

Support Vector Machine を用いた多重音の音源同定手法の検討 A Study on Sound Source Identification of Polyphonic Music Using SVM

北原 聡志† 野元 悠一† 甲藤 二郎†
Satoshi Kitahara Yuichi Nomoto Jiro Katto

1. まえがき

音源同定の技術は音楽のデータベース化と共に自動採譜システム、作曲支援、音楽検索システムなど様々なアプリケーションへの応用が期待される。我々はこれまでに楽器のカテゴリー分類を考慮して、2値分類の Support Vector Machine (SVM) を階層的に組合せることで音源同定を行う方法を提案し、単音入力に対して約 90[%] の認識率を得た [1]。しかしながら、この手法は和音入力を想定していないという問題が残っていた。そこで、本稿では最大 2 和音で構成された楽曲に対して音源同定を行う手法について述べる。その際、音源分離処理を施し単音に分けてから解析していく方法と和音のまま解析する方法の双方を試して精度比較を行ったので報告する。

2. 先行研究

音源同定の研究はその対象を単旋律から単一楽器の多重音、複数楽器による混合音へと複雑化させながら進歩してきた。現在、単音における研究では高い精度の報告もされているが、混合音を対象とした研究では未だ実用に至る精度は得られていない。例えば、単一楽器音を対象とした研究に Hidden Markov Model (HMM) を利用した手法がある [2]。ここでは、HMM の状態数と混合分布数を変化させながら最適なパラメータ値の検討を行っており、約 89[%] の認識率が報告されている。また、音源同定をパターン認識の応用と考えると、識別器に Gaussian Mixture Model (GMM) を利用した手法がある [3]。ここでは、5 種類の楽器の単音に対して平均 66[%]、二重奏入力に対して平均 49[%] の認識率が報告されている。その他、各楽器音データは楽器音の特徴空間上で基本周波数によって平均が変化する多次元正規分布に従うという仮定から、この分布のパラメータ推定法を提案し、音高による音色変化を基本周波数の関数として表現した F0 依存多次元正規分布 [4] など、様々な観点から研究が進められている。

3. 提案手法

3.1 Multiclass SVM の利用と楽器音特徴量

SVM の基本原理は 2 値クラスの識別器であるが、これを多クラスに拡張させた Multiclass SVM により和音の音源同定を実現する。本研究では One Versus the Rest で Multiclass SVM を構成する。現在研究対象としている楽器は、Classic guitar (Cg), Electric bass (Eb), Piano forte (Pf), Clarinet (Cl), Trumpet (Tr), Violin (Vn) であり、これら 6 種類の単音データベースから特徴量をそれぞれ抽出して学習させることで Multiclass SVM の楽器モデルを作成する。特徴量には Mel-Frequency Cepstral Coefficients (MFCC) を用いる。MFCC は音声認識の分野では一般的に 12~13 次元が用いられるが、

† 早稲田大学大学院理工学研究科

我々はこれまでの研究で楽器音の倍音構造や高周波成分の影響などを考慮した結果 26 次元分用いることで高い精度を得てきた [1]。また、フレーム一次差分の Δ MFCC および二次差分の Δ^2 MFCC を加えることで更に精度を高められることから、各々 26 次元の計 78 次元分を用いことにする。

3.2 音源分離を考慮した手法

音源分離処理では、「ハーモニック・クラスタリング」 [5] に基づいたクラスタリングを用いる。まず、基本周波数に対応する正規分布の平均 μ_k だけが自由度をもつ拘束付き混合正規分布によりモデル化し、EM アルゴリズムを適用することで、周波数軸上のスペクトルのクラスタリングを行う。次に、クラスタリングにより得られた各正規分布の平均 (基本周波数) μ_k と、各周波数成分が各正規分布に属する確率である帰属度 P_n^k を情報として分離を実現する。

$$P_n^k = \frac{w_n^k \cdot g(\omega | n\mu_k, \sigma^2)}{\sum_k \sum_n w_n^k \cdot g(\omega | n\mu_k, \sigma^2)} \quad (1)$$

[w :各正規分布の重み, g :正規分布]

ある正規分布に対する帰属度が高い周波数成分に注目し、その成分に対する累積確率とパワー値から、正規分布の母数を推定し、この母数と各成分に対する累積確率から各成分のパワー値を推定していく。

以上より得られた分離音 (単音) のそれぞれに対して、特徴量を順次計算していき、Multiclass SVM に通すことで識別を行う。

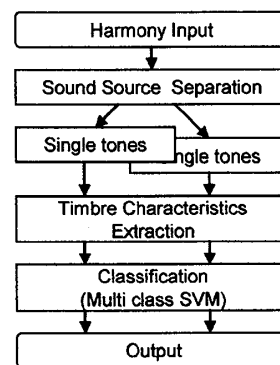


図1 音源分離を考慮した処理の流れ

3.3 和音解析による手法

続いて音源分離は行わず、和音のまま解析する手法について述べる。処理の流れを図 2 に示す。まず、曲の始めから終わりまで和音入力に対して特徴量を順次計算していく。これは、普通一曲を通して全ての楽器が同時に鳴り続けていることはなく、あるパートでは休符が挟まれ、一方の楽器のみが演奏している状況ができたり、楽器の種類によって音の立ち上がり時間や減衰時間に差があることから、曲

全体を見ることでその楽曲に使用されている楽器の特徴を万遍なく抜き出していき、次に Multiclass SVM に通すことで、楽器候補を同時に決定する。

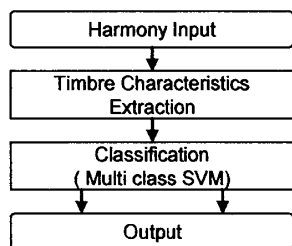


図2 和音解析による処理の流れ

3.4 補正処理

一般的に考えて、突然メロディー内で楽器の種類が変化することは少ないという音楽的知識に基づいた補正処理を加える。すなわち、SVM の出力結果を比較して、3.2 では最も多く認識された楽器にそれぞれ修正し、3.3 では最も多く認識された2種類の楽器に修正する。

4. 評価実験

4.1 予備実験

はじめに Multiclass SVM の性能を評価するために単音認識の予備実験を行った。

—実験条件—

学習用音源とテスト用音源は RWC 研究用音楽データベース、Sound Palette、McGill Master Samples の楽器音データベースより用意した。実験データの内訳を表1に示す。これらのデータベースには実際の楽器を演奏した音が単音で収録されており、楽器の個体差・音の強弱・奏法による音色の違いも含まれる。また、音域についても各楽器が奏でることのできる最低音から最高音までを選んでおり、A4=440[Hz]である。

表1 楽器音データベースの内訳

| 楽器 | Cq | Eb | Pf | Cl | Tr | Vn |
|--------|-------|-------|-------|-------|-------|-------|
| 音域 | E2-D5 | E1-E3 | E1-E7 | D3-D6 | B3-B5 | G3-G6 |
| 学習音源数 | 1296 | 473 | 655 | 185 | 143 | 370 |
| テスト音源数 | 652 | 224 | 245 | 106 | 123 | 142 |

—実験結果—

学習音源データとは異なるテスト音源データを用いて識別実験を行い、Multiclass SVM の性能を評価した結果、6楽器に対して平均 83.10[%]の正解率を得た。なお、評価方法は以下の式(2)で定義した。

$$\text{正解率} = \frac{\text{正解音源数}}{\text{テスト音源数}} \times 100 \quad (2)$$

4.2 音源同定実験

—実験条件—

予備実験で用いたテスト音源を繋ぎ合わせて、器楽の教科書に掲載されている既存の楽曲を作成し、入力メロディーとして評価実験を行った。各楽曲の使用楽器の組合せを表2にまとめる。各楽曲の曲長は30[s]-60[s]程度である。

—実験結果—

各手法の認識率の結果を表3にまとめる。音源分離を考慮した場合は 50.0[%]、音源分離を考慮しない場合は 81.25[%]の認識率となった。

—誤り率の検証—

(i) 音源分離ありのとき

音源分離ではそれぞれの単音を完全にクリアーな音として復元することは難しく、一方の楽器の音やノイズが少なからず含まれる。そのため分離した両方の音を同じ楽器と判断してしまうミスなどが見られた。

(ii) 音源分離なしのとき

実験で用いた6種類の楽器は大別すると減衰系(Cg,Eb,Pf)と持続系(Cl,Tr,Vn)から成る。音源分離せず解析する方法では、これら同一カテゴリー内の楽器同士で誤認識されるパターンが見られた。

表2 入力曲一覧

| 曲名 | 楽器1 | 楽器2 |
|----------|-----|-----|
| ソルフェージュ | Cq | Cl |
| | Pf | Tr |
| | Eb | Vn |
| 牧人の歌 | Pf | Tr |
| マルセリーノの歌 | Pf | Eb |
| | Cq | Eb |
| ムーンリバー | Pf | Eb |
| 喜びの歌 | Pf | Tr |

表3 認識率の比較

| | 音源分離あり | 音源分離なし |
|-----|--------|--------|
| 正解率 | 50.00% | 81.25% |

5. あとがき

本稿では識別器に SVM を利用し、和音楽曲の入力に対して音源同定を実現する手法として、音源分離を取り入れた方法と和音から解析していく方法をそれぞれ提案した。今後は和音の数をさらに増やすと共に、時間領域の特徴量を考慮することで精度の向上を図りたいと考えている。

参考文献

- [1] 北原聡志, 甲藤二郎, “楽器の階層的分類を考慮した SVM による音源同定”, 信学会 総合大会, 2006.
- [2] 馬場貴之, 山田武志, 北脇信彦, “HMM と MFCC を用いた楽器音の音源同定の検討”, 情報処理学会 研究報告 2005-MUS-61, pp.79-84, 2005
- [3] J. Eggink and G. J. Brown, “Instrument recognition in accompanied sonatas and concertos,” Proc. ICASSP 2003, pp. 553-556, 2003.
- [4] 北原鉄朗, 後藤真孝, 奥乃博, “音高による音色変化に着目した楽器音の音源同定: F0 依存多次元正規分布に基づく識別手法”, 情報処理学会論文誌, Vol.44, No.10, pp.2445-2455, Oct.2003
- [5] 亀岡弘和, 西本卓也, 嵯峨山茂樹, “ハーモニッククラスタリングによる多重音信号音高抽出における音源数とオクターブ位置推定”, 情処研報, 2003-MUS-51, pp.29-34, 2003.