Musical Instrument Identification Based on New Boosting Algorithm

○Jun Wu[1], Daisuke Saito[1], Hirokazu Kameoka[1,2] and Shigeki Sagayama[1]
[1] The University of Tokyo, [2] NTT CS Labs, NTT Corporation

1 Introduction

This paper describes an approach to multipitch estimation and musical instrument identification from multi-instrument polyphonic music signals.

In the past, instrument identification from isolated single notes was widely investigated, while only few studies have been done in the polyphonic case as it involves decomposition of polyphonic sounds into individual notes before identifying the instruments. Harmonic-Temporal-Timbral Clustering (HTTC) was proposed by Miyamoto *et al*. for solving this problem without any prior knowledge about instruments [4]. In this paper, we discuss utilization of prior knowledge about instruments in combination with Harmonic-Attack model [1].

2 Attack-Harmonic model of single notes in polyphonic music

2.1 Model-based decomposition into notes

To identify musical instrument for each note in polyphonic music signals, we need to decompose the signal into individual notes. The model of single note sound should contain rich information for identifying the instrument, such as Attack-Harmonic model which we proposed for multipitch analysis [1]. It models both harmonic and attack parts of individual notes using Gaussian mixtures as shown in Figure 1.

The spectrogram of input polyphonic music signal is modeled by a mixture of the above models whose parameters can be estimated by the EM algorithm.

2.2 Feature extraction

The estimated parameters through the Attack-Harmonic model contain rich information for identifying the instrument including note energy, relative partial energy, partial bandwidth, harmonic temporal envelope and attack spectral envelope. Principal Component Analysis (PCA) can be applied to reduce the high dimension of these features in the instrument identifying stage.

In polyphonic music, however, notes often overlap that cause imperfect decomposition into

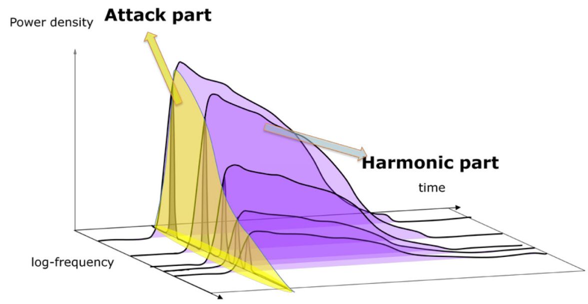

Figure 1. Attack-Harmonic model of single note

single notes which lead to unreliable feature parameters for identifying instruments. Weighted PCA can be applied to cope with this problem by giving different weights to overlapping and non-overlapping notes based on a certain criterion of judging whether overlapping or not.

3 New boosting for identification

The features extracted from the above sections are used in identification of instruments. A new boosting algorithm is proposed based on probabilistic decision where decision tree classifiers based on concept of information entropy [6] are used as weak classifier for generating probabilistic decision.

Boosting algorithm combines several weak classifiers to build a robust classification system. It increases the probability of choosing the training data which is easy to be misclassified. It uses updating rules to increase the probability assigned to those training data for which the classifier makes poor classification results and generates deterministic decision in every training iteration. In contrast, the new boosting algorithm uses probabilistic decisions for every classifier at the iterations of the boosting scheme by selecting the data from the training dataset, and combines them. It uses distribution of weights over the training events, at every iteration the weight of misclassified events is changed according to the accuracy of the classifier, forcing weak classifier

* ブースティング・アルゴリズムに基づく楽器認識、呉軍 [1]、齋藤大輔 [1]、亀岡弘和 [1,2]、嵯峨山茂樹 [1] (1:東大・情報理工,2:NTT・CS 研).

to focus on the events which are easily to be misclassified in the training set. In this way, it becomes more robust against unreliable features caused by overlapping notes.

New boosting algorithm:

Step 1. Assign initial weights $w = \{w_j = 1/N \mid j=1, 2, \ldots, N\}$ to the distribution of weights over the N training data.

Step 2. Choose T, the number of boosting iterations.

For i=1 to T:

Step 3. Generate a new classifier $h_i: X \rightarrow Y$ using the data set, and set M_{iy} the probability of h_i.

Step 4. For j=1 to N:

$$E_j = \begin{cases} w(x_j) \cdot M_{iy}(x_j), & y_i \neq h_i \\ w(x_j) \cdot \left(1 - M_{iy}(x_j)\right), & y_i = h_i \end{cases},$$

error rate $\varepsilon_i = \frac{1}{N} \sum_{j=1}^{N} E_j$

Step 5. If $\varepsilon_i > 1/2$, set $w = \{w_j = 1/N \mid j = 1, 2, \ldots, N\}$ and go back to step 3.

Step 6. $\alpha_i = \log((1 - \varepsilon_i)/\varepsilon_i)/2$

$$w_{i+1}(x_j) = \begin{cases} \frac{w_i(x_j)}{z_j} \cdot \exp(\alpha_i), & y_i \neq f_i(x_j) \\ \frac{w_i(x_j)}{z_j} \cdot \exp(-\alpha_i), & y_i = f_i(x_j) \end{cases}$$

Step 7.

$$f_{\text{FINAL}}(x) = \operatorname{argmax}_{y \in Y} \sum_{j=1}^{T} \alpha_i(x) M_{jy}(x)$$

4 Experimental evaluation

We evaluated the accuracy of the proposed algorithm for musical instrument identification using (1) synthetic data consisting of randomly mixed 271 music notes (32 altosax, 36 guitar, 88 piano, 45 violin, 36 flute and 34 oboe) and (2) real performed polyphonic music signal [5] by referring with associated MIDI files. Table 1 shows the accuracies by Support Vector Machine (SVM), AdaBoost [2], and new boosting algorithm with weak classifier in [6]. The proposed boosting algorithm significantly outperformed SVM and AdaBoost in all cases.

Table 1. Accuracy of musical instrument identification algorithms

	Synthetic data (%)			Real-world data (%)		
Number of Instruments	2	3	4	2	3	4
SVM	71.80	57.4	49.0	50.5	47.0	42.4
AdaBoost	76.90	63.7	52.3	53.4	49.4	45.7
New boosting	79.00	71.7	54.9	57.3	52.5	48.6

5 Conclusion

For musical instrument identification, this paper proposed an approach to first decompose polyphonic music to single notes, then extract features from the model, use classifier with probabilistic decision, finally apply a probabilistic boosting algorithm for identifying each note to a specific instrument. The proposed algorithm outperformed the comparison approaches.

References

[1] J. Wu, E. Vincent, S. A. Raczyński, T. Nishimoto, N. Ono, S. Sagayama, "Polyphonic pitch estimation and instrument identification by joint modeling of sustained and attack sounds", IEEE Journal of Topics in Selected Signal Processing,vol. 5, no. 6, pp. 1124-1132, Oct. 2011.

[2] M. Healy, R. Sourabh, and D. Anderson: " Effects of Varying Parameters in Asymmetric AdaBoost on the Accuracy of a Cascade Audio Classifier," In Proceedings of the Southeastern Conference of the Institute of Electrical and Electronics Engineers pp.169-172, 2004.

[3] H. Kameoka, T. Nishimoto, S. Sagayama, "A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering," IEEE Trans. on Audio, Speech and Language Processing, Vol. 15, No. 3, pp. 982-994, 2007.

[4] K. Miyamoto, H. Kameoka, T. Nishimoto, Nobutaka Ono, S. Sagayama, " Harmonic-Temporal-Timbral Clustering (HTTC) For the Analysis of Multi-instrument Polyphonic Music Signals," Proc. of ICASSP, pp.113-116, Apr., 2008.

[5] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: Popular, classical, and jazz music database," in Proc. ISMIR, 2002, pp. 287–288.

[6] J. Wu, Y. S. Kim, C. H. Song and W. D. Lee: "A New Classifier to Deal with Incomplete Data," The Ninth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD2008), pp.105-110, Thailand, 2008