E/O	O	O	**E**	**E**	**E**	2	E/O	E/O	O	O	O	E	E
3	E/O	E	O/E	O	E	E	O	3	E/O	E	O/E	O	**O**	E	O	3	E/O	E	O/E	O	E	E	O
4	O/E	O	**E**	E	O	**O**	**E**	4	O/E	O	**E**	E	**E**	**O**	**E**	4	O/E	O	**E**	E	O	**O**	**E**
Form-4A								*Form-4B*								*Form-4C*							
Character / Sentence	1	2	3	4	5	6	7	Character / Sentence	1	2	3	4	5	6	7	Character / Sentence	1	2	3	4	5	6	7
1	O/E	O	**E**	E	O	**O**	**E**	1	O/E	O	**E**	E	O	**O**	**E**	1	O/E	O	**E**	E	O	**O**	**E**
2	E/O	E	O/E	O	O	E	E	2	E/O	E	O/E	O	**E**	**E**	**E**	2	E/O	E	O/E	O	O	E	E
3	E/O	E	O/E	O	E	E	O	3	E/O	E	O/E	O	**O**	E	O	3	E/O	E	O/E	O	E	E	O
4	O/E	O	**E**	E	O	**O**	**E**	4	O/E	O	**E**	E	**E**	**O**	**E**	4	O/E	O	**E**	E	O	**O**	**E**

The four sentences within a poetry are called the opening (起 in Chinese, aka Chi), the continuation (承 in Chinese, aka Cheng), the transition (轉 in Chinese, aka Chuan or turning point), and the conclusion (合 in Chinese, aka Ho). Two adjacent sentences are a couplet (聯 in Chinese), and there are totally two couplets in a quatrain. In each couplet, opening and the transition (odd-numbered sentence) are called Chu Ju (出句 in Chinese), whereas the continuation and the conclusion (even-numbered sentence) are called Dui Ju (對句 in Chinese). The tonal relationship between two sentences within a couplet is called "antithesis" (對 in Chinese, aka Dui), in contrast, the tonal relationship between two couplets is called "consistency" (黏 in Chinese, aka Nian). Based on the relationships of antithesis and consistency, Wang indicated the three principle of Recent-style poetry as listed below (L. Wang, 1979, 2015).

(1) The repeated interchanges between level tones and oblique tones is necessary.

(2) The tones of two sentences within a couplet must be opposites, i.e. antithesis.

(3) The tones of 2nd characters in the turning point and the continuation must be identical, i.e. consistency.

The series of studies of Wang simplified the complex questions if license in the Recent-style poetry and provided a general principle of tonal patterns. However, Downer and Graham (1963) indicated that the dependency of tonal pattern is put on the even characters rather than the odd characters. They subdivided the 7-character sentence into three subsystems and produced a linear representation of 4 basic Tonal patterns. The first subsystem consists of the even-numbered positions, which is the basic structure of four basic tonal patterns, whether level or oblique, by the letter A, and the other tone by letter B. The second subsystem is made up of free odd-numbered characters, i.e. Position 1 and 3, which "borrow"' their tones from the immediately succeeding syllable (namely, 2 and 4, respectively). Finally, the third subsystem is composed of the bound odd-numbered characters, namely Position 5 and 7, which is independent of the even-numbered characters. The tone of the rhymed character is labeled by “x” and the other tone by “y”. It must be opposites and takes the tones from the rhyme scheme. Moreover, if the first sentence enters the rhyme scheme, the replacement of these two symbols “x” and “y” is necessary.

	Position						
	1	2	3	4	5	6	7
Line 1		A		B	x	A	y
2		B		A	y	B	x
3		B		A	x	B	y
4		A		B	y	A	x

Fig. 2 Linear Representation of Tonal Patterns

Downer and Graham (1963) provided a linear representation to organize tonal patterns and found that (1) tonal patterns allow only four possibilities for the tone sequence of a line, excluding the situation that the 1st sentence is in rhyme; (2) the tone of the 2nd character of the 1st sentence plays an important role to determine the tone of each even-numbered character within a poetry. Although they proposed a simple structure of tonal patterns, there were still some problems.

Chen (1979) indicated five problems of Downer and Graham’s work as listed below:

(1) Subsystem 2 depends on subsystem 1, so it is redundant.

(2) Downer and Graham’s work did not provide the reason why subsystem 2 borrows their tones from the immediate succeeding characters, but subsystem 3 does not.

(3) The subsystems 1 and 3 do not fall into two natural classes.

(4) The tones of subsystem 2 actually must be opposites.

(5) There is not unified principle which would account for the different sequencing of the pattern alternations in subsystem 1 as contrasted with that of subsystem 3.

Chen proposed a metrical structure organized on the principle of hierarchical representations with exclusively binary branching. Chen also mentioned that the various patterns of alternating sequences of tones in a metrical line are all based on one single principle of prosodic opposition: at each level of the metrical analysis, the sister constituents must carry opposite tones. The basic scheme of tonal opposition is a general structure for each sentence in a 7-character quatrain and a 5-character quatrain as shown in **Fig. 3** and **Fig. 4**. T stands for either a level tone or an oblique tone, and the other tone is presented by letter $\overline{\mathrm{T}}$. In addition, the tone-switching (E to O, and vice-versa) is applied to the 3rd sentence to prevent it from violating Tonotactic Condition, which does not allow four consecutive characters to carry an identical tone, i.e. four Es or Os. Chen explained the unsolved problems of Downer and Graham, and Wnag's studies, finally organized 12 tonal patterns in total as shown in Table 7, Table 8, Table 9, and Table 10, and pointed out some rules at issue in extracanonical lines:

(1) Each sentence permits two free characters.

(2) The first character is always free.

(3) The 5th position (the 5th character in a 7-character sentence) and the 3rd position (the 3rd character in a 5-character sentence) is free only when it is an oblique tone and followed by another oblique tone.

(4) In a 7-character sentence, Position 3 is free, except when position 5 has run out of the quota.

(5) Even-numbered and finial positions are rigid (the tones in these positions are definitely unchangeable).

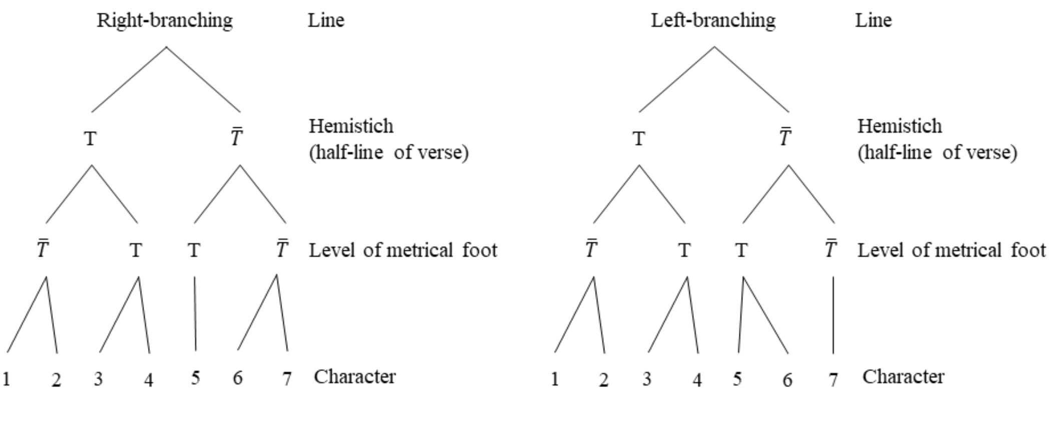

Fig. 3 The Basic Scheme of Tonal Opposition by Matthew Y. Chen (7-Character Quatrain)

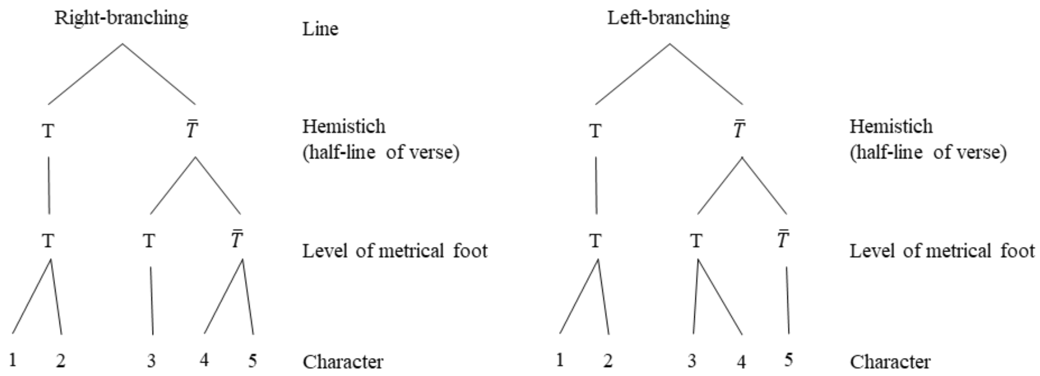

Fig. 4 The Basic Scheme of Tonal Opposition by Matthew Y. Chen (5-Character Quatrain)

Table 7

Tonal Pattern of 5-Character Quatrain by Matthew Y. Chen (Level Start)

Matthew Y. Chen 5-CHARACTER-LEVEL START					
Normal					
Sentence \ Character	1	2	3	4	5
1	E	E	E	O	O
2	O	O	O	E	E
3	O	O	E	E	O
4	E	E	O	O	E
1st sentence in rhyme					
Sentence \ Character	1	2	3	4	5
1	E	E	O	O	E
2	O	O	O	E	E
3	O	O	E	E	O
4	E	E	O	O	E
Extracanonical Lines					
Sentence \ Character	1	2	3	4	5
1	E/O	E	E	O	O
2	O/E	O	O	E	E
3	O/E	O	E	E	O
4	E	E	O/E	O	E

Table 8

Tonal Pattern of 5-Character Quatrain by Matthew Y. Chen (Oblique Start)

Matthew Y. Chen 5-CHARACTER-OBLIQUE START					
Normal					
Sentence \ Character	1	2	3	4	5
1	O	O	E	E	O
2	E	E	O	O	E
3	E	E	E	O	O
4	O	O	O	E	E
1st sentence in rhyme					
Sentence \ Character	1	2	3	4	5
1	O	O	O	E	E
2	E	E	O	O	E
3	E	E	E	O	O
4	O	O	O	E	E
Extracanonical Lines					
Sentence \ Character	1	2	3	4	5
1	O/E	O	E	E	O
2	E	E	O/E	O	E
3	E/O	E	E	O	O
4	O/E	O	O	E	E

Table 9

Tonal Pattern of 7-Character Quatrain by Matthew Y. Chen (Level Start)

Matthew Y. Chen 7-CHARACTER-LEVEL START							
Normal							
Sentence \ Character	1	2	3	4	5	6	7
1	E	E	O	O	E	E	O
2	O	O	E	E	O	O	E
3	O	O	E	E	E	O	O
4	E	E	O	O	O	E	E
1st sentence in rhyme							
Sentence \ Character	1	2	3	4	5	6	7
1	E	E	O	O	O	E	E
2	O	O	E	E	O	O	E
3	O	O	E	E	E	O	O
4	E	E	O	O	O	E	E
Extracanonical Lines							
Sentence \ Character	1	2	3	4	5	6	7
1	E/O	E	O/E	O	E	E	O
2	O/E	O	E	E	O/E	O	E
3	O/E	O	E/O	E	E	O	O
4	E/O	E	O/E	O	O	E	E

Table 10

Tonal Pattern of 7-Character Quatrain by Matthew Y. Chen (Oblique Start)

Matthew Y. Chen 7-CHARACTER-OBLIQUE START							
Normal							
Character / Sentence	1	2	3	4	5	6	7
1	O	O	E	E	E	O	O
2	E	E	O	O	O	E	E
3	E	E	O	O	E	E	O
4	O	O	E	E	O	O	E
1st sentence in rhyme							
Character / Sentence	1	2	3	4	5	6	7
1	O	O	E	E	O	O	E
2	E	E	O	O	O	E	E
3	E	E	O	O	E	E	O
4	O	O	E	E	O	O	E
Extracanonical Lines							
Character / Sentence	1	2	3	4	5	6	7
1	O/E	O	E/O	E	E	O	O
2	E/O	E	O/E	O	O	E	E
3	E/O	E	O/E	O	E	E	O
4	O/E	O	E	E	O/E	O	E

2.5 Poetry Generation

The concept of automatic poetry generation was proposed by Bailey (1974), which attracted the interest of many researchers in training computers to compose a poetry as experts. The fundamental 3 criteria (H. Manurung, 2004) which researches in this automatic poetry generation domain are eager to fulfill are as follows:

(1) Grammaticality (the generated sentences must be readable and grammatically correct).

(2) Meaningfulness (the sentences should convey something related to the theme).

(3) Poeticness (generated poems must have poetic features).

There are a couple of methods for automatic poetry generation, for example, rule-based model (Gervás, 2001), statistical machine translation (SMT) (He et al., 2012; Jiang & Zhou, 2008), and genetic algorithm (R. Manurung, Ritchie, & Thompson, 2012). However, there are some shortcomings of them as follows:

(1) Rule-based model

Gervás (2001) used rule-based model to generate poetry by setting constraints on sentence pattern and grammar, and also found that both topic and keyword play critical roles in poetry generation. However, it lacks semantic analysis, and the underlying meanings of the existing poetries cannot be well extracted. As a result, it does not reflect diversity of the poetries, is poor at meaningfulness, and violates poeticness.

(2) SMT

Compared with rule-based model, SMT relies on corpus and puts more emphasis on relationships among article, sentence, phrase, and vocabularies. It finds the n-best candidate sentences from corpus depending on the statistical probability of vocabulary by a character-based trigram language model. However, it highly increases the risk of poeticness violation, since term frequency and constraint setting still cannot express the coherences between sentences effectively.

(3) Genetic algorithm

R. Manurung et al. (2012) applied a linguistic representation based on Lexicalized Tree Adjoining Grammar for evolutionary algorithms based natural language generation to deal with the problems of semantics and grammar of poetries. Although the result has thematic meaning, the quality of poeticness would be affected by random combination.

In recent years, machine learning (ML) and deep learning (DL) have applied to the automatic poetry generation. Z. Wang et al. (2016) trained 76,859 five and seven-character quatrains with Bi-LSTM and the keywords are extended by RNN model. Yi et al. (2017) separated the sentences into 3 pairs to train 398,391 poems from Tang dynasty with Bi-directional RNN Encoder-Decoder to generate quatrains with 4 sub-models and attention mechanism for capturing association between 2 characters. D. Liu, Guo, Li, and Lv (2018) trained 149,524 poetries with RNN Encoder-Decoder and also used LDA to extend relevant keywords. However, the topic words generally affect the first sentence in the poetry, and any unconstrained keywords may result in lower quality of the generated poetry.

Huang et al. (2020) trained 90,208 quatrains of Tang and Song dynasty with LSTM Encoder-Decoder and designed 4 sub-models for each consecutive sentence within a quatrain to compose poetries. Each sentence of 5-character and 7-character Chinese quatrains are segmented into words according to the 6 specific templates by Classical Chinese Poetry Forms (CCPF). They applied TextRank to extract the keywords of each sentence, and LDA to classify the segmented words into the corresponding topics. After that, these segmented words were stored into the corresponding dictionaries of the QGM as the basis for poetry generation. The QGM was trained by Encoder-Decoder LSTM to learn the rules of each consecutive sentence of the poetry to automatically compose a Chinese classical poetry which meets the 3 elements of poetry generation defined by H. Manurung (2004), i.e. meaningfulness, grammaticality and poeticness.

2.6 Classical Chinese Poetry Forms

N.-J. Wang, Tseng, Yang, Su, and Lo (2015)'s research indicated that each sentence in both 5-character and 7-character quatrains starts with a double-character, and the last three characters will be segmented according to the metrical rules.

Classical Chinese Poetry Forms (CCPF) is a word segmentation which contain 6 specific templates as shown in Table 11 (Huang, 2019). It focuses on tern frequencies of the combinations of the last 3 characters of each sentence. There are three combinations: (A) 3 characters in a group, (B) preceding 2 characters in a group (2+1), and (C) succeeding 2 characters in a group (1+2). If the term frequency of the combination A is greater than 1, it means the word is meaningful, and the template of 5-character quatrain is 2+3, and template 2+2+3 is for 7-character quatrain. Otherwise, we continue comparing the term frequencies of template 2+1 to template 1+2. If the term frequency of template 2+1 is higher, the sentence will be segmented into 2+2+1 for five-character quatrain and 2+2 +2+1 for seven-character quatrain, and vice versa.

Table 11

CCPF Segmentation Template

Quatrain	Segmentation Template
Five-character	(2+3)、(2+2+1)、(2+1+2)
Seven-character	(2+2+3)、(2+2+2+1)、(2+2+1+2)

2.7 TextRank

TextRank is a graph-based ranking model for text processing proposed by Mihalcea and Tarau (2004). Graph-based ranking algorithm is a way of deciding the importance of a vertex within a graph. The core idea of it is voting or recommendation. A vertex links to another one, and meanwhile it votes for that another vertex. The vertex with higher number of the vote has a higher importance and also has a more powerful vote than others.

In TextRank, the graphs consist of natural language texts, and text units with flexible length, including words, collocations, sentences, or others, can be added as vertices in the graph. TextRank identifies the connections between text units in a text, and the connection between two vertices is called edge. The strength of connection is seen as a weight and be added to corresponding edge. It also implements the concept of recommendation. A text unit recommends other relevant text units, and the strength of the recommendation is recursively computed based on the importance of the units making the recommendation. The algorithm continues computing until convergence. Eventually, rank the vertices in order of their final score. Therefore, TextRank draws information from a text and identifies the most important sentences in the text.

The formula of TextRank which integrates the score of a vertex V_i (namely S(Vi)) with vertex weights is shown in Equation 1. WS(Vi) is the weight of vertex i, and V represents the set of vertices. $V_j \in In(V_i)$ is the vertex V_j which points to the set of vertices $In(V_i)$, whereas $V_k \in Out(V_j)$ is the vertex V_k which is the set of vertices pointed by the vertex V_j. "w_{ji}" is the edge weight that the vertex V_j points to the vertex V_i, whereas "w_{jk}" is the edge weight that the vertex V_j points to the vertex V_k. "d" is a damping factor within the range between 0 and 1, usually 0.85. This damping factor is the probability of switching from a particular vertex to another one in the graph.

$$WS(V_i) = (1-d) + d * \sum_{V_j \in In(V_i)} \frac{w_{ji}}{\sum_{V_k \in Out(V_j)} w_{jk}} WS(V_j) \qquad \text{Equation 1}$$

2.8 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) is proposed by Blei, Ng, and Jordan (2003). It is an unsupervised generative probabilistic model for collections of discrete data to discover (1) the number of topics in corpus, (2) the distribution of topics in each document, and (3) the distribution of words in each topic. The two basic ideas of LDA are (1) Each document consists of several latent topics, and (2) Each topic is described by several important words, and these words may appear in different topics simultaneously. The structure of LDA is graphically shown in **Fig. 5**, and the equation with the joint distribution of the variables is shown in Equation 2."W" stands for a word in the document, "Z" means a topic in a document, "N" represents number of words in the document, "M" is number of document to analyze, and "K" means number of topics. "α" is per-document topic distribution, whereas "β" is per topic word distribution. "θ" is topic probability per document (topic distribution for a document), whereas φ is topic probability per word (words distribution for a topic). W_{ij} represents j-th word in i-th document, and Z_{ij} means topic assignment for W_{ij}.

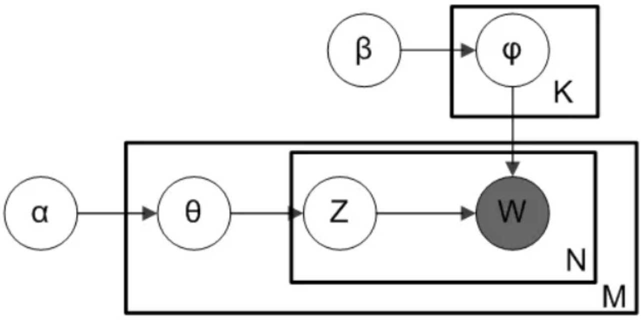

Fig. 5 LDA Structure

Resource: Anindya Naskar. (2019). *Latent Dirichlet Allocation for Beginners: A high level overview*. Retrieved from https://thinkinfi.com/latent-dirichlet-allocation-for-beginners-a-high-level-overview/

$$p(W_i, z_i, \theta_i, \psi | a, \beta) = \prod_{j=1}^{N} p(\theta_i | a) p(z_{i,j} | \theta_i) p(\psi | \beta) p(W_{i,j} | \theta_{zi,j}) \qquad \text{Equation 2}$$

2.9 Word to Vector

Word embedding and one hot encoding are two vector representations in Natural language processing, which changes the form of word into vector, since computer does not understand plain texts. One Hot encoding is a representation of categorical variables as binary vectors, either 0 or 1. Each word has its value in a vector, which is represented as a binary vector that is all zero values except the index of the word with 1. Once the amount of data is extremely huge, it produces a sparse matrix. The cons of one-hot representation are that (1) data in high dimensions may cause the curse of dimensionality, and (2) it loses the semantic relationships between words because all the words are encoded separately.

Word Embedding takes context into account, and gives words, which have similar meaning, a similar value for a specific feature, and it solves the problems of one-hot representation. The core idea of word embedding is to changes each word into word vector and place them to a vector space in low dimension and high information density (compared with one-hot representation). The similarity between words is measured by distance; in other word, the words with similar context occupy close spatial positions.

Word to Vector (Word2vec) is one of Word embedding proposed by Mikolov, Chen, Corrado, and Dean (2013) to predict the word based on the given information. There are two models of it, Common Bag of Words (CBOW) and Skip Gram (**Fig. 6**). The CBOW predicts the current word with the given context, whereas the Skip-gram predicts surrounding words with the given current word. Both of them have input layer, projection layer, and output layer. In CBOW, the vectors of the input layer are averaged in projection layer.

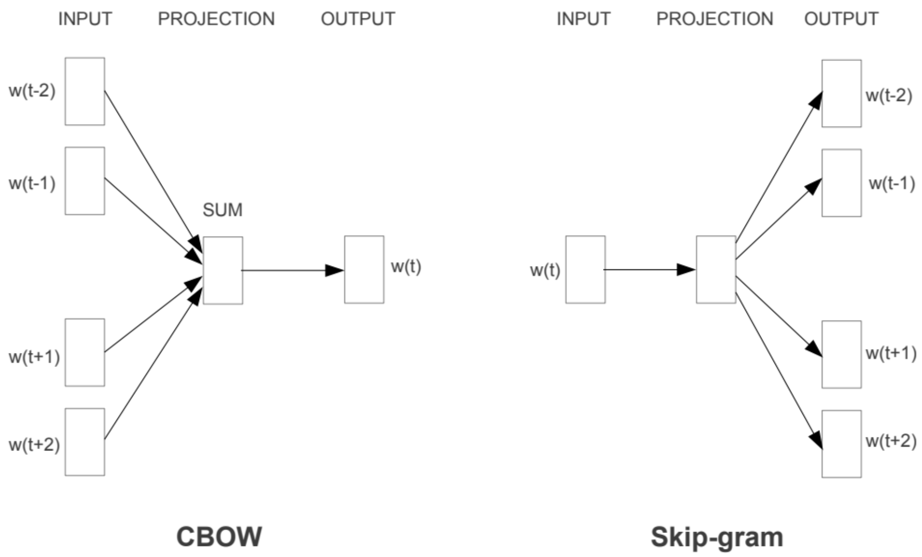

Fig. 6 Structure of the CBOW and Skip-Gram Model

Resource: Tomas Mikolov, Kai Chen, Greg Corrado, & Jeffrey Dean. (2013). *Efficient Estimation of Word Representations in Vector Space*. Retrieved from https://arxiv.org/abs/1301.3781

2.10 Long Short-Term Memory

Long Short-Term Memory (LSTM) is an improved model based on Recurrent Neural Networks (RNN). The idea of RNN is to make use of sequence data, which features time steps, for prediction. The output of previous time step is the input of the subsequent time step and meanwhile updates the current hidden state for prediction; therefore, RNN has a memory which captures the historical information calculated up to now. The structure of RNN is shown in **Fig. 7**, and the equations are shown in Equation 3 and Equation 4. RNN consist of three layers, i.e. input layer(x), hidden layer, and output layer. $\mathrm{X_t}$ in the input layer is an input at time step *t*, $\mathrm{S_t}$ is the hidden state at time step *t* in the hidden layer, and O_t is the output at time step *t* in the output layer. “V”, “U”, and “W” are shared weights in RNN across all steps, and “b” means a bias.

A loop allows information to be passed from one step of the network to the next. Equation 3 indicates that S_t refers to the value of S_{t-1} and the current input with an activation function *f*, which is a nonlinearity, e.g. tanh, ReLU, and sigmoid. Equation 4 indicates that Output is determined by the current hidden state with an activation function, usually Softmax.

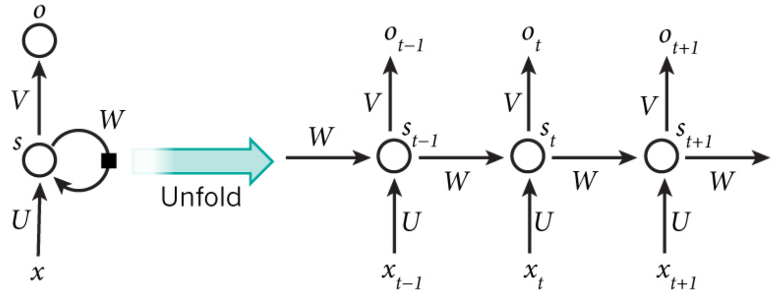

Fig. 7 RNN Structure

Resource: Denny Britz. (2015). *Recurrent Neural Networks Tutorial, Part 1 – Introduction to RNNs.* Retrieved from http://www.wildml.com/2015/09/recurrent-neural-networks-tutorial-part-1-introduction-to-rnns/

$$S_t = f(WS_{t-1} + US_t + b) \qquad \text{Equation 3}$$

$$O_t = softmax(VS_t + b) \qquad \text{Equation 4}$$

RNN is good at dealing with sequence data, and it is able to connect information to the current task. However, it in fact is limited to capture the information too many time steps ago; in other words, it does not perform well on the task which requires a long-term dependency. Moreover, it would suffer vanishing gradient problem when Backpropagation, which mea ns that RNN is unable to turn back to update the weight in previous few layers. Therefore, LSTM is proposed to solve the problems of RNN.

LSTM is proposed by (Hochreiter & Schmidhuber), and the structure of LSTM is shown in Fig. 8. LSTM contains four interactive layers to persists or remove data with a long-term dependency to the cell state, and controls weights with gates, namely forget gate (*f*), input gate (*i*), and output gate (*o*). Gates with a sigmoid function, which outputs numbers between 0 and1, are the decider to let information in (value equals 1) or not (value equals 0).

The *C* stands for cell state, and *t* means time step. f_t is a sigmoid layer called forget gate to keep and throw away the information from the cell state, and the equation is shown in Equation 5. W_f is the weight matrix of the forget gate, h_{t-1} is the output of time step *t-1*, x_t is the input of current time step *t*, and b_f means the bias of forget gate.

There are two layers as deciders for optionally storing new information in the cell state. The first one is Sigmoid layer called input gate which is the value at the time step t, i.e. i_t, and it decides which values to be kept. The other one is $\tilde{C}_t$ calculated by an activation function "tanh", which creates a vector between -1 to 1 of new candidate values that could be added to the state. In other words, $\tilde{C}_t$ is the new information to be stored. The equations are shown in Equation 6 and Equation 7.

Equation 8 shows the formula to update the old cell state $\tilde{C}_{t-1}$ to new one C_t, which is the new candidate values for the next cycle. The previous cell state $\tilde{C}_{t-1}$ times f_t, forgetting something the forget gate decided not to remember, and then plus the result of i_t multiplied by $\tilde{C}_t$.

There are also two layers as filters for optionally delivering information to the next time step and also producing an output. The first one is the output gate with a sigmoid function, and the purpose of it is to decide which part of the cell state to be output, the equation is shown in Equation 9. The other one is tanh function which pushes the value of C_t between -1 and 1, and then times o_t to produce the final result

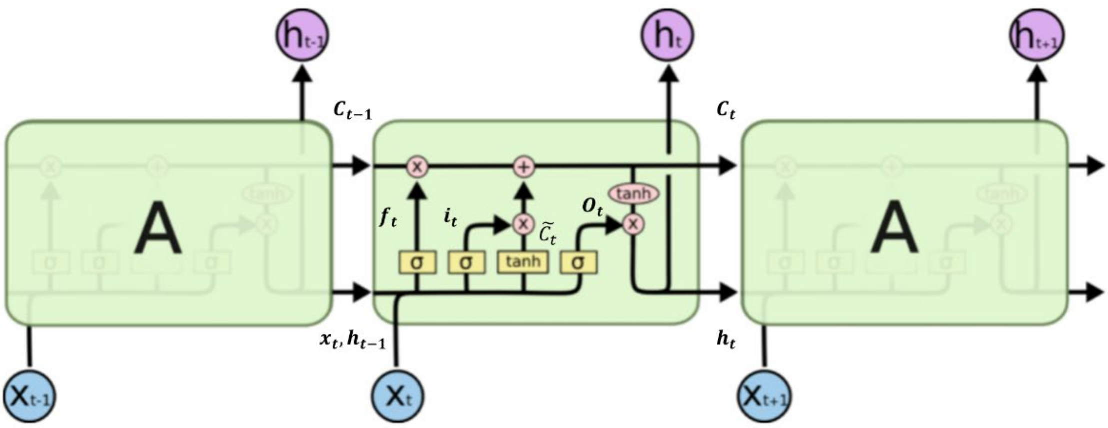

Fig. 8 LSTM Structure

Recourse: Colah. (2015). *Understanding LSTM Networks.* Retrieved from *https://colah.github.io/posts/2015-08-Understanding-LSTMs/*

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \quad \text{Equation 5}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad \text{Equation 6}$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad \text{Equation 7}$$

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad \text{Equation 8}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad \text{Equation 9}$$

$$h_t = o_t \cdot tanh(C_t) \quad \text{Equation 10}$$

LSTM Encoder-Decoder is one of the Sequence to Sequence model (Seq2seq) which succeeds in machine translation (Cho et al., 2014; Sutskever, Vinyals, & Le, 2014). Encoder-Decoder is a model which learns the mapping from a sequence of an arbitrary length to another sequence of an arbitrary length; in other word, the length of input and output sequence might be different marked by T and T', separately. Besides, it learns the conditional probability distribution over a variable-length sequence conditioned on yet another variable-length sequence, as shown in **Fig. 9**.

LSTM Encoder-Decoder is composed of three elements, they are encoder, context vector, and decoder. The encoder encodes a variable-length sequence (X_1, X_2, …, X_T) into a fixed-length vector representation, and also reads each symbol of an input sequence x sequentially. Meanwhile, the hidden state at time *t*, i.e. h_t, changes according to Equation 11. The hidden state becomes a summary *C* of the entire input sequence, i.e. context vector, after encoder reads the ending symbol of the sequence. The decoder decodes a given context vector *C* in a fixed-length back into a variable-length sequence. The hidden state of decoder at time t is calculated by Equation 12, which refers to the hidden state and output at previous time *t-1*. The decoder generates the output sequence (y_1, y_2, …, $\mathrm{y}_{\mathrm{T}'}$) by predicting y_t, and the conditional distribution of the next symbol is shown in Equation 13. “f” and “g” are activation functions, e.g. softmax.

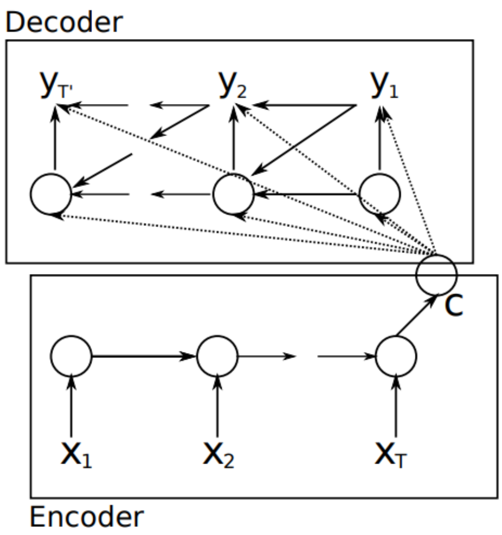

Fig. 9 Encoder-Decoder Illustration

Resource: Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, Yoshua Bengio. (2014). *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.* Retrieved from https://arxiv.org/abs/1406.1078

$$h_t = f(h_{t-1}, y_{t-1}) \tag{Equation 11}$$

$$h_t = f(h_{t-1}, y_{t-1}, c) \tag{Equation 12}$$

$$p(y_t | y_{t-1}, y_{t-2}, \dots, y_1, C) = g(h_t, y_{t-1}, C) \tag{Equation 13}$$

Chapter 3 Research Method

This research is going to examine the most common Chinese tonal patterns, i.e. the four basic tonal patterns, their variations, and the particular tonal patterns, of extant 5-characer quatrains, and remain the normal quatrains which are compliant with the most common tonal patterns as the dataset of this study. We apply CCPF Segmentation to segment the sentence into words with one to three characters in length. These segmented words will not only be trained by Word2Vec, but also be clustered into 28 topics by LDA, and later stored into four-word dictionaries in order of the sentence. Moreover, the keywords, which represents the core of each sentence, will be extracted from sentences by TextRank. The training model takes the keyword and sentences as input to learn the knowledge of quatrains, and performs automatic poetry generation with the four words stored in word dictionaries and the previous generated sentences. The research structure is shown in **Fig. 10**.

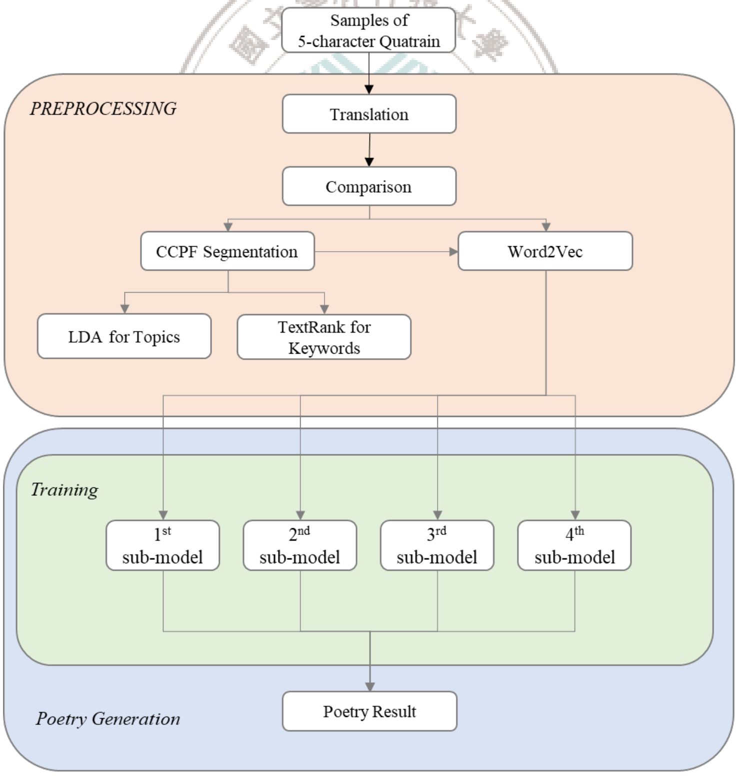

Fig. 10 Research Structure

The preprocessing will be described in section 3.1, and the details in training phase is described in section 3.2. Finally, in section 3.3, we are going to talk about the structure of poetry generative model and poetry generation.

3.1. Preprocessing

3.3.1. Resource of Data

For examining the fitness of existing five-character quatrains and the most common tonal patterns, this study firstly listed possible combinations of 5-character quatrains' tonal patterns and later collected the tones of Chinese characters. We totally organized 345 five-character quatrain tonal patterns based on the previous researches (Chen, 1979; Downer & Graham, 1963; Graham, 1980; L. Wang, 1979). Besides, the open resources of tones are derived from the website "Souyun" (搜韻 in Chinese) ("搜韻," 2009). The total number of characters is 20,739, but the characters in use are 4,910. The words exist in extant quatrains but without the data of rhyme data will be seen as Null and removed from the candidates for poetry generation.

This study contains 12,486 five-character quatrains of the Tang and Song dynasties derived from National Library of China (中國國家圖書館). Since the quatrains were recorded in simplified Chinese, this study used OpenCC to translate them into Mandarin and obtained eventually 12,398 five-character quatrains. In comparison with the 345 tonal patterns, this study remained 7,599 five-character quatrains as training samples. The details of quatrain are shown in Table 12.

Table 12

Details of 5-Character Quatrain

Dynasty	Tang Dynasty 5-character Quatrain	Song Dynasty 5-character Quatrain	Total
Number	1,900	5,699	7,599

3.3.2. CCPF, TextRank, and LDA

We applied CCPF to do word segmentation and obtained totally 53,417 words which have one to three characters in length. The example of CCPF Segmentation is shown in Table 13. The number of segmented words from individual sentences are shown in Table 14. Moreover, TextRank would select a segmented word from each individual sentence as a keyword to express or represent the core of the sentence. The segmented word with the highest term frequency is seen as a substitute of keyword if keyword is not found by TextRank.

Table 13

Example of CCPF Segmentation

Segmentation Template	(2+3)		(2+2+1)			(2+1+2)		
Sentence	床前	明月光	床前	明月	光	床前	明	月光

Table 14

Number of Words by CCPF Segmentation

Sentence	The first sentences	The second sentences	The third sentences	The fourth sentences	Total
Number of Words	13,955	13,609	12,896	12,957	53,417

Because the topics of Chinese classical poetry were classified to 28 categories by experts whom Yuntech Software Engineering Laboratory invited, the segmented words were clustered into 28 topics by LDA. These segmented words from individual sentences were stored into corresponding word dictionary for poetry generation. The illustration of LDA Word Dictionary is shown in **Fig. 11**.

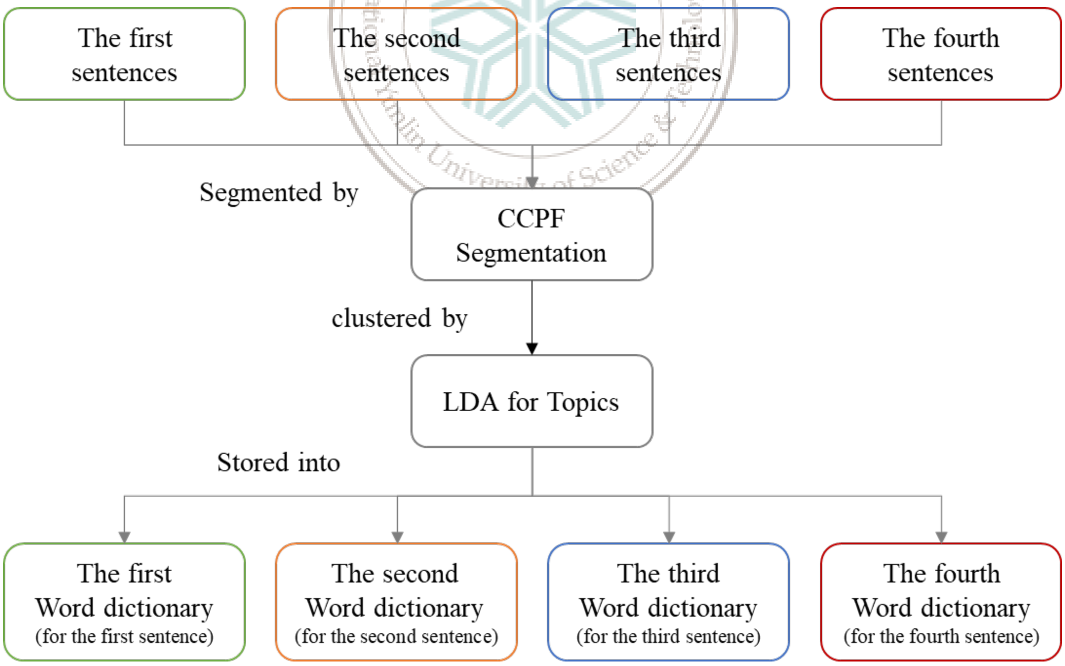

Fig. 11 The Illustration of LDA Word Dictionary

3.3.3. Word2Vec

This study applied skip-gram as the algorithm to embed the words for (1) training and poetry generation, and (2) comparison of the similarity between LDA words in the same topic. For training and poetry generation, we made use of 4 kinds of word segmentation, to wit: CCPF segmentation, Unigram, CSCP segmentation, and Jieba, to increase the flexibility of model in poetry generation. For comparison of the similarity, we only used CCPF segmentation since the words stored in word dictionaries are segmented previously by CCPF segmentation.

The parameter of Word2Vec is listed in Table 15. Because of the big number of words segmented by CCPF segmentation, this study changed the size from 100 into 200. This study presumed that each character in a sentence has a close relationship with each other, so the window was set to 3. In order to remain the words with low term frequency, the mini_count was set to 1.

Table 15

Details of Word2Vec Parameters

Parameter	Description	Value
sentence	The dataset	5-character Chinese quatrain
size	The vector of Word embedding, default = 100	200
window	the size of the window is a parameter that can be tuned for each problem.	3
min_count	The threshold of term frequency, default = 5	1
sg	Algorithm, 1:skip-gram, 0:CBOW	1

3.2. Training

The targets in training phase are to let training model learn the meaning of each sentence, sentence patterns, and tonal patterns. In order to learn the knowledge of four consecutive sentences in a five-character quatrain, this study used LSTM Encoder-decoder as the research model, which consists of 4 sub-models for each sentence. We gave a particular sentence(s) with its keyword(s), which were processed by TextRank, as an input to encoder, and also told training model that the sentence immediately after the given sentence is the output of decoder.

In the 1^{st} sub-model, the input of encoder is a keyword obtained by TextRank from the first sentence, and the output of decoder is the first sentence. The inputs of encoders in the following three sub-models consist of the keywords from their own sentence and the sentence generated by the sub-model immediately before them. Hence, the inputs of encoder in the 2^{nd} sub-model are (1) a keyword obtained by TextRank from the second sentence and (2) the first sentence, and the output of decoder is the second sentence. The inputs of encoder in 3^{rd} sub-model consist of (1) a keyword obtained by TextRank from the third sentence, (2) the first sentence, and (3) the second sentence, and the output is the third sentence, whereas the 4^{th} sub-model takes (1) a keyword obtained by TextRank from the fourth sentence, (2) the second sentence, and (3) the third sentence, and afterwards output the fourth sentence. Table 16 is an example of the inputs and outputs of 4 sub-models.

Table 16

Example of Inputs and Outputs of Training Sub-Models

Sub-model	The 1^{st} sub-model	The 2^{nd} sub-model	The 3^{rd} sub-model	The 4^{th} sub-model
Input	The 1^{st} keyword	The 2^{nd} keyword	The 3^{rd} keyword	The 4^{th} keyword
		The 1^{st} sentence	The 1^{st} sentence	The 2^{nd} sentence
			The 2^{nd} sentence	The 3^{rd} sentence
Example of input	走卻	移將	世間	八萬
		走卻坐禪客	走卻坐禪客	移將不動尊
			移將不動尊	世間顛倒事
Output	The 1^{st} sentence	The 2^{nd} sentence	The 3^{rd} sentence	The 4^{th} sentence
Example of output	走卻坐禪客	移將不動尊	世間顛倒事	八萬四千門

3.2.1. The 1^{st} Sub-model

The 5-character quatrains were segmented by CCPF segmentation, so the max length of words is three. In word embedding layer, the size of word embedding is 200, so the size of matrix is 3 times 200. The output would be the first sentences made up of 5 characters with two annotations “[BOS]” (beginning of the sentence) and “[EOS]” (end of the sentence), so we set 7 vectors to generate sentence and softmax as the activation function. The details of parameter settings are shown in Table 17.

Table 17

Parameters of the 1st Sub-model in Training Phase

Layer	Description	Value
Embedding	Sending the data of Word2Vec to embedding layer	3*200
Encoder LSTM	Setting the number of Encoder unit	200
Decoder LSTM	Setting the number of Decoder unit	200
Output Layer	Setting the number of word in glossary and activation function	4910*7, softmax

3.2.2. The 2nd, 3rd, and 4th Sub-model

In the 2nd sub-model, the inputs include the 2nd keyword which has one to three characters in length, a section break [SEP] in the middle, and the 1st sentence, so the max length of words is nine (= 3+1+5). Similarly, the max length of inputs in the 3rd and 4th sub-model become fifteen (=3+1+5+1+5) accumulated by their own keywords, two section breaks between 2 sentences immediately before them. The rest value of parameters is same as the 1st sub-model. Table 18 and Table 19 shows the parameters of the 2nd, 3rd. 4th sub-models.

Table 18

Parameters of the 2nd Sub-model in Training Phase

Layer	Description	Value
Embedding	Sending the data of Word2Vec to embedding layer	9*200
Encoder LSTM	Setting the number of Encoder unit	200
Decoder LSTM	Setting the number of Decoder unit	200
Output Layer	Setting the number of word in glossary and activation function	4910*7, softmax

Table 19

Parameters of the 3rd and 4th Sub-model in Training Phase

Layer	Description	Value
Embedding	Sending the data of Word2Vec to embedding layer	15*200
Encoder LSTM	Setting the number of Encoder unit	200
Decoder LSTM	Setting the number of Decoder unit	200
Output Layer	Setting the number of word in glossary and activation function	4910*7, softmax

3.3. Poetry Generation

The first words play an important role in the quatrain generation because it determines the topic of the generated quatrain, and the generative model will search the other three keywords from respective LDA dictionaries, which have the highest similarity computed by Word2Vec. The generative model is shown in **Fig. 12**.

The approach of quatrain generation is similar to the training model. The generative model takes the first keyword to generate the 1st sentence. The second sub-model takes the first generated sentence and the second keyword as input to generate the second sentence. The third and fourth sub-models takes the two sentences immediately before them and their own keyword to output the third and fourth sentence. Eventually, the production of a generated quatrain is done.

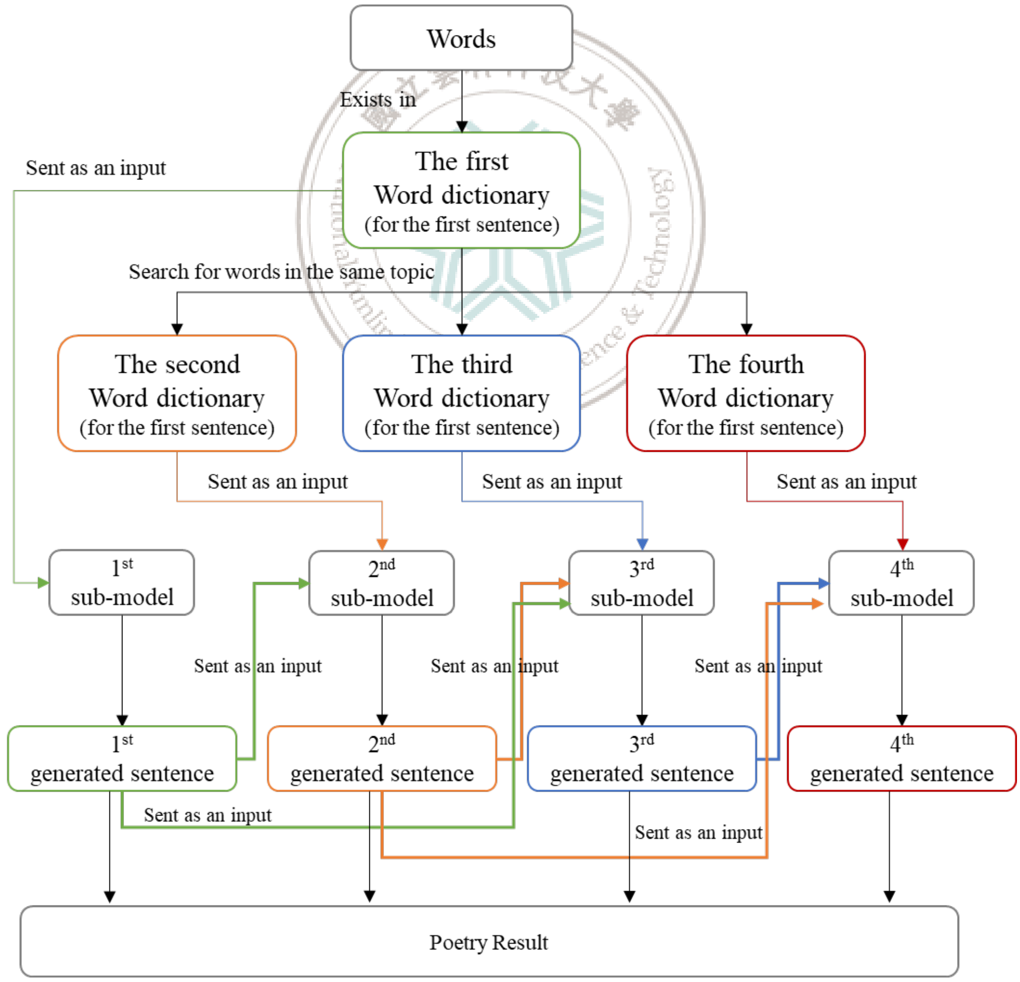

Fig. 12 Generative Model

Chapter 4 Experimental Result

This study used the extant 7,599 five-character quatrains which are compliant with the most common 345 tonal patterns as the training sample, and employed the 1,712 combination of words and their extended words from word dictionaries, which share the same topic, to eventually generate 660 Chinese 5-character quatrains following tonal patterns.

4.1. Tonal Patterns

This study organized 345 five-character quatrain tonal patterns excluding counterbalance based on the studies of (Graham, 1980; L. Wang, 1979). The number of 5-character Chinese quatrains in the Tang and Song Dynasties is 12,398, but only 7,599 five-character quatrains are compliant with the tonal patterns. Among 7,599 five-character quatrains, the number of the quatrains written in the four basic tonal patterns is 5,629, whereas the number of the quatrains written the variations based on the four basic tonal patterns is 1,236. The number of quatrains written with particular tonal patterns is 734. The comparison of these three-writing format is shown in Table 20. The 135 tonal patterns are provided in appendix.

Table 20

Comparison of Three kinds of The Most Common Tonal Pattern

Tonal Pattern	The four basic tonal patterns	The variation of the four basic tonal patterns	The Particular Tonal Patterns
Number	5, 629	1, 236	734

4.2. Model Evaluation

This study took whole 7,599 five-character quatrains to be the training dataset because of concise wording. There are very few cases that quatrains have extremely high similarities, so the testing set and validation set cannot present the generalization of our model.

The training results of four sub-model indicated that the models are convergent. The accuracy is for word to word poetry generation, which means that the model can correctly predict the word verbatim, and the average of accuracies is 0.98. Among them, the accuracy of the first sub-model is the lowest one, and the key problem is the 1^{st}

keyword. Some extant quatrains were written with the same 1st keyword, but the consecutive sentences were totally different from each other. The first sub-model learns the first sentence with great difficulty because it has to understand the difference among the first sentences which shares the same keyword. Therefore, the accuracy of it is lower than the others. Besides, the loss function we applied is Cross Entropy, which the value becomes smaller when the predicted values is similar to actual value. The accuracies of four sub-models are listed in Table 21. The training result is graphically shown in Fig. 13 to Fig. 20.

Table 21

Accuracies of Training Sub-Models

Sub-model	The first sub-model	The second sub-model	The third sub-model	The fourth sub-model
Accuracy	0.96	0.99	0.99	0.99

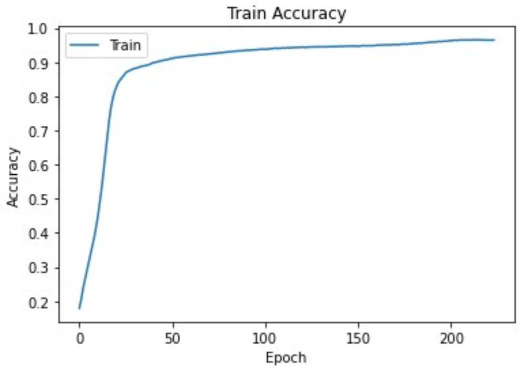

Fig. 13 Accuracy of the First Sub-model

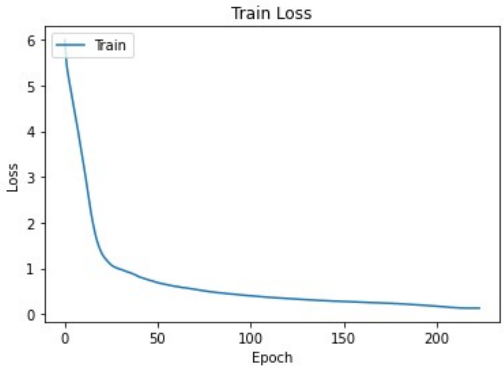

Fig. 14 Loss of the First Sub-model

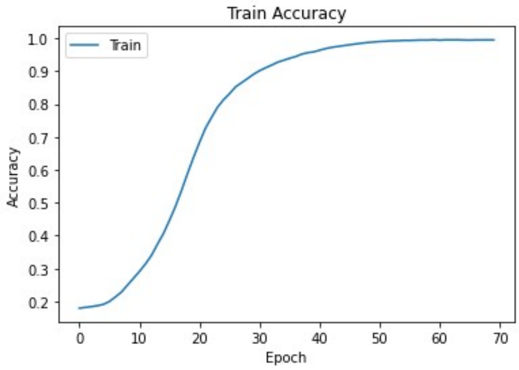

Fig. 15 Accuracy of the Second Sub-model

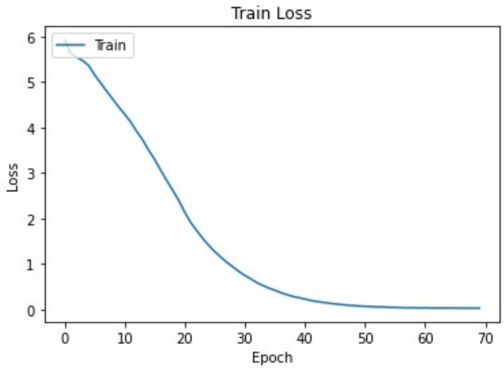

Fig. 16 Loss of the Second Sub-model

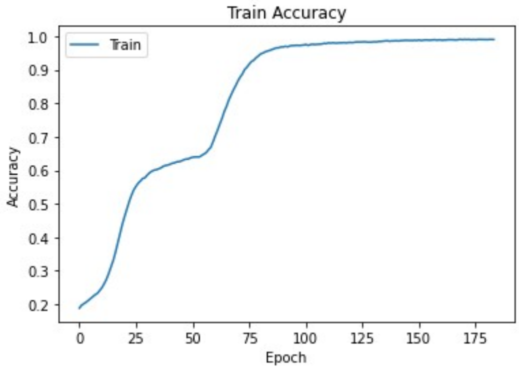

Fig. 17 Accuracy of the Third Sub-model

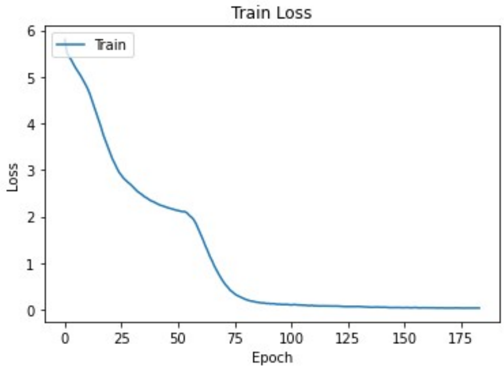

Fig. 18 Loss of the Third Sub-model

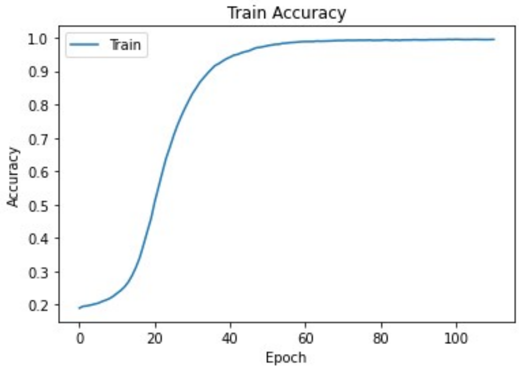

Fig. 19 Accuracy of the Fourth Sub-model

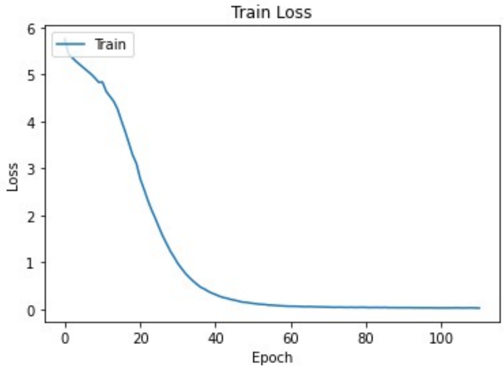

Fig. 20 Loss of the Fourth Sub-model

4.3. Quatrain Generation Result

This study employed 7599 five-character quatrains, which is compliant with the tonal pattern excluding counterbalance, as the training samples. The generative model freely chose a word stored in the first word dictionary, and later searched the other three words, which belong to the same topic and also have the highest similarities, from the other word dictionaries. After that, the generative model obtained 1,712 combinations, and eventually generated 1,712 five-character generated quatrains in total. Among the 1,712 five-character generated quatrains, there are 72 defective quatrains containing padding (symbolized as "[PAD]") which fills the blank to the max length. The problems occur when result of prediction, which is exactly the output of Softmax activation function, is that all probabilities are the same.

Among the rest 1,640 quatrains, there are 660 (around 40%) quatrains acting in compliance with tonal patterns. Moreover, the number of ending characters in the even-numbered lines containing an oblique tone is 22. The reason of these mismatched result probably is the unconstrained word position. The words were segmented into one to three characters in length and were located without any constraints. Therefore, the research model has a high tendency to misplace the words without considering the tones patterns.

This study compared the result to the research of Huang et al. (2020) since the research structures are similar. Huang et al. (2020) took 90,208 five and seven-character quatrains, including regulated quatrain and ancient-style quatrain, as the training sample without the examination of tonal patterns, and generated 572 quatrains in total. We found that our accuracy of tonal patterns for 5-character generated quatrain is approximately 30% higher than their accuracy of 5-character quatrains. Besides, our percentage of ending characters in the even-numbered lines containing an oblique tone is 0.02, but theirs is 0.1. The accuracy of tonal patterns for 5-character generated quatrain and the percentage of ending characters in the even-numbered lines containing an oblique tone prove the importance of training samples. The details of comparison are shown in Table 22. The five-character generated regulated quatrain is abbreviated to “5-char. GQ”, and the ending characters in the even-numbered lines containing an oblique tone is abbreviated to “QO”.

Table 22

Comparison of Quatrain Generation Result

Study	Training Sample	The Number of 5-char. GQ	The Number of 5-char. GQ Matching Tonal Patterns	Accuracy of 5-char. GQ	The Number of QO	The Percentage of QO
Huang et al. (2020)	90,208 Five and seven-character quatrains	74	9	0.12	7	0.10
This study	7,599 Five-character quatrains following the tonal patterns excluding counterbalance	1,640	660	0.40	22	0.02

There are three examples of result shown in Table 23. The first example of 5-chracter generated quatrain not only be compliant with tonal patterns, but also by in rhyme with the rhyme groups “麻”. The rhyming characters had been highlighted in red. The second example indicates that the positions of the word selected from the word dictionaries are random. The third example demonstrates that our generative model is able to alter the word dynamically to generate a smooth and coherent quatrain.

Table 23

Examples of 5-Chracter Generated Quatrain

No.	Words from Word Dictionary	5-Character Generated Quatrain	Tonal Pattern
1	塞北、江南、征帆、應似	*塞北* 亢喉地，*江南* 是妾家。 *征帆* 才及岸，*應似* 早春花。	仄 仄 ＊ 平 仄 ， 平 平 仄 仄 平 。 平 ＊ 平 仄 仄 ， ＊ 仄 仄 平 平 。
2	夜月、傷心、人不見、今日	*夜月* 三杯色，*傷心* 欲自迷。 泉存 *人不見*，*今日* 自魂銷。	仄 仄 ＊ 平 仄 ， 平 平 仄 仄 平 。 平 平 平 ＊ 仄 ， 平 仄 仄 平 平 。
3	綵縷、薰風、誰知、更不	*綵縷* 新纏臂，*柔風* 動賞心。 *誰知* 今牆外，*更不* 向諸家。	仄 仄 平 ＊ 仄 ， 平 平 仄 仄 平 。 平 ＊ 平 ＊ 仄 ， ＊ ＊ 仄 平 平 。

4.4. BLEU Evaluation

BLEU is a popular evaluation for the quality of machine translation which compares the similarities between the translation results of machine and human, and gives a score in the range 0-1.

He et al. (2012) assumed that the sentences sharing the same or similar keyword would have similar succeeding sentences. Considering the majority of words are segmented into 2 characters in length by CCPF segmentation, this study conducted BLEU-2 evaluation with reference to the research of (He et al., 2012; Huang et al., 2020; Zhang & Lapata, 2014). The keywords in a 5-character sentence would be “the 1st and 2nd character”, “the 3rd to 5th character”, “the 3rd and 4th character”, “the 5th character”, “the 3rd character”, and “the 4th and 5th character”. For example, “床前明月光” will be separated into “床前”, “明月光”, “明月”, “光”, “明”, and “月光”. Besides, the combinations of comparison are (1) the 1st sentence to 2nd sentence, (2) the 2nd sentence to 3rd sentence, and (3) the 3rd to 4th sentence.

Table 24 shows the score of BLEU-2 in comparison with the other three researches. This study got the highest scores in each combination. The possible reason is the different varieties of training sample. The research of Lu (2020) indicated that the total number of 5-character quatrains is relatively far lower than 7-character quatrains, and the most widely used words are from 7-character quatrains. This disparity results in 5-character quatrain’ lower production and also lower performance on BLEU-2 evaluation. This study only applied 5-character quatrains which must follow the tonal patterns, so our model concentrates on learning the knowledge of the 5-cahracter quatrains, and afterwards it increases the BLEU-score. Although it is the highest score, there's still room for improvement. In order to prevent the model from generating incoherent and non-semantic quatrains, we limit the number of topics to only one. However, it is possible that a quatrain consists of more than one topic. The model learns different topics within a quatrain, but it has to generate quatrains in a restrictive way. Therefore, this study suggests to extend the topic constraint on the generative model.

Table 24

Comparison of BLEU-2 Evaluation

Sentence / Models	1-2	2-3	3-4	Average
SMT	0.0559	0.0410	0.0547	0.0505
RNNPG	0.0561	0.0515	0.0572	0.0549
QGM	0.0316	0.0823	0.0776	0.0638
This Study	0.1605	0.1838	0.1619	0.1687

4.5. Human Evaluation

Referring to the evaluation criteria in (He et al., 2012; Huang et al., 2020; Yi et al., 2017; Zhang & Lapata, 2014), this study designed three criteria: Fluency, Coherence, and Meaning, as listed below. We particularly excluded the criterion "poeticness "because it is for tonal pattern and rhyme testing.

(1) Fluency: Is the poetry grammatically correct?

(2) Coherence: Is the poetry thematically structured?

(3) Meaning: Does the poetry convey a meaningful message to reader?

We invited 11 senior high school students, who just finished General Scholastic Ability Test (GSAT, 大學入學學科能力測驗 in Chinese), to evaluate 10 generated quatrains. The reasons are that (1) senior high school students read more Chinese classical poetries than the students in junior high school, (2) The research domain of college students, graduate student, and experts probably is not Chinese classical poetry, but senior high school students have to read all kinds of Chinese literature in order to deal with the GSAT. Consequently, we chose senior high school students to be the testers.

They scored each criterion from 1 (totally disagree) to 5 (totally agree). The results of the human evaluation are shown in Table 25. Except for human, this study has a better score in each criterion. The possible reason is that previous researches trained the 5-character and 7- character quatrains together in on model. Their model cannot pay fully attention to learn the feature and knowledge of 5-character quatrains and then to add insult to injury, the most widely used words are from 7-character; therefore, it has a poor performance on poetry generation, BLEU-2 evaluation, and human evaluation. Table 26 shows the top 3 five-character generated quatrains.

Table 25

Comparison of Human Evaluation

Models	**Fluency**	**Coherence**	**Meaning**	**Average**
SMT	2.81	2.47	2.33	2.54
RNNPG	4.01	3.18	3.20	3.46
QGM	3.43	3.24	3.31	3.33
This Study	3.45	3.58	3.71	3.58
Human	4.31	3.81	3.61	3.91

Table 26

Top 3 Five-Character Generated Quatrain (Human Evaluation)

No.	Words from Word Dictionary	5-Character Generated Quatrain
1	鷗鳥、漁舟、 憑君、喚起	***鷗鳥*** 來何事，***漁舟*** 到亦迷。 ***憑君*** 莫拋擲，***喚起*** 故園心。
2	雲氣、山陰、 仙翁、但覺	***雲氣*** 蓬萊近，***山陰*** 草樹香。 ***仙翁*** 行道處，***但覺*** 綠沈沈。
3	夜月、傷心、 人不見、今日	***夜月*** 三杯色，***傷心*** 欲自迷。 泉存 ***人不見***，***今日*** 自魂銷。

Chapter 5 Conclusion and Future Work

5.1. Conclusion

Generating Chinese quatrains is a big challenge in NLP field and gains a lot of attractions. This study applied the qualified 5-character quatrains which follow the most common three-category tonal patterns and segmented the sentence by CCPF segmentation. We used the keyword extracted from each sentence by TextRank and sentences to train the model to learn the semantic meaning and sentence patterns. The generative model generates a 5-character quatrain according to 4 words sharing the same topic, which is processed by LDA, from individual word dictionaries. The result shows a great growth of the accuracy of tonal patterns for 5-character generated quatrain. Besides, the performances on BLEU-2 Evaluation and human evaluation are better than pervious researches.

There are three contributions of this study. First, we provided a beneficial standard of tonal pattern evaluation for 5-character quatrains. Second, we collected the tones and rhymes of 20,739 Chinese characters. Consequently, the non-professional populace is able to do metrical rules evaluation with reference to it and will not be led astray by some incomplete open sources and confused about difference of traditional Chinese and Mandarin. Finally, the result of this study demonstrated that the existing generative models probably are unable to well process 5-character quatrains and 7-character quatrains together since the wording and tonal patterns of them are different.

5.2. Future Work

This study proved the importance of the training sample that the quatrains with correct tonal patterns can actually increase the accuracy of the generated quatrains. However, there are three future works for the improvement of performance.

(1) The Distribution of Words in Each Sentence

Since the amount of words processed by CCPF Segmentation from 4 consecutive sentences are similar, it is worth to checking their distributions and intersections. We will understand which words are frequently or only placed in which sentence and redesign a better model structure.

(2) Paying Attention to the Tones and Position of Words

The segmented words with one to three characters in length would be misplaced by the model because the model put emphasis on learning the relationship and meaning between sentence, but does not fully understand the tones of each character. Therefore, this study presumed that adding the data of tones probably will increase the accuracy of tonal patterns for 5-character generated quatrain.

(3) Topic Constraint

Although the poetries are clustered into 28 categories, they actually might have more than one topic. It is worth a shot that allowing the generative model to have two or more topics in poetry generations.

References

Bailey, R. W. (1974). Computer-assisted poetry: the writing machine is for everybody. *Computers in the Humanities*, 283-295.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *the Journal of machine Learning research, 3*, 993-1022.

BYVoid. (2011). 韻典網. Retrieved from https://ytenx.org/

Chen, M. Y. (1979). Metrical structure: Evidence from Chinese poetry. *Linguistic Inquiry, 10*(3), 371-420.

Chinese Text Project (中國哲學書電子化計劃 in Chinese). (2006). Retrieved from https://ctext.org/zh

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Downer, G. B., & Graham, A. C. (1963). Tone patterns in Chinese poetry. *Bulletin of the School of Oriental and African Studies, University of London, 26*(1), 145-148.

Gervás, P. (2001). An expert system for the composition of formal spanish poetry. In *Applications and Innovations in Intelligent Systems VIII* (pp. 19-32): Springer.

Graham, A. C. (1980). STRUCTURE AND LICENSE IN CHINESE REGULATED VERSE/律诗的声调结构. *Journal of Chinese Linguistics*, 42-58.

He, J., Zhou, M., & Jiang, L. (2012). *Generating chinese classical poems with statistical machine translation models.* Paper presented at the Proceedings of the AAAI Conference on Artificial Intelligence.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation, 9*(8), 1735-1780.

Huang, C.-M. (2019). Exploring and Analyzing Ancient Poetry with LDA Topic Model Based on Lexical Chain and Regulated Verse Segmentation Methods (基於語彙鏈, 格律斷詞方法以主題模型進行古詩詞探勘與分析 in Chinese). *資訊管理學報, 26*(3), 275-306.

Huang, C.-M., Lu, K.-L., Cheng, Y.-Y., & Peng, Y.-C. (2020). *Generating Chinese Classical Poetry with Quatrain Generation Model (QGM) Using Encoder-Decoder LSTM.* Paper presented at the 2020 IEEE International Conference on Big Data (Big Data).

Jiang, L., & Zhou, M. (2008). *Generating Chinese couplets using a statistical MT approach.* Paper presented at the Proceedings of the 22nd International Conference on Computational Linguistics (Coling 2008).

Lin, S.-f. (2014). *The Nature of the Quatrain from the Late Han to the High T'ang*: Princeton University Press.

Liu, D., Guo, Q., Li, W., & Lv, J. (2018). *A multi-modal chinese poetry generation model.* Paper presented at the 2018 International Joint Conference on Neural Networks (IJCNN).

Liu, L., Wan, X., & Guo, Z. (2018). *Images2poem: Generating chinese poetry from image streams.* Paper presented at the Proceedings of the 26th ACM international conference on Multimedia.

LU, K.-L. (2020). *Generating Chinese Classical Poetry with Topic Using LSTM Encoder-Decoder.* (Master). National University of Science and Technology, Retrieved from https://hdl.handle.net/11296/5w3rvj

Manurung, H. (2004). An evolutionary algorithm approach to poetry generation.

Manurung, R., Ritchie, G., & Thompson, H. (2012). Using genetic algorithms to create meaningful poetic text. *Journal of Experimental & Theoretical Artificial Intelligence, 24*(1), 43-64.

Mihalcea, R., & Tarau, P. (2004). *Textrank: Bringing order into text.* Paper presented at the Proceedings of the 2004 conference on empirical methods in natural language processing.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.

Pulleyblank, E. G. (1999). Chinese traditional phonology. *Asia Major*, 101-137.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). *Sequence to sequence learning with neural networks.* Paper presented at the Advances in neural information processing systems.

Wang, L. (1979). *Shi Ci Ge Lu Gai Yao (《詩詞格律 詩詞格律概要》 in Chinese)*.

Wang, L. (1989). *Han yu shi lu xue (《汉语诗律学》 in Chinese, aka The Chinese Versification)*.

Wang, L. (2014). *Shi Ci Ge Lu Gai Yao (《詩詞格律 詩詞格律概要》 in Chinese)*.

Wang, L. (2015). *Han yu shi lu xue (《汉语诗律学》 in Chinese, aka The Chinese Versification)*.

Wang, N.-J., Tseng, S.-S., Yang, C.-C., Su, J.-M., & Lo, F.-J. (2015). *〈詩風規則之研究-以唐朝近體詩為例〉*.

Wang, Z., He, W., Wu, H., Wu, H., Li, W., Wang, H., & Chen, E. (2016). Chinese poetry generation with planning based neural network. *arXiv preprint arXiv:1610.09889*.

Yi, X., Li, R., & Sun, M. (2017). Generating chinese classical poems with rnn encoder-decoder. In *Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data* (pp. 211-223): Springer.

Zhang, X., & Lapata, M. (2014). *Chinese poetry generation with recurrent neural networks.* Paper presented at the Proceedings of the 2014 Conference on

Empirical Methods in Natural Language Processing (EMNLP).
搜韻. (2009). Retrieved from https://sou-yun.cn/about.aspx
漢字古今音資料庫. (2011). Retrieved from https://xiaoxue.iis.sinica.edu.tw/ccr

Appendices

A. Table of Proper Noun Translations

Chinese	English	Description
古體詩	Ancient-style poetry	Aka Ku-ti-shih
近體詩	Recent-style poetry	Aka Chin-ti-shih, modern-style poetry
絕句	Quatrain	Aka Chueh-chu One of Recent-style poetry
律絕	Regulated quatrain	One of quatrain
古絕	Ancient-style quatrain	One of quatrain
律詩	Regulated verse	Aka Lu-shih, a double quatrain One of Recent-style poetry
平仄格式	Tonal pattern	The core of the metrical rules
平聲	Level tone	Aka even tone or Ping sheng
上平聲	Shang Ping sheng	One of Level tone
下平聲	Xia Ping sheng	One of Level tone
仄聲	Oblique tone	Aka deflected tone or Ze sheng
上聲	Rising tone	Aka Shang sheng One of oblique tone
去聲	Departing tone	One of oblique tone, aka Qu sheng
入聲	Entering tone	One of oblique tone, aka Ru sheng
一聲	Tone 1	Aka high level tone, e.g. 媽 (ma) One of the substitutes for Level tone in Mandarin
二聲	Tone 2	Aka rising tone, e.g. 麻 (maˊ) One of the substitutes for Level tone in Mandarin
三聲	Tone 3	Aka falling-rising tone, e.g. 馬 (maˇ) The substitute for Rising tone in Mandarin
四聲	Tone 4	Aka falling tone, e.g. 罵 (maˋ) The substitute for Departing tone in Mandarin
輕聲	Light tone	Aka Qing sheng or neutral tone, e.g. 嗎（ma˙）in Mandarin
韻書	Rhyme book	A dictionary organized by rhymes
廣韻	Extended Rhymes	Aka Guangyun One of rhyme book

平水韻	Pingshui Rhymes	One of rhyme book
韻目	Rhyme group	Aka rhyme heading
小韻	Homophone group	The sub-groups of rhyme group
拗救	Counterbalance	A remedy for irregular misplaced tones
聯	Couplet	Two adjacent sentences are a couplet. There are 2 couplets in a quatrain.
對	Antithesis	The tonal relationship between two sentences within a couplet
黏	Consistency	The tonal relationship between two couplets

B. 唐宋五言絕句之常見平仄格式 (基本四式、變格與特殊格式)

編號	平仄符號
1	平平平仄仄 仄仄仄平平 仄仄平平仄 平平仄仄平
2	平平平仄仄 仄仄仄平平 平仄平平仄 平平仄仄平
3	平平平仄仄 平仄仄平平 仄仄平平仄 平平仄仄平
4	平平平仄仄 平仄仄平平 平仄平平仄 平平仄仄平
5	仄平平仄仄 仄仄仄平平 仄仄平平仄 平平仄仄平
6	仄平平仄仄 仄仄仄平平 平仄平平仄 平平仄仄平
7	仄平平仄仄 平仄仄平平 仄仄平平仄 平平仄仄平
8	仄平平仄仄 平仄仄平平 平仄平平仄 平平仄仄平
9	平平仄仄平 仄仄仄平平 仄仄平平仄 平平仄仄平
10	平平仄仄平 仄仄仄平平 平仄平平仄 平平仄仄平
11	平平仄仄平 平仄仄平平 仄仄平平仄 平平仄仄平
12	平平仄仄平 平仄仄平平 平仄平平仄 平平仄仄平
13	平平平仄仄 仄仄仄平平 仄仄仄平仄 平平平仄平
14	平平平仄仄 仄仄仄平平 平仄仄平仄 平平平仄平
15	平平平仄仄 平仄仄平平 仄仄仄平仄 平平平仄平
16	平平平仄仄 平仄仄平平 平仄仄平仄 平平平仄平
17	仄平平仄仄 仄仄仄平平 仄仄仄平仄 平平平仄平
18	仄平平仄仄 仄仄仄平平 平仄仄平仄 平平平仄平
19	仄平平仄仄 平仄仄平平 仄仄仄平仄 平平平仄平
20	仄平平仄仄 平仄仄平平 平仄仄平仄 平平平仄平

21	平平仄仄平 仄仄仄平平 仄仄仄平仄 平平平仄平
22	平平仄仄平 仄仄仄平平 平仄仄平仄 平平平仄平
23	平平仄仄平 平仄仄平平 仄仄仄平仄 平平平仄平
24	平平仄仄平 平仄仄平平 平仄仄平仄 平平平仄平
25	平平仄仄仄 仄仄仄平平 仄仄仄平仄 平平平仄平
26	平平仄仄仄 仄仄仄平平 平仄仄平仄 平平平仄平
27	平平仄仄仄 平仄仄平平 仄仄仄平仄 平平平仄平
28	平平仄仄仄 平仄仄平平 平仄仄平仄 平平平仄平
29	平平平仄仄 仄仄平平平 仄仄仄平仄 平平平仄平
30	平平平仄仄 仄仄平平平 平仄仄平仄 平平平仄平
31	平平平仄仄 平仄平平平 仄仄仄平仄 平平平仄平
32	平平平仄仄 平仄平平平 平仄仄平仄 平平平仄平
33	仄平平仄仄 仄仄平平平 仄仄仄平仄 平平平仄平
34	仄平平仄仄 仄仄平平平 平仄仄平仄 平平平仄平
35	仄平平仄仄 平仄平平平 仄仄仄平仄 平平平仄平
36	仄平平仄仄 平仄平平平 平仄仄平仄 平平平仄平
37	平平仄仄平 仄仄平平平 仄仄仄平仄 平平平仄平
38	平平仄仄平 仄仄平平平 平仄仄平仄 平平平仄平
39	平平仄仄平 平仄平平平 仄仄仄平仄 平平平仄平
40	平平仄仄平 平仄平平平 平仄仄平仄 平平平仄平
41	平平仄仄仄 仄仄平平平 仄仄仄平仄 平平平仄平
42	平平仄仄仄 仄仄平平平 平仄仄平仄 平平平仄平
43	平平仄仄仄 平仄平平平 仄仄仄平仄 平平平仄平

44	平平仄仄仄 平仄平平平 平仄仄平仄 平平平仄平
45	平平仄仄仄 仄仄仄平平 仄仄平平仄 平平仄仄平
46	平平仄仄仄 仄仄仄平平 平仄平平仄 平平仄仄平
47	平平仄仄仄 平仄仄平平 仄仄平平仄 平平仄仄平
48	平平仄仄仄 平仄仄平平 平仄平平仄 平平仄仄平
49	平平仄仄仄 仄仄平平平 仄仄平平仄 平平仄仄平
50	平平仄仄仄 仄仄平平平 平仄平平仄 平平仄仄平
51	平平仄仄仄 平仄平平平 仄仄平平仄 平平仄仄平
52	平平仄仄仄 平仄平平平 平仄平平仄 平平仄仄平
53	平平仄仄仄 仄仄仄平平 仄仄平平仄 平平平仄平
54	平平仄仄仄 仄仄仄平平 平仄平平仄 平平平仄平
55	平平仄仄仄 平仄仄平平 仄仄平平仄 平平平仄平
56	平平仄仄仄 平仄仄平平 平仄平平仄 平平平仄平
57	平平平仄仄 仄仄平平平 仄仄平平仄 平平仄仄平
58	平平平仄仄 仄仄平平平 平仄平平仄 平平仄仄平
59	平平平仄仄 平仄平平平 仄仄平平仄 平平仄仄平
60	平平平仄仄 平仄平平平 平仄平平仄 平平仄仄平
61	仄平平仄仄 仄仄平平平 仄仄平平仄 平平仄仄平
62	仄平平仄仄 仄仄平平平 平仄平平仄 平平仄仄平
63	仄平平仄仄 平仄平平平 仄仄平平仄 平平仄仄平
64	仄平平仄仄 平仄平平平 平仄平平仄 平平仄仄平
65	平平仄仄平 仄仄平平平 仄仄平平仄 平平仄仄平
66	平平仄仄平 仄仄平平平 平仄平平仄 平平仄仄平

67	平平仄仄平 平仄平平平 仄仄平平仄 平平仄仄平
68	平平仄仄平 平仄平平平 平仄平平仄 平平仄仄平
69	平平平仄仄 仄仄平平平 仄仄平平仄 平平平仄平
70	平平平仄仄 仄仄平平平 平仄平平仄 平平平仄平
71	平平平仄仄 平仄平平平 仄仄平平仄 平平平仄平
72	平平平仄仄 平仄平平平 平仄平平仄 平平平仄平
73	仄平平仄仄 仄仄平平平 仄仄平平仄 平平平仄平
74	仄平平仄仄 仄仄平平平 平仄平平仄 平平平仄平
75	仄平平仄仄 平仄平平平 仄仄平平仄 平平平仄平
76	仄平平仄仄 平仄平平平 平仄平平仄 平平平仄平
77	平平仄仄平 仄仄平平平 仄仄平平仄 平平平仄平
78	平平仄仄平 仄仄平平平 平仄平平仄 平平平仄平
79	平平仄仄平 平仄平平平 仄仄平平仄 平平平仄平
80	平平仄仄平 平仄平平平 平仄平平仄 平平平仄平
81	平平仄仄仄 仄仄平平平 仄仄平平仄 平平平仄平
82	平平仄仄仄 仄仄平平平 平仄平平仄 平平平仄平
83	平平仄仄仄 平仄平平平 仄仄平平仄 平平平仄平
84	平平仄仄仄 平仄平平平 平仄平平仄 平平平仄平
85	平平平仄仄 仄仄仄平平 仄仄平平仄 平平平仄平
86	平平平仄仄 仄仄仄平平 平仄平平仄 平平平仄平
87	平平平仄仄 平仄仄平平 仄仄平平仄 平平平仄平
88	平平平仄仄 平仄仄平平 平仄平平仄 平平平仄平
89	仄平平仄仄 仄仄仄平平 仄仄平平仄 平平平仄平

90	仄平平仄仄 仄仄仄平平 平仄平平仄 平平平仄平
91	仄平平仄仄 平仄仄平平 仄仄平平仄 平平平仄平
92	仄平平仄仄 平仄仄平平 平仄平平仄 平平平仄平
93	平平仄仄平 仄仄仄平平 仄仄平平仄 平平平仄平
94	平平仄仄平 仄仄仄平平 平仄平平仄 平平平仄平
95	平平仄仄平 平仄仄平平 仄仄平平仄 平平平仄平
96	平平仄仄平 平仄仄平平 平仄平平仄 平平平仄平
97	仄仄平平仄 平平仄仄平 平平平仄仄 仄仄仄平平
98	仄仄平平仄 平平仄仄平 平平平仄仄 平仄仄平平
99	仄仄平平仄 平平仄仄平 仄平平仄仄 仄仄仄平平
100	仄仄平平仄 平平仄仄平 仄平平仄仄 平仄仄平平
101	平仄平平仄 平平仄仄平 平平平仄仄 仄仄仄平平
102	平仄平平仄 平平仄仄平 平平平仄仄 平仄仄平平
103	平仄平平仄 平平仄仄平 仄平平仄仄 仄仄仄平平
104	平仄平平仄 平平仄仄平 仄平平仄仄 平仄仄平平
105	仄仄仄平平 平平仄仄平 平平平仄仄 仄仄仄平平
106	仄仄仄平平 平平仄仄平 平平平仄仄 平仄仄平平
107	仄仄仄平平 平平仄仄平 仄平平仄仄 仄仄仄平平
108	仄仄仄平平 平平仄仄平 仄平平仄仄 平仄仄平平
109	平仄仄平平 平平仄仄平 平平平仄仄 仄仄仄平平
110	平仄仄平平 平平仄仄平 平平平仄仄 平仄仄平平
111	平仄仄平平 平平仄仄平 仄平平仄仄 仄仄仄平平
112	平仄仄平平 平平仄仄平 仄平平仄仄 平仄仄平平

113	仄仄仄平仄 平平平仄平 平平平仄仄 仄仄仄平平
114	仄仄仄平仄 平平平仄平 平平平仄仄 平仄仄平平
115	仄仄仄平仄 平平平仄平 仄平平仄仄 仄仄仄平平
116	仄仄仄平仄 平平平仄平 仄平平仄仄 平仄仄平平
117	平仄仄平仄 平平平仄平 平平平仄仄 仄仄仄平平
118	平仄仄平仄 平平平仄平 平平平仄仄 平仄仄平平
119	平仄仄平仄 平平平仄平 仄平平仄仄 仄仄仄平平
120	平仄仄平仄 平平平仄平 仄平平仄仄 平仄仄平平
121	仄仄仄平仄 平平平仄平 平平仄仄仄 仄仄仄平平
122	仄仄仄平仄 平平平仄平 平平仄仄仄 平仄仄平平
123	平仄仄平仄 平平平仄平 平平仄仄仄 仄仄仄平平
124	平仄仄平仄 平平平仄平 平平仄仄仄 平仄仄平平
125	仄仄仄平仄 平平平仄平 平平平仄仄 仄仄平平平
126	仄仄仄平仄 平平平仄平 平平平仄仄 平仄平平平
127	仄仄仄平仄 平平平仄平 仄平平仄仄 仄仄平平平
128	仄仄仄平仄 平平平仄平 仄平平仄仄 平仄平平平
129	平仄仄平仄 平平平仄平 平平平仄仄 仄仄平平平
130	平仄仄平仄 平平平仄平 平平平仄仄 平仄平平平
131	平仄仄平仄 平平平仄平 仄平平仄仄 仄仄平平平
132	平仄仄平仄 平平平仄平 仄平平仄仄 平仄平平平
133	仄仄仄平仄 平平平仄平 平平仄仄仄 仄仄平平平
134	仄仄仄平仄 平平平仄平 平平仄仄仄 平仄平平平
135	平仄仄平仄 平平平仄平 平平仄仄仄 仄仄平平平

136	平仄仄平仄 平平平仄平 平平仄仄仄 平仄平平平
137	仄仄平平仄 平平仄仄平 平平仄仄仄 仄仄仄平平
138	仄仄平平仄 平平仄仄平 平平仄仄仄 平仄仄平平
139	平仄平平仄 平平仄仄平 平平仄仄仄 仄仄仄平平
140	平仄平平仄 平平仄仄平 平平仄仄仄 平仄仄平平
141	仄仄仄平平 平平仄仄平 平平仄仄仄 仄仄仄平平
142	仄仄仄平平 平平仄仄平 平平仄仄仄 平仄仄平平
143	平仄仄平平 平平仄仄平 平平仄仄仄 仄仄仄平平
144	平仄仄平平 平平仄仄平 平平仄仄仄 平仄仄平平
145	仄仄平平仄 平平平仄平 平平仄仄仄 仄仄仄平平
146	仄仄平平仄 平平平仄平 平平仄仄仄 平仄仄平平
147	平仄平平仄 平平平仄平 平平仄仄仄 仄仄仄平平
148	平仄平平仄 平平平仄平 平平仄仄仄 平仄仄平平
149	仄仄仄平平 平平平仄平 平平仄仄仄 仄仄仄平平
150	仄仄仄平平 平平平仄平 平平仄仄仄 平仄仄平平
151	平仄仄平平 平平平仄平 平平仄仄仄 仄仄仄平平
152	平仄仄平平 平平平仄平 平平仄仄仄 平仄仄平平
153	仄仄平平仄 平平仄仄平 平平平仄仄 仄仄平平平
154	仄仄平平仄 平平仄仄平 平平平仄仄 平仄平平平
155	仄仄平平仄 平平仄仄平 仄平平仄仄 仄仄平平平
156	仄仄平平仄 平平仄仄平 仄平平仄仄 平仄平平平
157	平仄平平仄 平平仄仄平 平平平仄仄 仄仄平平平
158	平仄平平仄 平平仄仄平 平平平仄仄 平仄平平平

159	平仄平平仄 平平仄仄平 仄平平仄仄 仄仄平平平
160	平仄平平仄 平平仄仄平 仄平平仄仄 平仄平平平
161	仄仄仄平平 平平仄仄平 平平平仄仄 仄仄平平平
162	仄仄平平平 平平仄仄平 平平平仄仄 仄仄仄平平
163	仄仄平平平 平平仄仄平 平平平仄仄 仄仄平平平
164	仄仄仄平平 平平仄仄平 平平平仄仄 平仄平平平
165	仄仄平平平 平平仄仄平 平平平仄仄 平仄仄平平
166	仄仄平平平 平平仄仄平 平平平仄仄 平仄平平平
167	仄仄仄平平 平平仄仄平 仄平平仄仄 仄仄平平平
168	仄仄平平平 平平仄仄平 仄平平仄仄 仄仄仄平平
169	仄仄平平平 平平仄仄平 仄平平仄仄 仄仄平平平
170	仄仄仄平平 平平仄仄平 仄平平仄仄 平仄平平平
171	仄仄平平平 平平仄仄平 仄平平仄仄 平仄仄平平
172	仄仄平平平 平平仄仄平 仄平平仄仄 平仄平平平
173	平仄仄平平 平平仄仄平 平平平仄仄 仄仄平平平
174	平仄平平平 平平仄仄平 平平平仄仄 仄仄仄平平
175	平仄平平平 平平仄仄平 平平平仄仄 仄仄平平平
176	平仄仄平平 平平仄仄平 平平平仄仄 平仄平平平
177	平仄平平平 平平仄仄平 平平平仄仄 平仄仄平平
178	平仄平平平 平平仄仄平 平平平仄仄 平仄平平平
179	平仄仄平平 平平仄仄平 仄平平仄仄 仄仄平平平
180	平仄平平平 平平仄仄平 仄平平仄仄 仄仄仄平平
181	平仄平平平 平平仄仄平 仄平平仄仄 仄仄平平平

182	平仄仄平平 平平仄仄平 仄平平仄仄 平仄平平平
183	平仄平平平 平平仄仄平 仄平平仄仄 平仄仄平平
184	平仄平平平 平平仄仄平 仄平平仄仄 平仄平平平
185	仄仄平平仄 平平仄仄平 平平仄仄仄 仄仄平平平
186	仄仄平平仄 平平仄仄平 平平仄仄仄 平仄平平平
187	平仄平平仄 平平仄仄平 平平仄仄仄 仄仄平平平
188	平仄平平仄 平平仄仄平 平平仄仄仄 平仄平平平
189	仄仄仄平平 平平仄仄平 平平仄仄仄 仄仄平平平
190	仄仄平平平 平平仄仄平 平平仄仄仄 仄仄仄平平
191	仄仄平平平 平平仄仄平 平平仄仄仄 仄仄平平平
192	仄仄仄平平 平平仄仄平 平平仄仄仄 平仄平平平
193	仄仄平平平 平平仄仄平 平平仄仄仄 平仄仄平平
194	仄仄平平平 平平仄仄平 平平仄仄仄 平仄平平平
195	平仄仄平平 平平仄仄平 平平仄仄仄 仄仄平平平
196	平仄平平平 平平仄仄平 平平仄仄仄 仄仄仄平平
197	平仄平平平 平平仄仄平 平平仄仄仄 仄仄平平平
198	平仄仄平平 平平仄仄平 平平仄仄仄 平仄平平平
199	平仄平平平 平平仄仄平 平平仄仄仄 平仄仄平平
200	平仄平平平 平平仄仄平 平平仄仄仄 平仄平平平
201	仄仄平平仄 平平平仄平 平平平仄仄 仄仄平平平
202	仄仄平平仄 平平平仄平 平平平仄仄 平仄平平平
203	仄仄平平仄 平平平仄平 仄平平仄仄 仄仄平平平
204	仄仄平平仄 平平平仄平 仄平平仄仄 平仄平平平

205	平仄平平仄 平平平仄平 平平平仄仄 仄仄平平平
206	平仄平平仄 平平平仄平 平平平仄仄 平仄平平平
207	平仄平平仄 平平平仄平 仄平平仄仄 仄仄平平平
208	平仄平平仄 平平平仄平 仄平平仄仄 平仄平平平
209	仄仄仄平平 平平平仄平 平平平仄仄 仄仄平平平
210	仄仄平平平 平平平仄平 平平平仄仄 仄仄仄平平
211	仄仄平平平 平平平仄平 平平平仄仄 仄仄平平平
212	仄仄仄平平 平平平仄平 平平平仄仄 平仄平平平
213	仄仄平平平 平平平仄平 平平平仄仄 平仄仄平平
214	仄仄平平平 平平平仄平 平平平仄仄 平仄平平平
215	仄仄仄平平 平平平仄平 仄平平仄仄 仄仄平平平
216	仄仄平平平 平平平仄平 仄平平仄仄 仄仄仄平平
217	仄仄平平平 平平平仄平 仄平平仄仄 仄仄平平平
218	仄仄仄平平 平平平仄平 仄平平仄仄 平仄平平平
219	仄仄平平平 平平平仄平 仄平平仄仄 平仄仄平平
220	仄仄平平平 平平平仄平 仄平平仄仄 平仄平平平
221	平仄仄平平 平平平仄平 平平平仄仄 仄仄平平平
222	平仄平平平 平平平仄平 平平平仄仄 仄仄仄平平
223	平仄平平平 平平平仄平 平平平仄仄 仄仄平平平
224	平仄仄平平 平平平仄平 平平平仄仄 平仄平平平
225	平仄平平平 平平平仄平 平平平仄仄 平仄仄平平
226	平仄平平平 平平平仄平 平平平仄仄 平仄平平平
227	平仄仄平平 平平平仄平 仄平平仄仄 仄仄平平平

228	平仄平平平 平平平仄平 仄平平仄仄 仄仄仄平平
229	平仄平平平 平平平仄平 仄平平仄仄 仄仄平平平
230	平仄仄平平 平平平仄平 仄平平仄仄 平仄平平平
231	平仄平平平 平平平仄平 仄平平仄仄 平仄仄平平
232	平仄平平平 平平平仄平 仄平平仄仄 平仄平平平
233	仄仄平平仄 平平平仄平 平平仄仄仄 仄仄平平平
234	仄仄平平仄 平平平仄平 平平仄仄仄 平仄平平平
235	平仄平平仄 平平平仄平 平平仄仄仄 仄仄平平平
236	平仄平平仄 平平平仄平 平平仄仄仄 平仄平平平
237	仄仄仄平平 平平平仄平 平平仄仄仄 仄仄平平平
238	仄仄平平平 平平平仄平 平平仄仄仄 仄仄仄平平
239	仄仄平平平 平平平仄平 平平仄仄仄 仄仄平平平
240	仄仄仄平平 平平平仄平 平平仄仄仄 平仄平平平
241	仄仄平平平 平平平仄平 平平仄仄仄 平仄仄平平
242	仄仄平平平 平平平仄平 平平仄仄仄 平仄平平平
243	平仄仄平平 平平平仄平 平平仄仄仄 仄仄平平平
244	平仄平平平 平平平仄平 平平仄仄仄 仄仄仄平平
245	平仄平平平 平平平仄平 平平仄仄仄 仄仄平平平
246	平仄仄平平 平平平仄平 平平仄仄仄 平仄平平平
247	平仄平平平 平平平仄平 平平仄仄仄 平仄仄平平
248	平仄平平平 平平平仄平 平平仄仄仄 平仄平平平
249	仄仄平平仄 平平平仄平 平平平仄仄 仄仄仄平平
250	仄仄平平仄 平平平仄平 平平平仄仄 平仄仄平平

251	仄仄平平仄 平平平仄平 仄平平仄仄 仄仄仄平平
252	仄仄平平仄 平平平仄平 仄平平仄仄 平仄仄平平
253	平仄平平仄 平平平仄平 平平平仄仄 仄仄仄平平
254	平仄平平仄 平平平仄平 平平平仄仄 平仄仄平平
255	平仄平平仄 平平平仄平 仄平平仄仄 仄仄仄平平
256	平仄平平仄 平平平仄平 仄平平仄仄 平仄仄平平
257	仄仄仄平平 平平平仄平 平平平仄仄 仄仄仄平平
258	仄仄仄平平 平平平仄平 平平平仄仄 平仄仄平平
259	仄仄仄平平 平平平仄平 仄平平仄仄 仄仄仄平平
260	仄仄仄平平 平平平仄平 仄平平仄仄 平仄仄平平
261	平仄仄平平 平平平仄平 平平平仄仄 仄仄仄平平
262	平仄仄平平 平平平仄平 平平平仄仄 平仄仄平平
263	平仄仄平平 平平平仄平 仄平平仄仄 仄仄仄平平
264	平仄仄平平 平平平仄平 仄平平仄仄 平仄仄平平
265	平平仄平仄 仄仄仄平平 仄仄平平仄 平平仄仄平
266	平平仄平仄 仄仄仄平平 平仄平平仄 平平仄仄平
267	平平仄平仄 平仄仄平平 仄仄平平仄 平平仄仄平
268	平平仄平仄 平仄仄平平 平仄平平仄 平平仄仄平
269	平平仄平仄 仄仄仄平平 仄仄仄平仄 平平平仄平
270	平平仄平仄 仄仄仄平平 平仄仄平仄 平平平仄平
271	平平仄平仄平仄仄平平 仄仄仄平仄 平平平仄平
272	平平仄平仄 平仄仄平平 平仄仄平仄 平平平仄平
273	平平仄平仄 仄仄平平平 仄仄仄平仄 平平平仄平

274	平平仄平仄 仄仄平平平 平仄仄平仄 平平平仄平
275	平平仄平仄 平仄平平平 仄仄仄平仄 平平平仄平
276	平平仄平仄 平仄平平平 平仄仄平仄 平平平仄平
277	平平仄平仄 仄仄平平平 仄仄平平仄 平平仄仄平
278	平平仄平仄 仄仄平平平 平仄平平仄 平平仄仄平
279	平平仄平仄 平仄平平平 仄仄平平仄 平平仄仄平
280	平平仄平仄 平仄平平平 平仄平平仄 平平仄仄平
281	平平仄平仄 仄仄仄平平 仄仄平平仄 平平平仄平
282	平平仄平仄 仄仄仄平平 平仄平平仄 平平平仄平
283	平平仄平仄 平仄仄平平 仄仄平平仄 平平平仄平
284	平平仄平仄 平仄仄平平 平仄平平仄 平平平仄平
285	平平仄平仄 仄仄平平平 仄仄平平仄 平平平仄平
286	平平仄平仄 仄仄平平平 平仄平平仄 平平平仄平
287	平平仄平仄 平仄平平平 仄仄平平仄 平平平仄平
288	平平仄平仄 平仄平平平 平仄平平仄 平平平仄平
289	仄仄平平仄 平平仄仄平 平平仄平仄 仄仄仄平平
290	仄仄平平仄 平平仄仄平 平平仄平仄 平仄仄平平
291	平仄平平仄 平平仄仄平 平平仄平仄 仄仄仄平平
292	平仄平平仄 平平仄仄平 平平仄平仄 平仄仄平平
293	仄仄仄平平 平平仄仄平 平平仄平仄 仄仄仄平平
294	仄仄仄平平 平平仄仄平 平平仄平仄 平仄仄平平
295	平仄仄平平 平平仄仄平 平平仄平仄 仄仄仄平平
296	平仄仄平平 平平仄仄平 平平仄平仄 平仄仄平平

297	仄仄仄平仄 平平平仄平 平平仄平仄 仄仄仄平平
298	仄仄仄平仄 平平平仄平 平平仄平仄 平仄仄平平
299	平仄仄平仄 平平平仄平 平平仄平仄 仄仄仄平平
300	平仄仄平仄 平平平仄平 平平仄平仄 平仄仄平平
301	仄仄仄平仄 平平平仄平 平平仄平仄 仄仄平平平
302	仄仄仄平仄 平平平仄平 平平仄平仄 平仄平平平
303	平仄仄平仄 平平平仄平 平平仄平仄 仄仄平平平
304	平仄仄平仄 平平平仄平 平平仄平仄 平仄平平平
305	仄仄平平仄 平平平仄平 平平仄平仄 仄仄仄平平
306	仄仄平平仄 平平平仄平 平平仄平仄 平仄仄平平
307	平仄平平仄 平平平仄平 平平仄平仄 仄仄仄平平
308	平仄平平仄 平平平仄平 平平仄平仄 平仄仄平平
309	仄仄仄平平 平平平仄平 平平仄平仄 仄仄仄平平
310	仄仄仄平平 平平平仄平 平平仄平仄 平仄仄平平
311	平仄仄平平 平平平仄平 平平仄平仄 仄仄仄平平
312	平仄仄平平 平平平仄平 平平仄平仄 平仄仄平平
313	仄仄平平仄 平平仄仄平 平平仄平仄 仄仄平平平
314	仄仄平平仄 平平仄仄平 平平仄平仄 平仄平平平
315	平仄平平仄 平平仄仄平 平平仄平仄 仄仄平平平
316	平仄平平仄 平平仄仄平 平平仄平仄 平仄平平平
317	仄仄仄平平 平平仄仄平 平平仄平仄 仄仄平平平
318	仄仄平平平 平平仄仄平 平平仄平仄 仄仄仄平平
319	仄仄平平平 平平仄仄平 平平仄平仄 仄仄平平平

320	仄仄仄平平 平平仄仄平 平平仄平仄 平仄平平平
321	仄仄平平平 平平仄仄平 平平仄平仄 平仄仄平平
322	仄仄平平平 平平仄仄平 平平仄平仄 平仄平平平
323	平仄仄平平 平平仄仄平 平平仄平仄 仄仄平平平
324	平仄平平平 平平仄仄平 平平仄平仄 仄仄仄平平
325	平仄平平平 平平仄仄平 平平仄平仄 仄仄平平平
326	平仄仄平平 平平仄仄平 平平仄平仄 平仄平平平
327	平仄平平平 平平仄仄平 平平仄平仄 平仄仄平平
328	平仄平平平 平平仄仄平 平平仄平仄 平仄平平平
329	仄仄平平仄 平平平仄平 平平仄平仄 仄仄平平平
330	仄仄平平仄 平平平仄平 平平仄平仄 平仄平平平
331	平仄平平仄 平平平仄平 平平仄平仄 仄仄平平平
332	平仄平平仄 平平平仄平 平平仄平仄 平仄平平平
333	仄仄仄平平 平平平仄平 平平仄平仄 仄仄平平平
334	仄仄平平平 平平平仄平 平平仄平仄 仄仄仄平平
335	仄仄平平平 平平平仄平 平平仄平仄 仄仄平平平
336	仄仄仄平平 平平平仄平 平平仄平仄平仄平平平
337	仄仄平平平 平平平仄平 平平仄平仄 平仄仄平平
338	仄仄平平平 平平平仄平 平平仄平仄 平仄平平平
339	仄仄仄平平 平平平仄平 平平仄平仄 平仄平平平
340	平仄仄平平 平平平仄平 平平仄平仄 仄仄平平平
341	平仄平平平 平平平仄平 平平仄平仄 仄仄仄平平
342	平仄平平平 平平平仄平 平平仄平仄 仄仄平平平

343	平仄仄平平 平平平仄平 平平仄平仄 平仄平平平
344	平仄平平平 平平平仄平 平平仄平仄 平仄仄平平
345	平仄平平平 平平平仄平 平平仄平仄 平仄平平平