Music Classification using Features based on Musical Knowledge

Takashi Hasegawa **Takuya Nishimoto** **Nobutaka Ono** **Shigeki Sagayama**

Graduate School of Information Science and Technology, The University of Tokyo

{hasegawa, nishimoto, onono, sagayama}@hil.t.u-tokyo.ac.jp

ABSTRACT

This extended abstract details submissions to the Music Information Retrieval Evaluation eXchange. Three or four feature sets based on musical knowledge[1] were calculated. Then an SVM with the ANOVA RBF kernel was used, or a canonical LDA was conducted with effective features chosen from the sets. AIC values or accuracy of cross validation of training data were calculated for the feature selection. Four submissions were made by combinations of feature sets and classifiers.

1. FEATURE SET

1.1 Sonority (Sound Expanse and Thickness) features

Sound expanse and thickness are one set of important features when listening to music. We defined expanse as deviation and thickness as entropy of pitches.

Sound Expanse: $$D_n = \sqrt{V_n} = \sqrt{\sum n_i^2 \frac{p_i}{P} - E_n^2} \quad (1)$$

Sound Thickness: $$H = -\frac{1}{M} \sum_i^{N_F/2} \frac{p_i}{P} \log \frac{p_i}{P} \quad (2)$$

where $n_i = \frac{12}{\log 2} \log \frac{f_i}{f_A} + n_A$ is the pitch number, p_i is the power of the pitch n_i, P is the sum of the powers, $E_n = \sum n_i \frac{p_i}{P}$ is the pitch expectation.

The mean, standard deviation, skewness and kurtosis were computed across the audio clip with weight of P.
The total number of features in this set is 8.

1.2 Timbre features

Timbre is also important for music, and we used Mel-Frequency Cepstral Coefficients (MFCC) as timber features. The mean and standard deviation over the time frames were computed using Marsyas[2]. Then the mean and standard deviation across the audio clip were taken again.
13 coefficients and 1 second time frame were used.
The total number of features in this set is 52.

1.3 Harmony features

Harmony is one of traditional musicological aspects of Western music. We defined harmony as the occupancy rate of chords, and obtained virtual chord information from chroma vectors. Since we did not aim at accurate chord recognition, chord information was derived from prominent powers in the chroma vector. Virtual chord information $c(t) = \{c_i(t)\}$ is defined as 12 bit code and each bit $c_i(t)$ represents whether corresponding chroma power is prominent or not.

$$c_i(t) = \begin{cases} 1 & (p_i(t) \geq p_T) \\ 0 & (p_i(t) < p_T) \end{cases}$$
$$p_T = \mu_p + k_T \sigma_p \quad P\{z \sim N(0,\ 1^2) \leq k_T\} = \frac{2}{3} \quad (3)$$

Then the occupancy rate of each code across the audio clip was computed with weight of log power.

1.4 Rhythm features

Rhythm is also one of fundamental musicological aspects. We defined rhythm as power modulation of each pitch because it is considered to be arisen when there is repeated fluctuation in power or pitch change.
The powers of frequencies around the pitch are added using triangular window.
Then Fourier transformation of each pitch power was performed.

2. CLASSIFICATION

2.1 Support Vector Machine (SVM)

An SVM with the ANOVA radial basis kernel (ANOVA RBF) was used because it achieved good results on preliminary trials.

2.2 Canonical Linear Dicriminarion Analysis (LDA)

The features were excerpted in order to prevent overfitting before conducting the LDA. The algorithm for selecting the features was as follows:

(1) calculate an effectiveness of each feature

(2) add the most effective feature

(3) repeat (1) and (2) until a terminal condition is conformed.

We adopted the following two combinations of the effectiveness and the terminal condition.

2.2.1 Wilks Λ and AIC

(a) Effectiveness: Wilks Λ

$$\Lambda(x) = \frac{\left| \sum_{k=1}^{g} \sum_{i=1}^{n_k} \left(x_i^{(k)} - \bar{x}^{(k)}\right)\left(x_i^{(k)} - \bar{x}^{(k)}\right)' \right|}{\left| \sum_{k=1}^{g} \sum_{i=1}^{n_k} \left(x_i^{(k)} - \bar{x}\right)\left(x_i^{(k)} - \bar{x}\right)' \right|} \quad (4)$$

This value indicates degree of class separation. The smaller it is, the better classes are expected to be separated.

(b) Terminal Condition: If the following AIC does not decrease after addition of any feature, the selection is terminated.

$$AIC = -2\log L\left(\hat{\theta}\right) + 2m = n\log\Lambda_w + 2p(g-1) \quad (5)$$

where n is the number of the audio clips, p is the number of features and g is the number of categories (composers, artists, genres, or moods).

2.2.2 LOOCV Accuracy

(a) Effectiveness: Accuracy of Leave-One-Out Cross-Validation of training data.
The greater it is, the greater accuracy for test data is expected.

(b) Terminal Condition: If the accuracy does not increase after addition of any feature, the selection is terminated.

3. SUBMISSIONS

Four submissions were made by combinations of feature sets and classifiers, (a)(b)(c)(d). One feature sets consists of Sonority, Timbre and Harmony, and the other consists of three feature sets and Rhythm. The specifications of each submission and corresponding shell script name are shown in Table 1.

Classifier \ Feature Sets		Sonority+Timbre +Harmony	Sonority+Timbre +Harmony+Rhythm
SVM		(a) “tcck”	(b) “tcpk”
LDA	AIC	(c) “tcca”	“tcpa”
	LOOCV	(d) “tccl”	“tcpl”

Table 1. Submission Specifications

4. IMPLEMENTATION

4.1 Feature Extractor

Each feature set is computed respectively.

(1) Sonority features extractor consists of one C++ program.

(2) Timbre features extractor consists of Marsyas and one C++ program.

(3) Harmony features extractor consists of two C++ programs.

(4) Rhythm features extractor consists of one C++ program and one R program.

4.2 Classifier

Each submission consists of one R program.

The function “ksvm” in the package “kernlab” was used as the SVM classifier, and the function “lda” in the package “MASS” was used as the LDA classifier.

The “kernlab” package needs to be installed in R before using it, while the “MASS” package does not because it is originally included in R.

5. REFERENCES

[1] Takashi Hasegawa, Takuya Nishimoto, Nobutaka Ono and Shigeki Sagayama: “Composer Identification from MIDI Data by Combination Features of Pitch and Duration based on Musical Knowledge,” *SIGMUS79, ISPJ*, pp. , 2009 (in Japanese).

[2] Graham Percival and George Tzanetakis: Marsyas User Manual, http://marsyas.sness.net/docs/manual/marsyas-user.pdf.

[3] Alexandros Karatzoglou, Alexandros Smola, Kurt Hornik and Achim Zeileis: “kernlab - An S4 Package for Kernel Methods in R,” *Journal of Statistical Software*, Vol. 11, Issue 9, Nov 2004.