Classification Research on Chinese Folk Songs

Lu Dong

Department of Computer Science & Technology, Xi'an Jiaotong University

Emails: donglu666@stu.xjtu.edu.cn

ABSTRACT

In the era when music management has become an automatic process with the help of machine learning, research about Chinese folk songs attracts so much attention. The global feature model or local probabilistic model showcases its weakness in deciphering folk song with efficiency and correctness. This paper puts forward a synthetic feature model including both global and local perspectives to improve classification performance. Through a carefully selected global feature sets and an improved Bag of N-Gram scheme, music data can be better carved by data. The experiments verify that the synthetic feature model achieves the best performance with an accuracy of 90.6%.

1. INTRODUCTION

Classification is one of the major research in Music Information Retrieve[1-3] (MIR). Traditional research of MIR tends to use the music global feature model[4] one method combining music theory and mathematical statistics knowledge to design features that reflect the character of the whole song. Specifically, it maps a tune into a feature vector, through which people would be able to learn the pattern, classify among different genres or evaluate the emotion expressions by machine learning algorithms.

Currently, several famous global feature sets based on MIDI, like Alicante[5] sets, Fantastic[6] sets and McKay[7] sets not appropriate folk songs. One reason is that folk songs are without instruments. So we need to figure out a feasible global feature sets for Chinese folk songs. Another reason is that these models barely take the local features into considerations. For music, local characters[8, 9] like the combination of pitch and rhythm are critically important, especially for such high similarity structure songs. In the following part, this paper will introduce a synthetic model to comprehensive map music into data, together with some experiments to evaluate the design.

2. SYNTHETIC FEATURE MODEL

The main idea of the synthetic feature model is to build a feature model combining both global features and local features of folk songs. So there are two parts of this model, one is global features strategies and the other one is local feature strategies. The former one is based on previous famous global feature sets to select the global features that satisfy our music. The latter one is to build our feature bags from four levels-pitch, interval, duration and duration rate, and meanwhile using an improved Expected Cross-Entropy (ECE) to reduce the dimension. The structure chat presents in Figure 1.

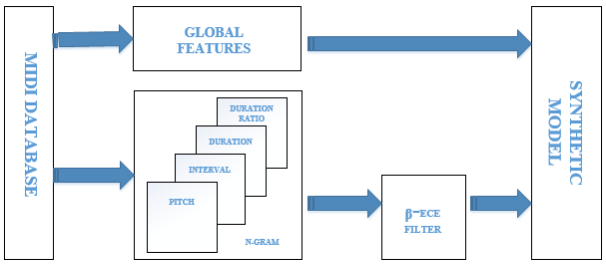

Figure 1. Synthetic Feature Model Structure.

3. GLOBAL FEATURE STRATEGIES

Since Chinese folk songs[9] are improvisational and survived through the mouth to mouth, not every feature mentioned before suits this kind of music. Thus, based on previous work, this paper designed abundant adaptable ones. The features such as musical interval with direction or without reflecting edges surpassed previous designs. In summary, these features can be divided into six categories: pitch class, duration class, neighbor class, interval class, rest class, and others. These items matched up with the feature name (FX) list in Table 1.

CLASS	FEATURES
Pitch	The maximum pitch(F1); The minimum pitch(F2); The most frequent pitch(F3); The most frequency pitch class(F4); Pitch range(F5); The average pitches(F6); Pitch standard deviation(F7); Pitch distribution(F8); Pitch entropy(F9).
Duration	The maximum duration(F10); The minimum duration(F11); The most frequent duration(F12); Duration range(F13); The average durations(F14); Duration standard deviation(F15); Duration distribution(F16); Duration entropy(F18-20)
Neighbor	Adjacent notes' start time range(F21); The average adjacent notes' start time (F22); Adjacent notes' start time standard deviation (F23); Adjacent notes' start time distribution(F24).
Interval	Interval range(F25); The average interval (F26);Interval standard deviation(F27); Interval distribution(F28); The most frequent interval with direction(F29); The most frequent interval without direction(F30); Interval entropy(F31); Interval Statistics(F32-39).
Rest	The maximum rest(F40); Rest range(F41); The average rest(F42); Rest standard deviation(F43); Rest distribution(F44).
Others	Notes density(F45); Melodic contour(F46).

Table 1. Global Feature Design

To judge these global features, the author applied Information Gain (IG) measurement. From the following histogram, the most IG ones could be distinguished for classification. It can not only help the individual who is also into this method but also can be used to crop the dimensions of the feature sets. But in this paper, due to the total number is not so large, the author still took all these features into considerations.

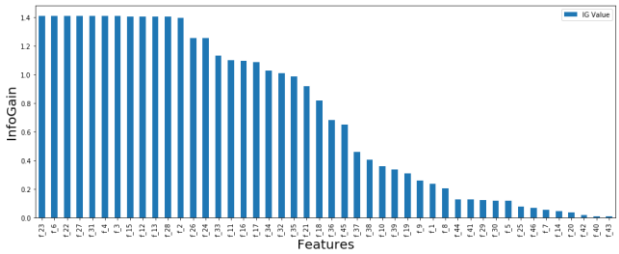

Figure 2. Global Feature Evaluation

4. LOCAL FEATURE STRATEGIES

The idea of "Bag of Words[10]" treats texts as combinations of words without order, because it is the words itself that matters rather than their positions. Meanwhile, the N-Gram algorithm[11] --the most popular statistical language model--is widely used in automatic speech recognition. It uses the contextual information to calculate a maximum probability to recognize sentences. These ideas pretty correspond to intuition towards music as some melodies always make people feel happy while some others drive people sad. Combining these two ideas, the author built a Bag of N-Gram model to decipher music local characteristics.

SVM	Pitch	Duration
1-GRAM Accuracy	0.763	0.743
2-GRAM Accuracy	0.779	0.748
3-GRAM Accuracy	0.724	0.731

Table 2. Comparing different N-Gram.

The main idea of Bag of N-Gram is to extract the symbolic sequences from MIDI files and use an N-length sliding window to obtain substrings as music words. A MIDI file saves a series of music events, each of which contains a pitch and its duration and consecutive notes form an interval. Some researchers have already used N-Gram of interval to achieve some success which triggers us to spare some attention to that. Thus, it is possible to apply an N-length window on four levels- pitch, duration, interval and duration ratio. As for the choice of N, many papers have demonstrated that 2 or 3 tend to be ideal because larger ones will degrade the efficiency and performance of classification. The author used an independent experiment to verify the performance among 1-Gram, 2-Gram, and 3-Gram and the result shows in Table 2. No doubt, the result indicates 2-Gram performs better than 1-Gram and 3-Gram, so this paper we choose N as 2

Since these combinations are still too large and sparse, the author chose the Expected Cross-Entropy (ECE) to reduce the dimension to denoise our data, because ECE not only considers the frequency of the words but also analyze the relationship between words. However, the climax part of music has a stronger power to affect our emotions, so it should not be given the same valuation. For this, this paper introduces an important factor β to lift the weights of climax parts. The equation is as followed.

$$\beta = \prod_{i=1}^{n} \frac{X_i}{\mu_c}$$

Xi is the unit of the sequences, where it can be pitch or duration, μ_c is the average value under class C, and the improved equation is as followed.

$$\beta\text{-}ECE(\mathrm{f}) = P(\mathrm{f}) \sum_{i=1}^{n} \beta_i p(\mathrm{C}_i \mid \mathrm{f}) \log \frac{P(\mathrm{C}_i \mid \mathrm{f})}{P(\mathrm{C}_i)}$$

In this way, the importance and frequency are both taken into consideration, which makes our bags more reasonable. We finally apply this method into our four levels-pitch, duration, interval, and duration ratio to build a less noise local feature model. We will show the changes before and after the dimension-reduced in Table 3.

2-GRAM	Pitch	Duration	Interval	Duration Ratio
Initial dimensions	281	138	403	240
ECE dimensions	163	83	237	133
β -ECE dimensions	178	85	256	143

Table 3. Dimension Changes

From the results, it can be found that after using ECE to reduce the dimension, the numbers almost half the initial dimensions, which means there are lots of noises. While the β -ECE improves a little bit dimensions which means better protect the climax parts.

5. EXPERIMENTS

5.1 Results of N-Apriori

All experimental data sets were subtracted from a Chinese folk song MIDI database, which was constructed in accordance with the scores of a national authority book named "The Collections of Chinese Folk Songs". The method of 10-fold cross-validation has been used in our data. We choose three popular classifiers, that is, Decision Tree, SVM, and Bayesian Network. Two measurements Accuracy and F-measure have been used to evaluate the results. In the experiments, the author compared four different models, the lobal model, N-GRAM probabilistic model, local Bag of N-GRAM model and Synthetic feature model.

The result in Table 4 shows that all the local feature models perform better than the global model, which means the local characteristics are more important than global characteristics in folk song classifications. Besides, the im-

proved local feature scheme performs better than the N-Gram Probabilistic model, which means only computing the probabilistic of the sequences cannot fully reflect the songs, and a varied of weights are of great importance. At last, the synthetic model achieves the best performance since it considers both global and local features that can summarize the music at a comprehensive view.

6. CONCLUSIONS

In this paper, the author reviewed the traditional resolution of music classification and analyzed its weakness for folk songs classification. One synthetic model is designed with both global and perspectives to improve classification performance. Through a fine selected global features and an improved Bag of N-Gram scheme, music characteristics can be better carved by data. The comparing experiments verify our ideas that the synthetic model could achieve the best performance with an accuracy of 90.6% using the SVM classifier on Chinese folk songs.

	Global model			N-Gram	Bag of N-Gram			Synthetic model		
Classifier	C4.5	SVM	B.N	----	C4.5	SVM	B.N	C4.5	SVM	B.N
Accuracy	0.685	0.829	0.523	0.843	0.786	0.864	0.846	0.813	**0.906**	0.887
F-measure	0.687	0.827	0.497	0.819	0.785	0.863	0.845	0.813	0.900	0.887

Table 4. Classification Results of Different Models

7. REFERENCES

[1] Mckay C, Fujinaga I. Automatic Genre Classification Using Large High-Level Musical Feature Sets.[C]. 2004.

[2] Cataltepe Z, Yaslan Y, Sonmez A. Music Genre Classification Using MIDI and Audio Features[J]. Eurasip Journal on Advances in Signal Processing. 2007, 2007(1): 150.

[3] Mandel M I, Dan E. Song-Level Features and Support Vector Machines for Music Classification.[C]. 2005.

[4] Shang C, Li M, Feng S, et al. Feature selection via maximizing global information gain for text classification[J]. Knowledge-Based Systems. 2013, 54(4): 298-309.

[5] León P J P D, I Esta J M. Statistical description models for melody analysis and characterization[C]. 2004.

[6] Cuthbert M S, Ariza C, Friedland L. Feature Extraction and Machine Learning on Symbolic Music using the music21 Toolkit.[J]. Ismir. 2011.

[7] Mckay C, Fujinaga I. Automatic Genre Classification Using Large High-Level Musical Feature Sets.[C]. 2004.

[8] Velarde G, Weyde T, Meredith D. An approach to melodic segmentation and classification based on filtering with the Haar-wavelet[J]. Journal of New Music Research. 2013, 42(4): 325-345.

[9] Pamjav H, Juhász Z, Zalán A, et al. A comparative phylogenetic study of genetics and folk music.[J]. Molecular Genetics & Genomics Mgg. 2012, 287(4): 337-349.

[10] Zhang Y, Jin R, Zhou Z. Understanding bag-of-words model: a statistical framework[J]. International Journal of Machine Learning and Cybernetics. 2010, 1(1-4): 43-52.

[11] Stolcke A. SRILM-an extensible language modeling toolkit[C]. 2002.