Musical instrument identification based on new boosting algorithm with probabilistic decisions

Jun Wu, Emmanuel Vincent, Stanislaw Raczynski, Takuya Nishimoto, Nobutaka Ono, Shigeki Sagayama

MUSICAL INSTRUMENT IDENTIFICATION BASED ON NEW BOOSTING ALGORITHM WITH PROBABILISTIC DECISIONS

Jun Wu[*], Emmanuel Vincent†, Stanisław Andrzej Raczyński[*], Takuya Nishimoto[*], Nobutaka Ono[*] and Shigeki Sagayama[*]
[*]The University of Tokyo, Tokyo 113–8656, Japan
E-mail: { wu, raczynski,nishi,onono,sagayama }@hil.t.u-tokyo.ac.jp
†INRIA, Centre de Rennes - Bretagne Atlantique, 35042 Rennes Cedex, France
E-mail: emmanuel.vincent@inria.fr

ABSTRACT

This paper describes a new approach in musical instrument identification, an important task in the field of Music Information Retrieval (MIR). It is based on our previously developed probabilistic model which approximates the input audio spectrogram with a mixture of Gaussians. The EM algorithm is used to estimate the model parameters and calculate our newly proposed Harmonic Temporal Timbre Energy Ratio and Harmonic Temporal Timbre Envelope Similarity features. We then use these features in a novel boosting algorithm to perform the instrument classification. Contrary to traditional boosting methods, like the very popular AdaBoost, our new method uses probabilistic decision-making for hypotheses in each iteration, which results in better noise handing and higher classification accuracy.

Keywords

Musical instrument identification, Probabilistic harmonic model, boosting algorithm.

1. Introduction

The Music Instrument Identification research is an important problem in MIR. It has both scientific and practical applications. Although it has been considered as difficult problem, some approaches dealing with single instrument identification have recently been developed such as using Cepstral coefficient [1], Temporal features [2], Spectral features[3]. For more difficult problem which is to identify the multi-instrumental polyphonic music, some previous research has been done such as: Using frequency component adaptation with given correct F0s[4]. Using Missing feature theory with given correct F0s[5]. And using feature weighting to minimize influence of sound overlaps with given correct F0s[6]. However, all of these researches need to have given correct F0 as the basic condition while in real application the correct F0 is not given actually. In this paper a model capable of estimating F0 and deriving features for instrument identification is proposed in our previous research[7]. For using supervised approach for instrument identification, AdaBoost algorithm is often used by researchers for Music Information Retrieval because it is the most famous boosting algorithm[8]. However, AdaBoost may not be the most suitable approach because it uses a deterministic decision method during the iterations. But actually the decision for musical instrument in the model is probabilistic. Therefore, a new boosting algorithm based is proposed in this paper.

2. Proposed Model

The proposed approach can be divided to be three parts in this paper: (1) musical instrument model (2) feature extraction (3) classification.

(1) In our previous work, a method called Harmonic-Temporal Clustering (HTC) is proposed for multipitch analysis. It achieved the highest score in the task of Multiple Fundamental Frequency Estimation at MIREX 2009. Later an extension to HTC called HTTC for the analysis of individual audio signals within a multi-instrument polyphonic music to estimate their pitch, onset time, power and duration of all the acoustic events was proposed. However, this unsupervised classification method does not guarantee high accuracy for identification of musical instruments. In this paper we propose a new probabilistic harmonic model which is capable of estimating F0 and extracting features for instrument identification. The proposed model decomposes the spectrogram of the input signal into a mixture of individual acoustic events. It is modeled with an acoustic model with a 2-dimensional harmonic and temporal structure. Unlike conventional frame-wise approaches, the proposed model deals with both harmonic and temporal structures simultaneously, which leads to high estimation accuracy.

(2) In polyphonic music, different signals are very often overlapped so the analysis and identification of each signal or each pitch are difficult. For solving this problem, we need to retrieve as much information from each signal or pitch as possible to find the specific instruments' patterns and identify them. The characteristic of instruments' spectral energy of each harmonic partial can be used for identifying specific instrument. There are many differences between the shapes of the harmonic partials and the temporal structure of different musical instruments. Therefore we consider that the characteristic in timbre of specific instrument is derived from the difference of harmonic temporal timbre energy and harmonic temporal timbre envelope shape. The shapes of acoustic events classified into the same timbre category or same instrument should look alike regardless of the pitch, power, onset timing and duration. Besides the spectral envelope features and temporal features, we define the Harmonic Temporal Timbre Energy Ratio (HTTER) and Harmonic Temporal Timbre Envelop Similarity (HTTES). HTTER defines the features of the energy ratio of the harmonic temporal timbres. HTTES defines the difference between the envelop shapes of the harmonic temporal timbres.

(3) To increase the accuracy of classification, MIR researchers usually choose AdaBoost. However, in the case of musical instrument identification, AdaBoost may not be the most suitable approach because it uses a deterministic decision making method. The primary benefit of using boosting systems is the reduction of variance and increase in confidence of the decision. The decision obtained by any given classifier may be different from each other even if the model structure is kept constant. Therefore, combining the outputs of

several such classifiers by certain means may reduce the risk of selecting a poorly performing classifier. The most popular boosting method, AdaBoost uses distribution of weights over the training events and, at successive iterations, the weight of misclassified events is changed according to the accuracy of the classifier. It forces the weak classifier to focus on the hard events in the training set. However, problems with such an approach appear when the training events contain much noise, event number is too small for learning, etc. These problems commonly occur in the field of musical instrument identification and in such cases AdaBoost does not produce sufficiently stable results. To cope with these problems, a new boosting algorithm based on probabilistic decision making is proposed instead of the original AdaBoost, which involves deterministic decision. The update rule reduces the probability assigned to those events for which the hypothesis makes good predictions and increases the probability of the events on which the prediction is poor. The proposed new boosting algorithm uses probabilistic decisions for every hypothesis at the iterations of the boosting scheme, selecting the data events from a dataset, and then combines them. It uses distribution of weights over the training events: at each iteration the weight of misclassified events is changed according to the accuracy of the classifier, forcing the weak learner to focus on the hard events in the training set. It is more robust to noise in the data set and able to deals with it efficiently.

New boosting algorithm:

Step 1. Initially assign weights w = {$w_j = 1/N$ | j=1, 2, …, N}to be the distribution of weights over the N training events.

Step 2. Choose k to be the number of the boosting rounds.

For i=1 to T do:

Step 3. Generate the new classifier using data sets. Get back a hypothesis $h_t: X \to Y$, we set M_{ty} to be the probability of h_t for every $y \in Y$, Y is the output space.

Step 4. Compute the error rate ε_t as

For j=1 to N do

If $y_i \neq h_t$

$E_j = w(x_j) \cdot M_{ty}(x_j)$

If $y_i = h_t$

$E_j = w(x_j) \cdot (1 - M_{ty}(x_j))$

End for

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

Step 5. If $\varepsilon_i > \frac{1}{2}$, then set $w = \left\{w_j = \frac{1}{N}\middle| j = 1, 2, \ldots, N\right\}$ and go back to step 3.

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

Step 7. For each x_j

If $y_i \neq f_i(x_j)$

Then $w_{i+1}(x_j) = w_i(x_j)/Z_j \cdot exp(\alpha_i)$

If $y_i = f_i(x_j)$

Then $w_{i+1}(x_j) = w_i(x_j)/Z_j \cdot exp(-\alpha_i)$

End for

End for

Step 8. $f_{FINAL}(x) = argmax_{y \in Y} \sum_{j=1}^{T} \alpha_i(x) M_{jy}(x)$

Let $\{(x_j, y_j)|$ j=1,2, …, $N\}$ denote a set of N training examples. New AdaBoost calls a given weak classifier repeatedly in a series of rounds t=1, 2…..T. The main idea of the algorithm is to maintain a distribution or set of weights over the training set. The weight of this distribution on training datum j on round I is denoted as $w_i(x_j)$. Initially, all weights are set equally, but after each round, the weights of incorrectly classified data are increased so that the weak learner is forced to focus on the examples which are more difficult to classify in the training set. The weak learner's job is to find a weak hypothesis f_i appropriate for the distribution w_i. The distribution is obtained by normalizing a set of weights assigned to each event based on the classification performance of the classifiers on that event (Step 1). Step 2 chooses the number of boosting rounds. The larger number of iteration may give higher accuracy but costs more time. In this paper the decision tree classifier capable of giving probabilistic decision for every hypothesis is used. We take $M_{ty}$to be the probability of h_t for every $y \in Y$ while Y is the output space. Then we generate a new classifier using data selected form the data set and get back a hypothesis h_t (Step 3). The goodness of a weak hypothesis is measured by its error rate. The importance of a base classifier f_i depends on its error rate, which is defined as in Step 4. After calculating all of the training data, the error rate ε_i is computed using formula $\varepsilon_i = \frac{1}{N} \sum_{j=1}^{N} E_j$. If $\varepsilon_i > 1/2$, current f_i is discarded, a new training subset is selected and a new f_i is generated. (Step 5) The importance of a classifier f_i is calculated in Step 6. The α_i parameter is also used to update the weight of the training samples. $w_i(x_j)$ denotes the weight assigned to datum (x_j, y_j) during the ith boosting round. The weight update mechanism is used in Step 7. Z_j is the normalization factor which is used to ensure that$\sum_j w_{i+1}(x_j) = 1$. The weight update mechanism increases the weights of incorrectly classified examples and decreases the weights of those correctly classified examples. The final hypothesis $f_{FINAL}(x)$ is a weighted majority vote of the T weak hypotheses where $\alpha_i(x)$ is the weight assigned to hypothesis h_t and $M_{jy}(x)$ is the probability of h_t for every $\mathrm{y} \in \mathrm{Y}$. (Step 8)

3. Experiments

Overall, the proposed algorithm is intuitive and efficient in dealing with the problem of musical instrument identification, which was shown by the experiments. Recognition accuracy of instrument identification when using 12-dimensional MFCCs and the proposed features is shown. The accuracy of identifying the correct instrument is calculated for each pitch from the polyphonic test signals for mixtures of 2 instruments, 3 instruments and 4 instruments. The proposed algorithm outperforms the MFCC features by 9.5% for 2 instruments task, 10.2% for 3 instruments task and 15.5% for 4 instruments task on average. Experiments also show the accuracy of musical instrument identification when using different algorithms: SVM, AdaBoost, and the proposed new boosting algorithm. Again, the proposed algorithm showed an improvement over the previously used techniques.

4. Conclusion

In this paper we have proposed new features that we calculate from a probabilistic harmonic model and use for instrument identification. We have also proposed a new boosting algorithm based on probabilistic decision making for every hypothesis at every iteration of the boosting scheme. The algorithm uses distribution of weights over the training events: at each iteration the weight of misclassified events is changed according to the accuracy of the classifier, forcing the weak learner to focus on the hard events in the training set. The proposed algorithm is intuitive and efficient in dealing with the problem of musical instrument identification, which was shown by the experiments.

Reference

[1] J C. Brown, “Computer identification of musical instruments using pattern recognition with cepstral coefficients as features,” Journal of the Acoustical Society of America, vol. 105, no. 3, pp. 1933–1941, 1999.

[2] A. Eronen and A. Klapuri, “Musical instrument recognition using cepstral coefficients and temporal features,” in 14 EURASIP Journal on Advances in Signal Processing Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, pp. 753–756, Istanbul, Turkey, June 2000.

[3] G. Agostini, M. Longari, and E. Pollastri, “Musical instrument timbres classification with spectral features,” EURASIP Journal on Applied Signal Processing, vol. 2003, no. 1, pp. 5–14, 2003.

[4] T. Kinoshita, S. Sakai, and H. Tanaka, “Musical sound source identification based on frequency component adaptation,” in Proceedings of IJCAI Workshop on Computational Auditory Scene Analysis, pp. 18–24, Stockholm, Sweden, July-August 1999.

[5] J. Eggink and G. J. Brown, “Application of missing feature theory to the recognition of musical instruments in polyphonic audio,” in Proceedings of International Symposium onMusic Information Retrieval, Baltimore, Md, USA, October 2003.

[6] T. Kitahara, M. Goto, K. Komatani, T. Ogata, and H G. Okuno, “Instrument Identification in Polyphonic Music: Feature Weighting to Minimize Influence of Sound Overlaps,” EURASIP Journal on Advances in Signal Processing, Vol.2007, Article ID 51979, 2007.

[7] J. Wu, Y. Kitano, S. Raczynski, S. Miyabe, T. Nishimoto, N. Ono, S. Sagayama, "Musical Instrumental Identification Based on Harmonic Temporal Timbre Features," in Proceedings of Workshop on Statistical and Perceptual Audition (SAPA), pp. 7-12, 2010.

[8] 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.