UNDERSTANDING THE EXPRESSIVE FUNCTIONS OF JINGJU METRICAL PATTERNS THROUGH LYRICS TEXT MINING

Shuo Zhang
Music Technology Group
Universitat Pompeu Fabra
ssz6@georgetown.edu

Rafael Caro Repetto
Music Technology Group
Universitat Pompeu Fabra
rafael.caro@upf.edu

Xavier Serra
Music Technology Group
Universitat Pompeu Fabra
xavier.serra@upf.edu

ABSTRACT

The emotional content of jingju (aka Beijing or Peking opera) arias is conveyed through pre-defined metrical patterns known as *banshi*, each of them associated with a specific expressive function. In this paper, we first report the work on a comprehensive corpus of jingju lyrics that we built, suitable for text mining and text analysis in a data-driven framework. Utilizing this corpus, we propose a novel approach to study the expressive functions of *banshi* by applying text analysis techniques on lyrics. First we apply topic modeling techniques to jingju lyrics text documents grouped at different levels according to the *banshi* they are associated with. We then experiment with several different document vector representations of lyrics in a series of document classification experiments. The topic modeling results showed that sentiment polarity (positive or negative) is better distinguished between different *shengqiang-banshi* (a more fine grained partition of *banshi*) than *banshi* alone, and we are able to achieve high accuracy scores in classifying lyrics documents into different *banshi* categories. We discuss the technical and musicological implications and possible future improvements.

1. INTRODUCTION

Traditionally, the emotional content of jingju (aka Beijing or Peking Opera) music is conveyed through pre-defined melodic and metrical patterns known as *shengqiang* and *banshi*. With the general absence of professional composers, the melodic material of jingju was taken from local tunes, and lyrics were arranged by performers according to their poetic structure. In order to convey different emotional contents, the original melodic outlines were transformed rhythmically, according to a pre-defined set of labelled metrical patterns. Each of the metrical patterns, known as *banshi*, is associated with an expressive function. Each of the melodic materials to which this metrical patterns were applied is known as *shengqiang*, and is also associated with emotional content at a larger scale.

There exists many general descriptions and rules for the expressive functions associated with each *banshi* in musicological literature [11, 15] and jingju textbooks [2, 3, 14]. However, the actual realization of these associativities across existing jingju repertoires has not been characterized in a clear manner. Such a task is well suited for a data-driven computational analysis.

In this work, we first report the work on constructing the Jingju Lyrics Collection data collection, a comprehensive corpus of jingju lyrics that we built through web scraping xikao.com, suitable for text mining and text analysis in a data-driven framework. We describe substructures of this data collection as well as relevant corpus statistics based on musicological entities and considerations. Utilizing this corpus, we propose a novel approach to study the expressive functions of *banshi* by applying text analytics techniques on lyrics.

The rest of the paper is organized as follows. Section 2 provides necessary musicological concepts and background that lead to the research questions we are concerned with, namely, understanding the emotional content of *banshi* metrical patterns through lyrics text analytics. Section 3 reports the construction of the JLC lyrics data collection and describes its substructures as well as relevant corpus statistics. Following the introduction of the JLC data collection, we then report text analytics experiments aimed at revealing different semantic content in different *banshi*, including topic modeling (Section 4) and document classification (Section 5). Finally we discuss the results and future directions.

2. BACKGROUND

As stated above, *shengqiang* (SQ) and *banshi* (BS) are the melodic and rhythmic devices used in arranging the music in jingju. They are selected in order to deliver the emotional content of lyrics, the psychological profile of the characters or the general atmosphere of the play. The two main *shengqiang* of jingju are *xipi* and *erhuang*. There are around twelve types of most common *banshi*, each interrelated with others. For example, *yuanban*, literally meaning 'original meter', is considered a default medium tempo meter, and the rest of *banshi* can be considered transformations of this one: *manban*, the result of slowing down *yuanban* in tempo and stretching it in meter; *kuaiban*, the result of speeding *yuanban* up in tempo and compressing it

Banshi	**Code**	**Postulated expressive function**	**Musical feature**
yuanban	YB	straightforward, unemotional, narration, facts and explanation	medium
manban	MB	peaceful, introspective	slow
kuaiban	KB	animated, excitement, anticipation	fast
yaoban	YAB	exterior calm and interior tension	free meter
Sheng -qiang	**Code**	**Postulated expressive function**	**Musical feature**
xipi	XP	sprightly, bright and clear, energetic, forceful, and purposeful	melodic skeleton
erhuang	EH	dark, deep and profound, heavy and meticulous	melodic skeleton

Table 1. List of common *banshi* (rhythmic) and *shengqiang* (melodic) types and their acronyms (Code) in this paper. The entity column contains 4 *banshi* types in the first 4 rows, and 2 *shengqiang* types in the last 2 rows.

in meter, etc. The combination of a *shengqiang* with a particular *banshi* results in a unique musical form (henthforce referred to as SQBS), which is referred to by combining both elements, such as *erhuang_yuanban*, *xipi_manban*, etc. In general, *shengqiang* are associated with general emotional frameworks, and *banshi* with specific expressive functions [11]. Table 1 lists the most common types of *banshi* and *shengqiang*, their musical and postulated expressive functions.

The goal of this paper is twofold: first, to introduce a jingju lyrics corpus that we constructed especially for computational text analysis in this domain; second, to understand the expressive functions associated with each *banshi* through large-scale text analytics. In the current context, we take the implicit assumption that the emotional content (i.e., the target of the expressive functions for a *banshi*) of the jingju can be represented by inspecting the semantic content expressed in lyrics. We define the following research questions: (1) What are the document-topic-word distributions that characterize the lyrics texts found in each type of *banshi*? (2) How distinct are these distributions among different *banshi*? (3) Are we able to distinguish between one *banshi* and another from lyrics? (classification) (4) How does the interplay between *shengqiang* and *banshi* affect this characterization?

In recent years, there has been a number of studies employing data-driven and computational approaches to various facets of jingju music [6–8, 12, 13]. However, all of the previous works rely on audio recording or score as their primary data source. To the best of our knowledge, the current work is novel in its use of large-scale lyrics text corpus for jingju and the application of state-of-the-art NLP and text mining algorithms to uncover the associations between jingju music and expressive functions.

3. BUILDING THE JINGJU LYRICS COLLECTION

3.1 Web Scraping Xikao Database

There are a limited number of traditional style plays in the jingju repertoire (as they are not being expanded much in modern times). In order to build a comprehensive corpus of collection of jingju lyrics, we have chosen to extract data from the well-maintained open source jingju libretto database website `xikao.com`. As this website provides jingju librettos in HTML and PDF formats not ready for corpus analysis purposes, we have crawled the website to extract all lyrics in plain text format through web scraping. All texts from this website is of Creative Commons License and is free to use for non-commercial purposes. We denote our overall collection of the lyrics data (including subsequence creation of substructures within the collection) as JLC (Jingju Lyrics Collection). We use this generic name to accommodate future possibilities of expanding the collection from other sources.

`xikao.com` is a community collaboration platform aimed at building the most comprehensive collection of jingju plays for jingju professionals and aficionados by collaboratively digitizing published jingju librettos available in prints. It is being actively maintained since its inception in 2000, and there has been a steady growth in the number of digitized librettos. At the time of writing, there are a total of 2163 published librettos in print being considered for digitization, whereas there are 850 works already digitized and proof-read/edited, and there are currently 360 plays at the various stages of being digitized by dozens of anonymous users/editors/annotators. Due to the dynamic growth of its content, we can also periodically re-apply our web scraping pipeline in order to expand our data collection to reflect the most comprehensive coverage to date.

Librettos in Xikao is organized by play as a basic unit. Metadata, *banshi* (metrical pattern), *shengqiang* (melodic skeleton), role type, as well as other information such as the instrumental interlude and oral delivery mode (spoken dialogue, singing) are also annotated in the digitized documents. Meanwhile, as noted above, the overall goal of Xikao is not oriented towards computational analysis, therefore we need to apply several transformations in order to create the most useful data sets for our study (detailed in Section 3.2 and 3.3). Several examples of these shortcomings are illustrated here. First, the organization by play may not be the most useful for analysis aimed at understanding *shengqiang* or *banshi* or other musicologically meaningful categories. Second, the meta data information are also spread within the documents, making it hard to retrieve in an straightforward way. Overall, the Xikao website in its original form (before or after we have scraped its contents and stored in plain text files) is considered unstructured data that needs to be re-structured and augmented in order to use for large scale data-drive text analysis.

3.2 Text Processing

In a post-processing stage to the web scraping, we extract all lyrics that are sung to a particular *banshi* type clearly indicated (there are a good portion of a play that are spoken dialogue). As part of the standard NLP pipeline for Chinese [1], we perform word segmentation [2] on all text documents using the state-of-the-art, Conditional Random Field-based Stanford Word Segmenter [3] [10]. The result of the segmentation is verified by hand by a native speaker of Chinese and deemed reasonable. [4] After the segmentation is obtained, we use Unicode based tokenization (splitting on whitespaces) in the study, where each token is defined as one or more Unicode character [5]. In all subsequent processing steps we remove 125 frequent single-character words using a standard stop word list of Chinese. All punctuations are removed as a normalization step for NLP pipeline. The resulting corpus contains lyrics text files for 818 plays, a quite large size to study considering the small number of jingju plays that are still being performed today.

3.3 Data Sets Permutation and Creation

Following preprocessing, we extract subsets of the data collection and restructure them in order to create musicologically meaningful datasets for computational text analysis of jingju lyrics. We consider the creation of several data sets within this framework.

SQBS Dataset: This general data set consists of lyrics from all lines in all plays that correspond to a SQBS, in a tabular format, where the first column indicates SQBS category, second contains the lyrics line. Here, it's worth pointing out a 'lyrics line' is referring to the longest unit an actor is singing continuously in the same SQBS without switching to or interrupted by any other SQBS. The total number of SQBS categories in this data set is 151. In this case, it is possible to observe the distribution of the frequencies of each SQBS category. In Figure 1 and Figure 2, we show the top 10 SQBS categories in the SQBS dataset by number of lines and by number of words/characters (where a 'line' is defined as above). With few exceptions, we can see that the top 10 most frequent categories are mostly consistent when considered by number of lines vs. by number of words/characters. Meanwhile, we note that *xipi_yaoban* (XPYAB) is the most frequent musical form in jingju by all measures, which is often used in singing in the middle of spoken dialogues.

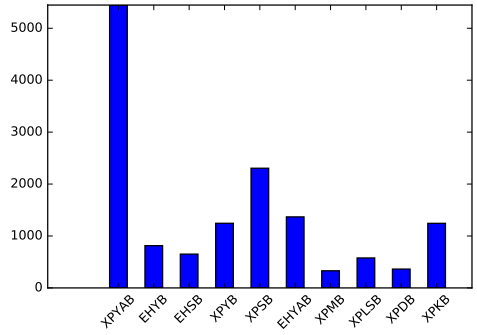

Figure 1. Distribution of top 10 SQBS categories by number of lines

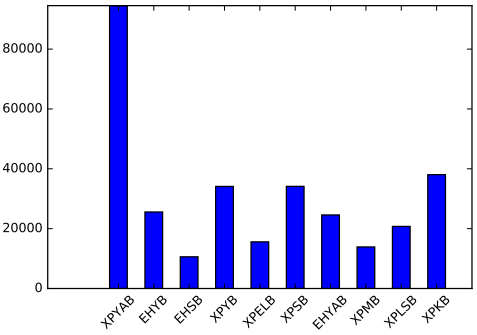

Figure 2. Distribution of SQBS top 10 categories by number of words/characters

SQBS7 Dataset: Among the 151 SQBS categories, we have selected a core set of 3 *banshi* types coupled with the two *shengqiang*, based on their musicological importance and the frequency of their occurrence in our corpus. Concretely, we consider the most basic *banshi* types for the two main *shengqiang*, that is *yuanban*, *manban*, and *kuaiban* for *xipi* and *erhuang* [6]. To also include a non metered *banshi*, we have also considered the one with a more frequent occurrence in our corpus, that is, *yaoban*, giving rise to 7 core SQBS categories that are most representative in analysis. We denote this data set as SQBS7, which is a subset of SQBS data set. All following data sets to be used in this study are transformed from the SQBS7 dataset.

PL* Datasets: The PL* data sets are grouped by play and one or two other musicological entities (BS or SQBS). First, we create the PLay-ShengQiang-BanShi (PLSQBS) data set, where a document is defined to be all the texts associated with a particular *shengqiang_banshi* within the same play. For example, all the *erhuang_ manban* texts from one play form one PLSQBS_document, whereas all the *erhuang_ yuanban* texts from the same play form another PLSQBS_document. This is aimed at looking at a particular combination of *shengqiang_banshi* type. Second, we collapse all *shengqiang* categories and create the

[1] Here we apply a shallow pipeline of word segmentation, tokenization, and stop-word removal.

[2] Since the Chinese language is written without spaces between characters and words, the word segmentation is a necessary and challenging task for any NLP or text mining analysis of Chinese text.

[3] Obtained at http://nlp.stanford.edu/software/segmenter.shtml.

[4] The linguistic style of the jingju lyrics is a mixed style that is typically similar to modern Chinese yet with occasional semi-classical style language. Therefore, unless there is a segmenter trained specifically on this language, using any pre-trained segmenter would not have yielded a perfect segmentation. To give an estimate of error, the original Stanford Segmenter paper [10] gives a overall F-score around 0.95, with a recall of known vocabulary greater than 0.95 and a recall of unknown vocabulary (OOV) in the range of 0.7s.

[5] In Chinese, a 'word' can be any number of characters, most commonly 1,2,3, or 4.

[6] It has to be noticed that in traditional plays *erhuang* was never set to *kuaiban*.

PLay-BanShi (PLBS) documents. Each document in this data set is defined to be all the texts associated with a particular *banshi* within the same play. Therefore, regardless of *shengqiang*, all the *manban* texts from one play form one PLBS document.

One potential problem with the PL* data sets is that the partition of documents may result in very short documents, creating a data sparseness problem in document modeling algorithms. Figure 4 shows the distribution of document length in the PLSQBS data set. We observe that there is a peak at less than 200 words per document, whereas there are also a few documents with more than 2000 or 4000 words. This may be taken into consideration when performing text mining experiments on these data sets.

AG* Datasets: We aggregate all PLSQBS documents to form 7 AGSQBS ('AG' for aggregate) documents (as there are 7 types of *shengqiang_banshi* considered in the current study), in order to study the characteristics in all texts associated with a particular *shengqiangbanshi*. By analogy, we aggregate all PLBS documents to form the 4 AGBS documents (i.e., all texts for a particular *banshi*).

We provide an overview of the relationships between data sets in Figure 3. Table 2 gives a more detailed description of the PL* and AG* data sets used in the subsequent experiments. The entire data collection is openly available through Github [7], and the datasets used in the experiments of this paper are available through CompMusic project website [8].

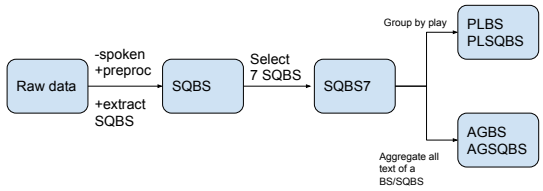

Figure 3. Block diagram overview of data sets created

Name	Description of document	#documents
PLSQBS	all the texts associated with a particular *shengqiang-banshi* within a particular play	1429
PLBS	all the texts associated with a particular *banshi* within the a particular play	1247
AGSQBS	all the texts associated with a particular *shengqiang-banshi*	7
AGBS	all the texts associated with a particular *banshi*	4

Table 2. Overview of data sets used in the experiments of this paper

[7] https://github.com/MTG/Jingju-Lyrics-Collection
[8] http://compmusic.upf.edu/jingju-lyrics-datasets

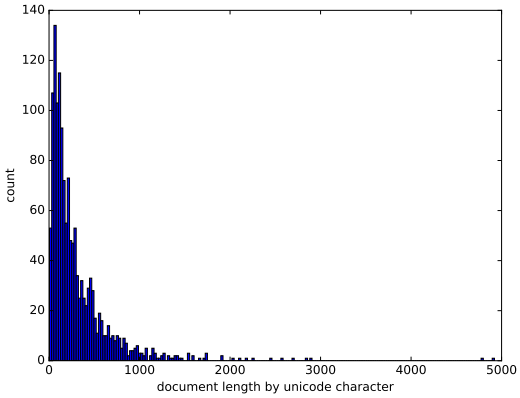

Figure 4. Distribution of document length in the PLSQBS dataset

4. EXPLORING TOPIC STRUCTURES OF SUBSECTIONS OF JLC DATA COLLECTION

In this section, we use probabilistic topic models to explore the topic structures of lyrics in different *banshi* in an unsupervised setting [9].

4.1 Topic Modeling

Topic model is a class of unsupervised statistical models for uncovering the underlying semantic structure of a document collection. The idea is to model each document as arising from multiple background (latent) topics, where a topic is defined to be a distribution over a fixed vocabulary of terms. Specifically, we assume that K topics are associated with a collection, and that each document exhibits these topics with different proportions. In a topic model, we typically obtain a document-topic distribution (the probabilistic distribution of all topics in a particular document), and a topic-word distribution (the distribution of the terms that are associated with one topic). In the current context, we are interested in the topics that characterize each *banshi*(BS) or *shengqiang_banshi*(SQBS) in the AGBS and AGSQBS data sets.

Here we consider a state-of-the-art topic modeling techniques known as Latent Dirichlet Allocation(LDA) [1]. In LDA, each topic z is associated with a multinomial distribution over the vocabulary Φ_z, which is drawn from a Dirichlet prior $Dir(\beta)$. A given document D_i is then generated by the following process:(1) Choose Θ_i ˜$Dir(\alpha)$, a topic distribution for D_i; (2) For each word $w_j \in D_i$: (a) Select a topic z_j ˜Θ_i (b) Select the word w_j ˜Φ_{zj}. We use collapsed Gibbs sampling implementation in Mallet [10] to infer the values of the latent variables Φ and Θ.

We compute the perplexity measure for a held-out data set defined in the LDA model to determine the optimal number of background latent topics in the current experiments. The perplexity, used by convention in language

[9] The code for all experiments in this paper is available at https://github.com/MTG/Jingju-Lyrics-Text-Analysis.
[10] Downloaded from http://mallet.cs.umass.edu/

YB	0:Hanxin (military General), youth, wealth, ruthless
KB	2:family, mother, husband-wife, brother; 16: war, military, mails, destruction
MB	4:princess, death, crime, Chang'an, brave; 19: wish, sir, madam, pain, defeat
YAB	7: sudden news, human head, revenge, affection

Table 3. Top topics for each *banshi* and their top words. Sorted by topic index number (0 is topic 0 assigned by the LDA model, etc.)

modeling, is monotonically decreasing in the likelihood of the test data, and is algebraically equivalent to the inverse of the geometric mean per-word likelihood [1]. A lower perplexity score indicates better generalization performance. More formally, for a test set of M documents, the perplexity is:

$$perplexity = exp\left\{-\frac{\sum_{d=1}^{M}\log p(w_d)}{\sum_{d=1}^{M} N_d}\right\} \qquad (1)$$

where w_d is a word in the document, and N_d is the length of the document (total number of words). In our experiments, we compute perplexity for each topic model given number of topics from 8 to 50. The result shows that using 20 topics results in the lowest perplexity and is the optimal choice.

4.2 Topic Modeling Results

First, we present the topic modeling results for the AGBS and AGSQBS data sets. In this case, the output from the MALLET LDA model contains a document-topic distribution and a topic-word distribution. The former shows the distribution of topics (number of topic K=20, as determined in Section 4.1) in each BS/SQBS (i.e., each document in the data set), and the latter shows the top 20 words associated with each topic. Since we are interested in characterizing the main topics found in each BS/SQBS, we show results for both of these components.

To better understand the topics that characterize each category, we extracted the most salient topics from the topic distribution of each BS or SQBS. In doing so, we removed common topics with high occurrences in all categories, and only select those with a high occurrences in each category. We present these topics and summarize their top words in English in Table 3 and Table 4. Many of these topics have to do with specific stories in Chinese history that are well known in jingju repertoires. Comparing these results to the general descriptions found in Table 1, we see a reasonable interpretation for each category - although we observe that the division between the positive and negative emotions of *xipi banshi* and *erhuang banshi* in Table 4 are much more salient than the topics that distinguish the four *banshi* types in Table 3.

EHYB	0: pity, old, heavy, prince, depart, pain, cold
EHMB	2: unfortunate, pity, worry; 14: war, military, courtesy, hero
EHYAB	4: tears, sir, madam, wish, leave, pain, hurt, stab; 18: life, run, death, brave, sword
XPYB	7: Kings from The Three Kingdoms, drink, happy
XPKB	10: traitor, laugh, believe
XPMB	12: youth, beautiful view, morning, world
XPYAB	11: general(military), prime minister, angry, military, step forward; 14: (see EHMB)

Table 4. Top topics for each *shengqiang-banshi* (SQBS) and their top words. Sorted by topic index number(0 is topic 0 assigned by the LDA model, etc.)

5. DOCUMENT CLASSIFICATION IN JLC DATA COLLECTION

In this section we propose a supervised document classification task with the aim of classifying lyrics documents in the PL* data sets into different *banshi* categories.

5.1 Document Vector Representation

In the vector-space model (VSM) of information retrieval, a text document is represented by a document-term vector where each attribute represents the frequency (count) with which a particular term $w_{k,i}$ (a word in the vocabulary) occurs in the document d_i (aka bag-of-words or BOW). A highly effective transformation of BOW word vector weights is tf-idf weighted vectorization. [11] However, a shortcoming of these types of traditional word vectors is that it is high-dimensional and very sparse.

Recent advances in NLP have concentrated on training word embeddings with neural networks that result in low-dimensional dense vector representations of words [5] by predicting target words from context (or vice versa). These high quality word embeddings also have the desired property of reflecting semantic similarity in vector space (semantic similarities can be captured by vector arithmetics). Expanding on this idea of word embeddings, [4] trained embeddings for sentences or longer units, which they denote "paragraph vectors". These paragraph vectors have the similar properties of reflecting semantic similarity above the word level, and is shown to be highly effective in a series of document and sentiment classification tasks.

5.2 Document Classification in JLC

To characterize the strength of the association between the lyrics text and its associated *banshi*, we define a document classification task: to what degree can we use textual features extracted from the lyrics to classify the documents in the PLBS and PLSQBS data sets into one of the four BS or seven SQBS classes?

[11] In the tf-idf (term frequency - inverse document frequency) weighting scheme [9], the term frequency count is compared to an inverse document frequency count, which measures the number of occurrences of a word in the entire corpus. Thus the tf-idf transforms the document into a weighted vector that assigns higher value to terms that have high occurrences in a small number of documents.

We use several varieties of vector representation for documents in our classification experiment, as described above: (1) BOW; (2) tf-idf BOW; (3) Paragraph Vectors (D2V); (4) document-topic distribution from the topic models (TM) we derived in Section 4. The document-topic distribution can be seen as a very low-dimension representation of a document, which has been shown to perform well in document classification tasks in place of BOW representation [1]. For Paragraph Vectors, we train document embeddings on our PL* data sets using the `Doc2Vec` (D2V) implementation available in the Python library `gensim`. The resulting embeddings has 100 dimensions for each document. We use Support Vector Machine (SVM) with RBF kernel for all classification experiments.

5.3 Document Classification Results

For the PLSQBS and PLBS data sets, we present the document classification accuracy [12] in Table 5. We observe that document classification accuracy scores are significantly above chance in both data sets (chance being $1/7 == 0.14$ in PLSQBS data set and $1/4 == 0.25$ in PLBS data set), with best accuracy scores of 0.41 and 0.53. This indicates that supervised learning is able to capture the important features that distinguish the different *banshi* or *shengqiang-banshi* classes, which in these particular JLC data sets, are somewhat unintuitive even for human judgments [13].

Contrasting the performance of different features, we note that tf-idf is more effective than BOW, as expected. The D2V Paragraph Vectors achieves comparable or lower results with the tf-idf (200 times higher in dimension). This is somewhat unexpected since these Paragraph Vectors are supposed to be a higher quality vector representation that captures both semantic similarity and word order that is absent from BOW representations such as tf-idf [4]. We attribute this under-performance of D2V to the smaller amount of training data available in the PL* data sets (comparing to much larger general-domain training corpora used in literature for D2V). In the mean time, we observe that the topic models features are ineffective at capturing the document-level distinctions.

Dataset	Feature	Accuracy
PLBS	BOW (20000)	0.481
PLBS	tf-idf (20000)	0.527
PLBS	**D2V** (100)	**0.528**
PLBS	TM (20)	0.274
PLSQBS	BOW (20000)	0.356
PLSQBS	**tf-idf** (20000)	**0.408**
PLSQBS	D2V (100)	0.347
PLSQBS	TM (20)	0.112

Table 5. Document classification with SVM RBF kernel, with dimensionality of features shown in parenthesis

6. CONCLUSION

In this work, we have introduced the Jingju Lyrics Collection, a comprehensive data collection of jingju lyrics enriched and re-structured with several extracted datasets based on musicological considerations. Utilizing this data, we performed topic modeling in order to explore the topic structures of the jingju lyrics as related to different *banshi* and SQBS types. The results show that while the topics are in general reasonable, the distinctions between different *banshi* are less contrastive than between different *shengqiang-banshi* [14]. Document classification experiments are carried out to further understand the association within a supervised setting. The strong results in document classification support the associations between the expressive functions (as expressed in the lyrics) and the *banshi* or *shengqiang-banshi* (SQBS) categories.

We observe that unexpectedly, neither D2V Paragraph Vectors nor the topic models are more effective at document classification than the high-dimension tf-idf vectors. We postulate that these have several implications (even though our goal and contribution in this work do not lie in the use of these more advanced representations). First, it shows that latent topics may not be the most effective way to capture the different expressive functions in different *banshi* types (as opposed to, e.g., sentiment). It maybe of interest to perform feature and error analysis to understand what components of the document classification have made it more effective (e.g., sentiment polarity words, etc). Second, in addition to the training size problem discussed in Section 5.3, we attribute lower performance of D2V/TM to the potential errors in the NLP pipeline applied to the corpus, especially the Chinese segmentation (as already mentioned in Section 3.2). This includes two aspects: first, the automatic segmentation may introduce errors even for standard Chinese text; second, the language of jingju falls somewhere between modern and archaic Chinese, making it more challenging to segment automatically using a standard segmenter trained on modern language. Our on-going and future work, therefore, includes making corrections to the segmentation in the JLC while keeping expanding the collection.

7. ACKNOWLEDGEMENTS

This research is funded by the European Research Council under the European Union's Seventh Framework Program (FP7/2007- 2013), as part of the CompMusic project (ERC grant agreement 267583).

[12] The datasets are well balanced in their class sizes therefore accuracy is an appropriate measure.

[13] Here we are referring to a layperson who is a native speaker of Chinese but may not be a jingju expert. We are yet to evaluate this task on jingju experts.

[14] Our main goal in this paper is to investigate the expressive functions of *banshi*. Even though the result shows *shengqiang*'s importance in distinguishing positive from negative sentiments, we note that *shengqiang-banshi* is still a unique form combining both *shengqiang* and *banshi*.

8. REFERENCES

[1] David M Blei, Andrew Y Ng, and Michael I Jordan. Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(4-5):993–1022, 2012.

[2] Cao B.. *Jingju changqiang banshi jiedu (Deciphering banshi in jingju singing)*. Renmin yinyue chubanshe, Beijing, 2010.

[3] Liu J.. *Jingju yinyue gailun(Introduction to jingju music)*. Renmin yinyue chubanshe, Beijing, 1998.

[4] Quoc V. Le and Tomas Mikolov. Distributed representations of sentences and documents. In *Proceedings of the 31th International Conference on Machine Learning, ICML 2014, Beijing, China, 21-26 June 2014*, pages 1188–1196, 2014.

[5] Tomas Mikolov, Ilya Sutskever, Kai Chen, Gregory S. Corrado, and Jeffrey Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems 26: 27th Annual Conference on Neural Information Processing Systems 2013. Proceedings of a meeting held December 5-8, 2013, Lake Tahoe, Nevada, United States.*, pages 3111–3119, 2013.

[6] Rafael Caro Repetto, Rong Gong, Nadine Kroher, and Xavier Serra. Comparison of the singing style of two jingju schools. In *16th International Society for Music Information Retrieval (ISMIR) Conference*, pages 507–513, Málaga, Spain, 26/10/2015 2015.

[7] Rafael Caro Repetto and Xavier Serra. Creating a corpus of jingju (beijing opera) music and possibilities for melodic analysis. In *15th International Society for Music Information Retrieval Conference*, pages 313–318, Taipei, Taiwan, 27/10/2014 2014.

[8] Rafael Caro Repetto and Xavier Serra. Melodic transformation processes in the arrangements of jingju banshi. In *Fourth International Conference On Analytical Approaches To World Music (AAWM 2016)*, New York, USA, 08/06/2016 2016.

[9] Gerard Salton and Michael J. McGill. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, NY, USA, 1986.

[10] Huihsin Tseng. A conditional random field word segmenter. In *In Fourth SIGHAN Workshop on Chinese Language Processing*, 2005.

[11] Elizabeth Wichmann. *Listening to theatre: the aural dimension of Beijing opera*. University of Hawaii Press, 1991.

[12] Shuo Zhang, Rafael Caro Repetto, and Xavier Serra. Study of the similarity between linguistic tones and melodic pitch contours in beijing opera singing. In *15th International Society for Music Information Retrieval Conference*, pages 343–348, Taipei, Taiwan, 27/10/2014 2014.

[13] Shuo Zhang, Rafael Caro Repetto, and Xavier Serra. Predicting pairwise pitch contour relations based on linguistic tone information in beijing opera singing. In *16th International Society for Music Information Retrieval (ISMIR) Conference*, pages 107–113, Malaga, Spain, 26/10/2015 2015.

[14] Zhang Z. . *Jingju chuantongxi pihuang changqiang jiegou fenxi (Structural analysis of pihuang singing in jingju traditional plays)*. Renmin yinyue chubanshe, Beijing, 1981.

[15] Jiang J. . *Zhongguo xiqu yinyue (Music of Chinese traditional opera)*. Renmin yinyue chubanshe, Beijing, 2000.