

Automatic Interpretation of Chinese Traditional Musical Notation Using Conditional Random Field

Rongfeng Li¹, Yelei Ding¹, Wenxin Li¹ and Minghui Bi²,

¹ Key Laboratory of Machine Perception (Ministry of Education), Peking University

²School of Arts, Peking University

rongfeng, dingyelei, lwx, biminghui@pku.edu.cn

Abstract. For the majority of Chinese people, Gongchepu, which is the Chinese traditional musical notation, is difficult to understand. Tragically, there are fewer and fewer experts who can read Gongchepu. Our work aims to interpret Gongchepu automatically into western musical notation-staff, which is more easily accepted by the public. The interpretation consists of two parts: pitch interpretation and rhythm interpretation. The pitch interpretation is easily to solve because there is a certain correspondence between the pitch notation of Gongchepu and staff. However, the rhythm notations of Gongchepu cannot be interpreted to the corresponding notations of staff because Gongchepu only denotes ban (strong-beat) and yan (off-beat), and the notations of duration are not taken down. In this paper, we proposed an automatic interpretation model based on Conditional Random Field. Our automatic interpretation method successfully achieves 96.81% precision and 90.59% oov precision on a database of published manually interpretation of Gongchepu.

Keywords: Musical notation, Gongchepu, interpretation, nature language processing, Conditional Random Field

1 Introduction

Chinese poetic songs are noted by *gongchepu*-Chinese traditional musical notation, once popular in ancient China and still used for traditional Chinese musical instruments and Chinese operas nowadays. A *Gongchepu* sample of Chinese poetic songs entitled 天淨沙 *Tian-jin-sha* is shown in Figure1.

As illustrated in Figure 1, the melodic notations of *Gongchepu* are noted at the right side of the lyrics, consisted of pitch notation and rhythm notations, which are the two basic characters of a musical notation. Therefore, the interpretation consists two sections, one is pitch interpretation and the other is rhythm interpretation.

This work is supported by the NSFC(No. 60933004).

The figure shows a vertical column of Chinese characters with Gongchepu notation above them. The notation consists of 10 characters: 合, 四, 一, 上, 尺, 工, 凡, 六, 五, 乙. The lyrics are: 天 涯 風 瘦 馬, 夕 陽 西 下, 斷 腸 人 在, 枯 藤 老 樹 昏 鴉, 小 橋 流 水 平 沙, 古. The characters '夕陽西下' and '流水' are highlighted with boxes. Arrows point from these boxes to the labels 'Notations of gongchepu' and 'Lyrics' respectively.

Figure 1. Gongchepu of Tian-jin-sha

For the pitch interpretation, we firstly introduce the details of pitch notations of *gongchepu*. Pitch of each note in *gongchepu* is denoted by 10 Chinese characters: 合 *hé*, 四 *sì* 一 *yī*, 上 *shàng*, 尺 *chě*, 工 *gōng*, 凡 *fán*, 六 *liù*, 五 *wǔ*, 乙 *yǐ*. They are equivalent to the notes of solfège system: sol, la, ti, do, re, mi, fa, sol, la, ti. 合 *hé*, 四 *sì* 一 *yī* are pitched an octave lower 六 *liù*, 五 *wǔ*, 乙 *yǐ*. *gongchepu* is named by the character 工 *gōng* and 尺 *chě*.

Once we take 上 *shàng* as the fixed pitch c^1 , the range of the 10 characters is $g-b^1$. *Gongchepu* uses the following notations to note other notes in different octaves:

- a) Octaves higher: a radical “亠” is added for one octave higher. For example, we use “𠂇” to represent an octave higher “上”. Similarly, the radical “亠” is added to represent two octaves higher.
- b) Octaves lower: an attached stroke is added to the ending of stroke of the character to note an octave lower. For example, we use “𠂇” to show an octave lower “上”.

Likely, two attached parts are added to represent two octaves lower.

Based on the rule above, the pitch notations of *gongchepu* can be interpreted directly to the corresponding notations of staff.

For the rhythm interpretation, we explain the rhythmic rules of *gongchepu*. *gongchepu* denote the beats by the following notations: The mark “、” represents the stronger-beat which is called *ban*, while the notation “。” represents the off-beat called *yan*. The marks are put at the upper right corner of the first note of a beat. Illustrated from Figure 2 which is written horizontally for convenient reading, we can see the notes separated into beats with the *ban* and *yan*.

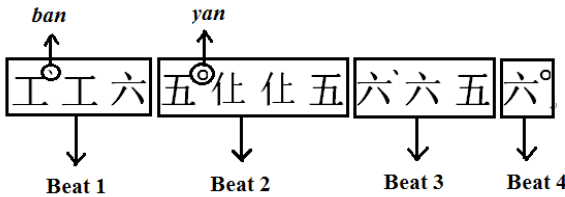




Figure 2. Ban and yan in *gongchepu*

Rhythmic structure of *gongchepu* is formed by the regular combination of *ban* and *yan*. For example, the cycle of 1 *ban* and 1 *yan* forms a 2/4 meter and cycle of 1 *ban* and 3 *yan* forms a 4/4 meter. However, the duration of each note, which should be noted in staff, cannot be specified by the rhythmic mark of *ban* and *yan*. In this case, the rhythm notations cannot be interpreted to the exclusive corresponding notations.

For example, if 2 notes are in 1 beat, it can be sung as . If 3

notes are in 1 beat, we could get 4 results: . But whichever should be sung is not restrict by the rhythmic rules of *gongchepu* and can be improvised by the singers. Does this mean that the rhythm in Chinese music is not important as Sachs [1] suggested in his studies of the rhythms of world music? Yang [2] corrects this misconception with the view that in order to perform the music in a proper way, the improvisations should have a certain fixed pattern. In other words, the rhythm of Chinese traditional music does have a certain pattern while the notation of duration of each note cannot be seen in the *gongchepu*.

Despite of all the analysis of the organizational structure of Chinese poetic songs in the past years, almost nothing has been published on the internal rhythmic structure. This is because there are few experts can read *gongchepu* nowadays, and they only teach a small group of students face to face.

In this paper, we proposed a stochastic model to interpret *gongchepu* into staff automatically. Dealing with the rhythm rules of *gongchepu*, the interpretation is similar to part-of-speech tagging in Natural Language Processing. This allows us to use Conditional Random Field to solve the interpretation problem. In recent years, a few musical notation researchers such as Qian [3] and Zhou [4] published their interpretation of the Chinese poetic songs collection, where the *gongchepu* is originally used. We implement our interpretation model on a database their published manually interpretation.

The rest of this paper is structured as follows. We begin with modeling the interpretation problem in section 2. Section 3 introduces the features for the statistical model. Section 4 provides the experimental settings and results. Finally, we draw the conclusion and future discussion in section 5.

2 Automatic Gongchepu Interpretation Model based on Conditional Random Field

In this section, we firstly formulate the interpretations problem. With the formulation, the interpretation problem is transform to a sequence tagging problem which is similar in natural language processing. Then we introduce the most widely used natural language processing model including Hidden Markov Model and Conditional Random Field to solve the interpretation problem.

2.1 Formulations of Rhythm Interpretation

We begin to formulate the interpretation problem by reviewing the rhythm rules of *gongchepu*. The rhythm marks including *ban* and *yan* are put at the upper right corner of the first note of a beat. Thus, notes are separated into beats with the *ban* and *yan*. We denote the beat sequence by B_1, B_2, \dots, B_n . Taking the ‘‘Tune of Fresh Flowers’’ as an example, beats separations are shown in Figure 3.

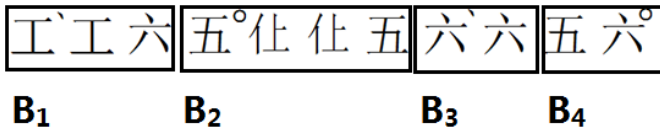


Figure 3. Beat separation by marks of *ban* and *yan*

However, the duration of each note, which should be noted in staff, cannot be specified by the rhythmic mark of *ban* and *yan*. In this case, the rhythm notations cannot be interpreted to the exclusive corresponding notations. For example, if 2

notes are in 1 beat, it can be sung as . We indicate the rhythm pattern of each beat by R_1, R_2, \dots, R_n .

Interpret the notes beat by beat, the interpretation task is illustrated in Figure 4.

In spite of the missing information of the duration of each note, the length of note duration in a beat is relatively fixed. Thus, rhythm patterns of each beat are limited. In

this paper, we conclude 37 patterns p_1, p_2, \dots, p_{37} which are used in Chinese poetic music. Thus, the value of $R_i, i=1, 2, \dots, n$ is limited in the patterns set $P=\{p_1, p_2, \dots, p_{37}\}$.

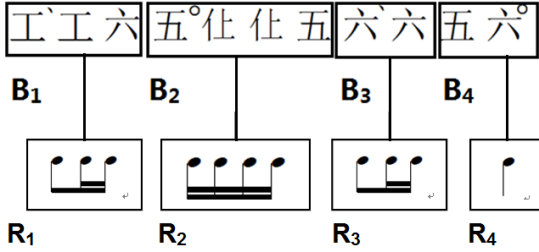


Figure 4. Interpret the rhythm beat by beat

By the above denotations, the interpretation transform to a tagging problem: when the beats sequence $\{B_1, B_2, \dots, B_n\}$ is observed, we are required to tag the sequence by the rhythm patterns from a limited set P . This is very similar to the sequence tagging problem in natural language processing.

Once the features $F(B_i)=\{f_1(B_i), f_2(B_i), \dots, f_m(B_i)\}$ of each beat are extracted, statistical language processing models such as Conditional Random Field can be applied to the interpretation.

2.2 Hidden Markov Model

HMM is well-understood, versatile and have been successful in handling text-based problem including POS tagging Kupiec[5], named entity recognition (Bikel[6]) and information extraction (Freitag & McCallum[7]). In the rhythm interpretation, the HMM is constructed based on the following assumptions: a) The rhythm pattern sequence $\{R_1, R_2, \dots, R_n\}$ forms a Markov Chain; b) The beats B_1, B_2, \dots, B_n are independent; c) for each rhythm pattern R_i , it only depends on its corresponding beat B_i . The graphical structure of HMM is shown in Figure 5.

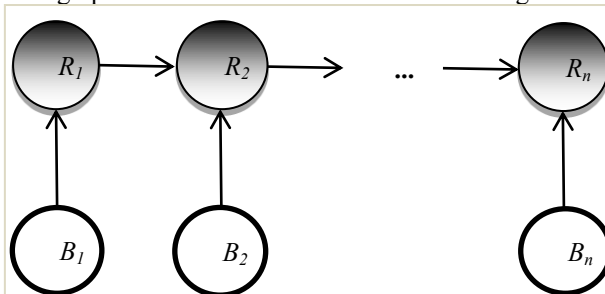


Figure 5. Graphical structure of HMM in rhythm interpretation

2.3 Conditional Random Field

Dealing with the multiple interacting features and long-range dependencies of observation problems, we would be inclined to use Conditional Random Field which is introduced by Lafferty et al [8]. Conditional Random Field have been proven to be efficient in handling different language POS tagging such as Chinese (Hong, Zhang, et al.[9]), Bengali(Ekbal, Haque, et al.[10]) and Tamil(Pandian & Geetha[11]), etc. Compare to HMM, CRF can handle the following undirected graphical structure which is shown in Figure 6.

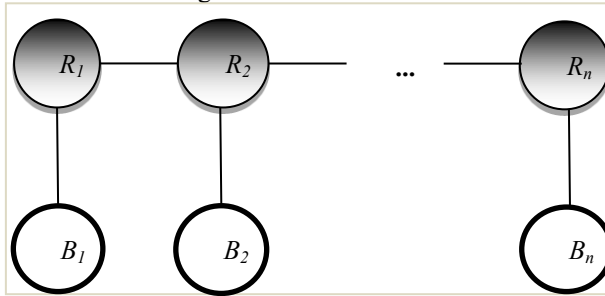


Figure 6. Graphical structure of CRF in rhythm interpretation

Conditional Random Fields are undirected graphic models. Giving an undirected graph $G=(V,E)$. Let C be the set of cliques (fully connected subsets) in the graph. Take the vertex of V as random variable we define the joint distribution of the vertex of V as follows:

$$P(V) = \frac{1}{Z} \prod_{c \in C} \Psi(X_c) \quad (1)$$

Here, X_c is the vertex set of a clique $c \in C$ and Z is the normalizing partition function. Ψ is called a potential function of c . The potential function can be described as the following exponential form:

$$\Psi(X_c) = \exp\left(\sum_i \lambda_i f_i(X_c)\right) \quad (2)$$

In the above model, the undirected graph consists of observations B_1, B_2, \dots, B_n and states R_1, R_2, \dots, R_n . Cliques from the above graph consist of two consecutive vertexes which are separated into two classifications: vertex of two consecutive states R_{i-1}, R_i and vertex of each states R_i and its corresponding observation B_i . Thus, the exponential form of potential functions can be denoted as the following two functions:

$$\Psi(R_{i-1}, R_i) = \exp\left(\sum_k \lambda_k f_k(R_{i-1}, R_i)\right) \quad (3)$$

and

$$\Psi(R_i, B_i) = \exp\left(\sum_k \mu_k g_k(R_i, B_i)\right) \quad (4)$$

According to the definition of (1), we get the conditional probability distribution:

$$P(R | B) = \frac{P(R, B)}{P(B)} = \frac{\frac{1}{Z} \prod_{i=2}^T \Psi(R_{i-1}, R_i) \prod_{j=2}^T \Psi(R_j, B_j)}{\sum_S \left\{ \frac{1}{Z} \prod_{i=2}^T \Psi(R_{i-1}, R_i) \prod_{j=2}^T \Psi(R_j, B_j) \right\}} \quad (5)$$

Denoting:

$$Z(B) = \sum_S \left\{ \frac{1}{Z} \prod_{i=2}^T \Psi(R_{i-1}, R_i) \prod_{j=2}^T \Psi(R_j, B_j) \right\} \quad (6)$$

(5) can be written as:



$$P(R | B) = \frac{1}{Z(B)} \exp\left\{ \sum_i \sum_k \lambda_k f_k(R_{i-1}, R_i) + \sum_k \mu_k g_k(R_i, B_i) \right\} \quad (7)$$

Here f_k is the feature function and g_k is the state feature functions. $\lambda_1, \lambda_2, \dots, \lambda_T, \mu_1, \mu_2, \dots, \mu_T$ are parameters to be estimated from training data.

To apply the above models, we should extract the features of each beat, which are discussed in the following section.

3 Feature Selection for Automatic Interpretation

Wise choice of the features is always vital to the performance of the statistical models. Chinese traditional music does not have harmony, polyphony, or texture. Thus, we only concern about the melody and select the proper features based on the opinions of the Chinese opera performance as follows.

- **Notes Sequence (NS):** The higher and lower octave symbols expand the 10 characters in *gongche* into 38 characters. Encoding these characters, we can get the original text features of the notes sequences.
- **Numbers of the Notes(NN):** Sequence of the notes numbers forms the approximately rhythmic structure. Rhythmic pattern is usually related to the notes number of previous beat. In the example of “Tune of Fresh Flowers” in figure 5, we consider the third beat “六六五” which is a three-note beat and the previous beat has four notes. Therefore, it preferred to determine the rhythmic pattern as  rather than  to avoid a too compact rhythmic structure.
- **Pitch Interval Direction and Position(PIDP):** The concept of “interval direction and position” is introduced by Williams(1997) for melodic analysis. Williams use “+” for rising direction of the pitch interval and “-” for the falling direction. Moreover, pitch interval is measured by

chromatic scale. For example, the pitch interval direction and position of the section of “Tune of Fresh Flowers” is illustrated in Figure 7.



Figure 7. Pitch interval direction and position of “Tune of Fresh Flowers”

4 Experimental Result

The experiments of *gongchepu* interpretation were based on the *gongchepu* of *Sui-jin-Ci-pu* collected by Xie[12] which collected poetic songs of Tang, Song and Yuan Dynasties of ancient China. *Sui-jin-Ci-pu* collected over 800 songs, but only a few of them have been interpreted. We trained our statistical models based on Qian [5]’s manually interpretation. We selected 60 songs from the 96 of Qian’s interpretation to set up our database. The database included 969 melody segments and amounted to 6347 beats. According to the different number of notes within a beat, the beats were separated into 6 types. The dataset was randomly divided into two parts with similar distribution of different types of beats. 3174 beats were used as training data while the left 3173 were reserved for test.

Table 1: Data size of *gongchepu*

Numbers of notes with in a beat	Trainin g data size	Testing data size	Total data size
1	1187	1017	2204
2	1110	1322	2432
3	647	676	1323
4	210	152	362
5	19	5	24
6	1	1	2
Total	3174	3173	6347



Table 1 shows the data size of the *gongchepu* for training and testing. In the table, we can see there are only 24 beats with 5 notes and 2 beats with 6 notes. 99.59% of beats in the dataset have more than 4 notes.

Two method Hidden Markov Model (HMM) and Conditional Random Field (CRF) which were introduced in Section 2 are applied using three single features: notes sequence (NS), numbers of notes (NN), pitch interval position and direction (PIDP) and their combinations: NS+NN, NN+PIDP, NS+PIDP, NS+NN+PIDP. The experimental results of interpretation precision and oov precisions are shown in Table 2.

Table 2. Interpretation precision and oov precisions

Features	precision		oov precision	
	HMM	CRF	HMM	CRF
NS	84.34%	87.86%	47.85%	67.62%
NN	83.43%	85.55%	68.43%	78.84%
PIDP	84.82%	85.97%	57.92%	77.53%
NS+NN	85.64%	89.67%	75.67%	80.23%
NN+PIDP	86.74%	89.56%	77.28%	81.55%
NS+PIDP	85.49%	89.89%	76.42%	79.88%
NS+NN+PIDP	87.38%	90.05%	78.27%	82.03%

The results from table 2 shows that CRF get better performance than HMM and achieve 90.05% precision and 82.03% oov precisions using the combination feature of NS+NN+PIDP.

We analyzed the oov beat and found that most interpretation error occurred in handling the beats which have 3 notes. For example,  is always misinterpreted into .

After rhythmic pattern tagging, we can interpret *gongchepu* automatically. The interpreted staff of the *gongchepu* of 天净沙 Tian-jin-sha in Figure 1 is shown in Figure 8.



Figure 8. Interpretation of Tian-jin-sha

5 Conclusions and Future Discussions

This paper proposed an automatic interpretation of *gongchepu*. We apply Hidden Markov Model and Conditional Random Field to solve the interpretation problem. Three single features: notes sequence (NS), numbers of notes (NN), pitch interval position and direction (PIDP) and their combinations: NS+NN, NN+PIDP, NS+PIDP, NS+NN+PIDP are selected for the interpretation model.

Experimental results showed that the precision of interpretation by CRF achieved 90.05% and the oov precision was 82.03%. It will be very helpful for reading and singing the Chinese poetic songs noted in *gongchepu*. Furthermore, our work will have positive influence on the protection of the ancient Chinese traditional culture, for the number of the experts who are able to read *gongchepu* is decreasing and the way of singing Chinese traditional poetic songs will most likely fade in the following generations.

Obviously, the sample size of the *gongchepu* database (6347 beats) is much smaller than the corpus in NLP. However, music is more abstract than natural language, and music is an easier way for listener to understand and accept, while natural language may cause many unpredictable misunderstandings. Thus our work, training on the musical notation database, which is much smaller than the NLP corpus, is still credible.

Melodic features only bring a superficial knowledge in understanding the rhythm of *gongchepu*. Actually, Chinese language plays an important role in the development of Chinese music. Thus in the further research, we will take the linguistic features in consideration.

References

1. Curt Sachs: Chinese Tune-Title Lyrics. The Rise of Music in the Ancient World. London (1943)
2. Yinliu Yang: Gongchepu-qian-shuo "Introduction of gongchepu". Renmin yinyue chubanshe. Beijing (1962)
3. Rengkang Qian: Qing-jun-shi-chang-qian-chao-qu "Interpretation of Suijin cipu". Shanghai yinyue chubanshe, Shanghai(2006)
4. Xuehua Zhou: Nashu-ying-qu-pu-jian-pu-ban "Interpretation of nashu". Shanghai jiaoyu chubanshe. Shanghai (2008)
5. Julian Kupiec: Robust part-of-speech tagging using a hidden Markov model. Computer Speech and Language, 6, 225–242. (1992)
6. Daniel M.Bikel, Richard Schwartz, & Ralph M.Weischedel: An Algorithm that Learns what's in a name. Machine Learning Journal, 34, 211–231. (1999)
7. Dayne Freitag & Andrew McCallum: Information Extraction Using HMMs and Shrinkage. In Papers from the AAAI-99 Workshop on Machine Learning for Information Extration, pp. 31–36 Menlo Park, California. AAAI. (1999)
8. John Lafferty, Andrew McCallum and Fernando Pereira: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In Proceedings of the Eighteenth International Conference on Machine Learning. (2001)
9. Mingcai Hong, Kuo Zhang, Jie Tang & Zijuan Li: A Chinese Part-of-speech Tagging Approach Using Conditional Random Fields. Computer Science, Vol. 33, No. 10, pp. 148-152. (2006)
10. Ekbal Asif, Rejwanul Haque, and Sivaji Bandyopadhyay: Bengali Part of Speech Tagging using Conditional Random Field. In Proceedings of Seventh Inter-national Symposium on Natural Language Processing. Thailand (2007)
11. S. Lakshmana Pandian, T. V. Geetha: CRF Models for Tamil Part of Speech Tagging and Chunking. Proceedings of the 22nd International Conference on Computer Processing of Oriental Languages, 11-22. 42. (2009)
12. Yuanhuai Xie: Sui-jin-ci-pu "A Collection of Song". (1844)