POP-SONG STRUCTURAL ANALYSIS with LYRICS

Chang-Hung Hsieh
IISR Lab
s971506@gmail.com

Pin-Chu Wen
IISR Lab
joe791023@gmail.com

Richard Tzong-Han Tsai*
IISR Lab
National Central University
thtsai@csie.ntu.edu.tw

*corresponding author

ABSTRACT

In this paper, we propose a method for the automatic extraction of musical structure in popular music songs using conditional random fields (CRFs) to analyze song lyrics. We design song lyric features that predict the particular sections of a song (verse, chorus, bridge, etc.) from lyrics text. Using preexisting lyric timing data, we create a feature that can locate each lyric line's relative position in the song. Our system achieves high performance in terms of F1 measure (0.8308). In addition, we have released a free demo application for public use. By providing an .MP3 song file and an .LRC lyrics timing file, users can analyze and predict the structure of any Mandarin pop song.

1. INTRODUCTION

Music structure discovery (MSD) aims to characterize the temporal structure of songs. In the case of popular music, this means classifying segments of a piece of music into parts such as intro, verse, pre-chorus, chorus, bridge, collision, instrumental solo, ad lib, or outro. From the lyrics of a song, we can predict the verse (V), pre-chorus (P), chorus (C), and bridge (B), collectively referred to as VPCB. With this song structure data, one can develop new applications or functions such as the ability to navigate a song by section (e.g., skip a verse or pre-chorus), generate song excerpts, or abbreviate a song. Such functions could be useful in karaoke systems or in broadcasting for example.

2. PREVIOUS WORK

Structure in music can be defined as the organization of different musical forms or parts through time. How we define musical forms and what cements our perception of these forms is an open question. However, previous MSD studies have classified songs into finite sections [1-3].

Benward and Saker [2], for example, classify song structure into introduction, verse, pre-chorus, chorus, bridge, instrumental solo, and ad-lib. In our system, we analyze only those sections that can be predicted by lyrics:

Verse: When two or more sections of the song have basically identical music and different lyrics, each section is considered one verse.

Pre-chorus: The pre-chorus functions to connect the verse to the chorus with intermediary material, typically using subdominant or similar transitional harmonies.

Chorus: The element of the song that repeats at least once both musically and lyrically. It is almost always of greater musical and emotional intensity than the verse.

Bridge: In music, especially for popular music, a bridge is a contrasting section that prepares for the return of the original material section (the B in AABA).

Our task is to classify each paragraph of a song's lyrics as either V, P, C, or B by analyzing the song's lyrics text file, which is segmented into unclassified paragraphs, and the song's LRC file, which gives the start time of each lyric line in [mm:ss:xx] (minutes: seconds: hundredths of a second).

3. PROPOSED METHOD

In this section each individual block of the system is described.

3.1 LRC and Lyrics

We use lyrics text files from Mojim1, an online lyrics collection site. Our corpus of LRC files is compiled from KKBOX's 2006-2013 monthly top-100 Chinese pop music charts.

3.2 Pre-Processing

Before we can use the lyrics data from Mojim, we must pre-process the text files to normalize their format. After we have the formatted lyrics, we annotate each paragraph as the label composed of one of the two section positions (beginning of a section and inside a section) and one of five sections (verse, pre-chorus, chorus, bridge.. For example, the first paragraph of the verse section is annotated as B-V, meaning beginning-of-the-verse. The following paragraph in the verse section is annotated as V-I, meaning inside-the-verse.

3.3 Problem Formulation and the Model

A lyric of a song is a sequence of paragraphs. The section tags of neighboring tags are dependent. Because most songs follow similar structural patterns (VPC or VPCBC, etc.), we formulate VPCB song section prediction as a sequence labeling task and use conditional random fields

(CRFs) [4] to model this task since CRF performs very well in other text segmentation tasks such as paper abstract segmentation [5]. We use the CRF++ package.

3.4 Features

We employ the following three features computed from the lyrics texts.

Title appearance feature (TA). If the given paragraph contains the song title, its TA value is set to 1. Otherwise, if the given paragraph's next paragraph contains the title, its TA value is set to 2. Otherwise, its TA value is set to 0.

Prefix frequency rank feature (PFR). The value of this feature is set to the frequency rank of the given paragraph's two-character prefix if the rank is less than or equal to three. Otherwise, the PFR value is set to 0.

Rank of paragraph length feature (RPL). The value of this feature is set to the rank of the given paragraph's length if the rank is less than or equal to three. Otherwise, the RPL value is set to 0.

In addition, we design the following feature computed from the timing information files.

Relative starting time feature (RST). Given a paragraph p, p's RST feature is defined as follows:

$$RST(p) = \left\lfloor \frac{p's\ starting\ time}{the\ last\ paragraph's\ starting\ time} \right\rfloor$$

4. EXPERIMENTS AND CONCLUSIONS

4.1 Data set

Our evaluation dataset consists of 696 song lyrics text files from Mojim[1] and their corresponding LRC files from KKBOX. The files were annotated by fourteen annotators with pop-song background according to Benward and Saker's definitions [2].

4.2 Results and Conclusion

We evaluate our result using precision (P), recall (R) and F-measure (F):

$$\mathrm{P} = \frac{\#\text{ of sections predicted correctly}}{\#\text{ of sections predicted}}$$

$$\mathrm{R} = \frac{\#\text{ of sections predicted correctly}}{\#\text{ of sections}}$$

$$\mathrm{F} = \frac{2\mathrm{PR}}{\mathrm{P} + \mathrm{R}}$$

Table 1 shows the experimental results. We can see that in the verse, pre-chorus, and chorus the F1 measure is over 0.8. In contrast, the bridge's F1 measure is only 0.74. There are two possible reasons for this performance drop: One, bridge is the least common type of song section. Two, we observed that in the confusion matrix the entries corresponding to bridge and pre-chorus are high, so we can assume that their definitions are similar. With a larger corpus, we could likely improve the bridge's F1 measure

[1] mojim.com

	Precision	Recall	F1
V	0.8275	0.8537	0.8404
P	0.8398	0.8675	0.8534
C	0.8359	0.8399	0.8379
B	0.7812	0.7092	0.7435
ALL	0.8290	0.8327	0.8308

Table 1. Experimental results

5. REFERENCES

[1] Anderson, M. *Michael Anderson's Little Black Book of Songwriting*, 2006.

[2] Benward, B. and Saker, M. *Music in Theory and Practice* 2009.

[3] Bimbot, F., Deruty, E., Sargent, G. and Vincent, E. *Semiotic structure labeling of music pieces: Concepts, methods and annotation conventions*. City, 2012.

[4] Lafferty, J. D., McCallum, A. and Pereira, F. C. N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In *Proceedings of the Proceedings of the Eighteenth International Conference on Machine Learning* (2001). Morgan Kaufmann Publishers Inc., [insert City of Publication],[insert 2001 of Publication].

[5] Lin, R. T. K., Dai, H.-J., Bow, Y.-Y., Chiu, J. L.-T. and Tsai, R. T.-H. Using conditional random fields for result identification in biomedical abstracts. *Integr. Comput.-Aided Eng.*, 16, 4 2009), 339-352.