情報処理学会論文誌 Vol. 50 No. 12 2910–2922 (Dec. 2009)

Random Forestを用いた音楽ジャンル分類

新妻雅弘[†1,*1] 斎藤博昭[†1]

本論文はRandom Forestを分類器とし，高次元の特徴量を用いた音楽ジャンル分類について述べる．従来の研究は，特定の音楽ジャンルに限られた特徴量，単旋律に関わる特徴量のみを用いており，広範な音楽ジャンル分類には対応できない．また，特徴量すべてについてチューニングを行うことで高い精度を得ている研究も多いが，ある音楽ジャンル分類に最適化したチューニングは過学習の問題からそのほかの音楽ジャンル分類にはそのまま適用できない．そこで本研究では，広範なジャンルを考慮した1,022次元の特徴量を用い，高次元，低サンプル数の問題に強いとされるRandom Forestを分類器とすることでチューニングをすることなく幅広い音楽ジャンルの分類に対応する手法を提案する．また，Out-Of-Bagデータを用いることで分類の際に重要な役割を果たしている特徴量を分析し，音楽ジャンル分類というタスクのメカニズムの解明に足がかりを与える．実験の結果，提案手法による分類精度はRootデータセットに対して97%，Leafデータセットに対して84%となり，チューニングを行っている先行研究と同等以上となった．また，訓練データに用いるサンプル数と精度の関係を検証した結果，提案手法がサンプル数の減少に対して優れた性能を持つことが示された．さらに，Out-Of-Bagデータを用いた変数の重要性推定に基づいた特徴量選択により，RootデータセットおよびLeafデータセットの各々のジャンル分類に固有な特徴量を分析できることを示した．

Music Genre Classification Using Random Forest

Masahiro Niitsuma[†1,*1] and Hiroaki Saito[†1]

This paper proposes a music genre classification method using high-dimensional features and Random Forest. Most of the previous work uses features designed to be adapted to only a limited range of music genres because of the data sparseness problem and tunes parameters in order to achieve high accuracy. However, feature selection often leads to over fitting and tuning parameters each time is not practical. In this paper, we propose music genre classification method using 1,022-dimensional features designed to be adapted to a wide range of music genres and Random Forest that easily handles high-dimensional features. Moreover, variable importance is estimated using out-of-bag data to clarify the mechanism of music genre classification. The proposed method has achieved 97% accuracy in root genre classification and 84% in leaf genre classification, which are favorable compared to other methods especially when the number of training samples is small. Moreover, variable importance estimation using out-of-bag data sheds new light on mechanism of music genre classification.

1. はじめに

近年のインターネットの発展および，mp3などの音楽圧縮技術の発達にともない，ネットワーク上に存在する音楽データの数が急増している．また，iPodなどのポータブルミュージックプレイヤの普及にともない，個人レベルで所有する音楽データの数も増加している．このような背景から，大量の音楽データを組織的に管理する必要性や大量の音楽データを対象に柔軟な検索を行いたいという需要が高まっている．そのような需要を受け，音楽情報検索（Music Information Retrieval）の研究は急速に発展している．古典的なコンピュータベースの音楽分析は文法ベースのテクニックかストリングマッチングのように非常に限られた手法で行われていたが，音楽情報検索研究団体の研究者は現代のパターン認識の技術を応用することで，高い成果をあげている[1)–6)]．なかでも音楽ジャンル分類の研究は近年非常にさかんになっており，Music Information Retrieval Evaluation eXchange（MIREX）のようなコンテストも毎年行われている．

しかし，従来の音楽ジャンル分類に関わる研究では，対象となるデータセットに依存した特徴量を用いたり，特徴量やその抽出区間の組合せをパラメータとし，訓練データにフィットさせたりすることで高い精度を達成しているものが多かった．これらの手法は，限られた音楽ジャンルを分類するには適しているが，より広い音楽ジャンルや未知の音楽ジャンルを対象とする場合，そのつどパラメータのチューニングを行わなければならず実用的でない．

そこで本論文では，広範なジャンルを考慮した1,022次元の特徴量を用い，高次元，低サンプル数の問題に強いとされるRandom Forestを分類器とすることでチューニングをすることなく幅広い音楽ジャンルの分類に対応する手法を提案する．さらに，Out-Of-Bagデータを用いることで分類の際に重要な役割を果たしている特徴量を分析し，音楽ジャンル分類

†1 慶應義塾大学大学院理工学研究科開放環境科学専攻
School of Science for Open and Environmental Systems, Keio University
*1 現在，英国クイーンズ大学博士課程
Presently with School of Music and Sonic Arts, Queen's University

のメカニズムに対する分析手法を提案する．

以下 2 章で関連研究について，3 章で提案手法について，4 章で実験および実験結果の考察について述べ，最後に 5 章でまとめを述べる．

2. 関 連 研 究

音楽ジャンル分類の研究は，MIDI[7] などの記号データを対象とするものと WAV などの音響データを対象とするものに大別される．これまでの音楽ジャンル分類の研究には，音響データを対象としたものが多い．確かに音響データの研究は，CD などの音源に対して直接適応できるという点で価値が高い．しかし，音響データからの音楽的特徴量の抽出は現在の技術では困難である．音楽的特徴量を用いることができるという点で，記号データを用いた研究にも多くのメリットがある．また，音響データを対象とした際に用いられる特徴量と，記号データを対象とした際に用いられる特徴量は異なっているが，そのほかの手法は一致しており，この特徴量の差異は近年の音響信号処理技術の発達にともない減少している．したがって，記号データを対象とした分類研究は音響データを対象とした研究に援用できる．

分類手法としては，主に教師なし学習と教師あり学習の 2 つがある．教師なし学習は，類似性を測定することにほかならず，音楽ジャンル分類以外の様々なアプリケーションに応用できる．たとえば，ユーザが特定のジャンルの音楽を探しているのではなく，特定の曲と近い音楽を探しているという場合には，教師なし学習を用いたシステムは大量の音楽データをナビゲートするために有用であると考えられる．しかしながら実際には，そのような研究は非常に限られた音楽ジャンルにしか適用されておらず，より大きなデータのより複雑な音楽ジャンル分類の場合には，教師なし学習による分類と人間による分類が一致する可能性は低い．そのため，ここでは主に教師あり学習を用いた研究を取り上げる．以下に分類対象とするデータごとに先行研究の概要を述べる．

まず，音響データを対象とした関連研究について述べる．Tzanetakis ら[8] は，音色とリズムに関わる特徴量を用いガウス過程に対する最尤推定（ガウス分類器）により 6 つの音楽ジャンル分類に取り組み，62%の精度を報告している．Girmaldi ら[9] は離散ウェーブレット変換を用いて時間，周波数に関わる 143 次元の特徴量を抽出し，分類器としては，k 近傍法のアンサンブルを用いて 73.3%の精度を報告している．Deshpande ら[10] は，k 近傍法を用いて Rock，Classical，Jazz の音楽ジャンル分類を行い，75%の精度を報告している．McKinney ら[11] は，4 つの異なる特徴量セットを用い，ガウス分類器により Jazz Folk，Electronica，R&B，Rock，Reggae，Vocal の音楽ジャンル分類に取り組み，61%から 74%の精度を報告している．Xu ら[12] は，サポートベクターマシン（SVM）を用いて Pops，Classical，Jazz，Rock の分類に取り組み，93%の精度を報告している．また，混合ガウスモデルや隠れマルコフモデルと比較して，SVM を用いた分類が高い精度となることを報告している．Xin ら[13] は，決定木をベース分類器とした Random Forest を用いて，それぞれ Country と Folk，Hip hop と Jazz，Blues と Jazz，Metal と Punk の分類を行い，RBF network，SVM，k 近傍法の各性能の比較をしている．また，特徴量の次元圧縮のために，主成分分析（PCA）だけでなく自己組織化マップ（SOM）を用いている．Lei ら[14] は，Random Forest と多層パーセプトロンの組合せを用い，Classical，Electronic，Jazz，Blues，Metal の分類を報告しており，精度としては最良で 83.7%となっている．特徴量として，MFCC および MPEG-7 記述子が用いられ，PCA により次元圧縮がなされている．

次に，記号データを対象とした関連研究について述べる．Chai ら[15] は，単旋律の動きに関わる情報を特徴量とし，隠れマルコフモデルを用いて民謡の分類に取り組み，63%の精度を得ている．Pedro ら[16] は単旋律の情報から旋律の動き，和音，リズムの情報を抽出し，SOM を用いて Jazz と Classical の分類に取り組み，77%の精度を得ている．また，最近では，単旋律における音高の幅や音価の平均値などの情報を用い，k 近傍法，SOM，ベイズ分類器をそれぞれ用いて Jazz，Pops を分類している[17]．Conklin[18] は，単旋律を分析することで得られる特徴量を用い，ベイズ分類器によりバッハのコラールと民謡の分類を行っている．Random Forest を用いた記号データに関わる研究としては，Rizo ら[19] が，CART をベースとした Random Forest を用いた主旋律分類について報告している．Classical，Jazz，Popular の 3 つのジャンルのデータセットを用いて実験を行い，すべてのジャンルのデータを用いて訓練を行った結果，すべてのジャンルで 80%以上，Jazz においては 98%の精度を報告している．使われている特徴量は旋律およびリズムに関わる 34 の特徴量に限られている．

以上のように，それぞれの研究で使用されている特徴量や分類手法は多岐にわたっており，その分類精度にも大きな幅がある．評価手法に一貫性がないため，単純に精度を比較することはできないが，多くの研究は特定の音楽ジャンルに限られた特徴量，特に単旋律に関わる特徴量のみを用いており，広範な音楽ジャンル分類には対応できない．また，どの特徴量が音楽ジャンル分類に寄与しているかは分からないため，用いる特徴量すべてについてチューニングを行うことで高い精度を得ている研究も多いが，ある音楽ジャンル分類に最適化したチューニングは過学習の問題からそのほかの音楽ジャンル分類にそのまま適用できない．また，様々な音楽ジャンルの分類を行う際にそのつどチューニングを行うことも実用的でない．

そこで本研究では，分類対象である音楽ジャンルが変化しても，特徴量の変更や新たな

チューニングの必要がない，幅広い音楽ジャンルに対応できる分類手法を提案する．そのためには，広範な音楽ジャンルを視野に入れた特徴量を用いる必要があるがそのような特徴量は一般に高次元になりがちであり，高次元の特徴量を用いる際に問題となるのが，データスパースネスの問題である．たとえば本研究で用いる McKay ら[20),21)] の提案する特徴量をすべて用いると，次元数は 1,022 にものぼり，特に細かい音楽ジャンル分類をする際には十分なサンプル数を得ることが困難である．そこで，本研究では，高次元，低サンプル数の問題に強いとされる Random Forest を用いることで，1,022 次元すべての特徴量を用い，チューニングをすることなく幅広い音楽ジャンルを適切に分類できることを示す．特に高次元，低サンプルの状況においても高い精度で音楽ジャンル分類が達成できることを他の分類手法と比較して示す．

Random Forest を用いた音楽ジャンル研究は従来いくつか報告されているが，いずれも分類前に次元圧縮を行うか，旋律やリズムに限られた低次元の特徴量のみを用いている．次元圧縮は多くの場合有効な手法であるが，分類に必要な情報が失われてしまう可能性もある．本研究では Random Forest を分類器とし，次元圧縮を行わずに音色などを含む高次元の特徴量をすべて用いることで，どれだけ高い精度が得られるかを検証する．さらに，Out-Of-Bag データを用いることで分類の際に重要な役割を果たしている特徴量を分析し，音楽ジャンル分類に固有な特徴量を考察する．

3. 提 案 手 法

本章では，高次元の特徴量を用い，Random Forest を分類器とした音楽ジャンル分類手法について述べる．以下，提案手法が用いる特徴量および分類手法を詳説する．

3.1 特　徴　量

音楽ジャンル分類に用いられる特徴量としては，音響信号から直接的に導出されるスペクトル情報や時間領域の情報などの低レベルの特徴量と，音楽的な訓練を受けた人間にとって意味を持つ音楽的な情報である高レベルの特徴量の 2 つがある．従来の音楽ジャンル分類研究においては，低レベルの特徴量を用いたものが多かったが，分類する粒度が細かくなればなるほど，より高レベルな特徴量が必要となる．

高レベルな特徴量の理論的な分析としてはいくつもの研究がなされている．Tzanetakis ら[8)] と LaRue[22)] は基本的な理論分析のサーベイに触れている．Cooper ら[23)] はリズムやメロディの分析について言及している．Reti[24)] は音楽における動機要素の分析について触れている．Tarasti[25)] は記号言語を用いた音楽分析について触れている．Temperley[26)] などの研究者によって開発された，高レベルな特徴量の分析システムも多く存在する．

しかしながら，これらの分析結果を特徴量として用いるためには多くの難点がある．第 1 に，これらの分析には，主観的な判断が介入する可能性が高い．たとえば，シェンカ分析や GTTM などの分析においては 2 人の専門家がいれば，2 つの異なった結果が出る可能性がある．第 2 に，計算量が膨大なものとなる．第 3 に，これらの特徴量は特定の音楽ジャンルのために用いられるものであり，広範な音楽ジャンルには適用できない．たとえば，和声分析の結果を主な特徴量として用いれば，和声を基礎とした音楽を分類することはできるが，和声を基礎としない民族音楽には適用できない．

一方で，民族音楽学者による既存の研究[27)–29)] は，特定の音楽ジャンルや分析手法に偏っていないという点で特徴量を設計する際に役立つ可能性が高い．そのような特徴量として，計算機への実装を考えて拡張された McKay ら[20),21)] の提案する特徴量がある．これらの特徴量は網羅的であり冗長さを持つ．McKay らによれば，これらすべての特徴量を用いることはデータスパースネスを引き起こすため，特徴量の選択が必要である．しかし，本研究の目的は広範な音楽ジャンルをチューニングなしに分類することであり，できるだけ漏れのない特徴量を用いる必要がある．そこで，本研究ではこの特徴量をすべて用い，3.2 節で述べる分類手法によりデータスパースネスの問題を解決する．以下に特徴量の概要を示す．

Instrumentation:

この特徴量は，General MIDI における 128 の音高を持つ楽器と 47 の打楽器のパッチ情報に基づく特徴量である．このパッチ情報をもとに，どのような楽器が使われているか，各時間にどのような楽器が発音しているか，打楽器の分布などが計算される．

Texture:

MIDI におけるチャネルやトラックの情報に基づく特徴量である．これらをもとに，いくつ独立した声部があるか，声部ごとの重要性の違いなどが計算される．

Rhythm:

ビートヒストグラムに基づく特徴量である．リズムに関わる特徴量を計算するためには，単純なビートトラッキングシステムを使う方法もある．しかしながら，ビートトラッキングシステムが提供する情報だけでは不十分なため，ここでは，McKay らによって修正された Tzanetakis らによるビートヒストグラム[30)] が用いられ，このビートヒストグラムをもとに，最も音量の大きい拍や，1 秒あたりの音価の平均などが計算される．

Dynamics:

音量に基づく特徴量であり，ベロシティをもとに計算される．

Pitch-Statistics:

音高の統計情報に基づく特徴量である．ここでは，ピッチヒストグラム[8] を基礎とし，それをもとに音域や，最も頻繁に現れた音高などが計算される．後述の Melody や Chords とは，音符の一時的な位置を考慮せず，曲全体での情報を用いる点が異なる．

Melody:

これは音高が現れる順序に基づく特徴量であり，連続する 2 音の音程のヒストグラムをもとに計算される．

Chords:

同時に発音する音程に基づく特徴量である．垂直音高のヒストグラムや，和音の種類のヒストグラムなどからなる．

特徴量の詳細については，McKay ら[21] を参照されたい．

なお，提案手法では，1 つの曲を 1 つの区間としてこれらの特徴量を抽出する．従来の音響ファイルを用いた研究では，特徴量の抽出区間も 1 つのパラメータとして，このパラメータと分類精度の関係を報告しているものもある[31]．記号ファイルを用いた研究でも，抽出区間の長さやオーバラップの区間をパラメータとし，分類精度がどう変化するかについて多くの研究がなされてきたが[17),32]，そのほとんどにおいて曲全体を抽出区間とすることの有効性が論じられている．そこで本研究では曲全体を特徴量の抽出区間とする．

3.2 分類手法

特徴量の抽出後，分類器による分類が行われる．本研究では 3.1 節で述べたように高次元のデータを特徴量とするため，従来用いられていた分類器を用いても高い精度は期待できない．

そこで，本研究では分類手法として，強力な集団学習の一種であり近年多くの応用研究が報告されている Random Forest を用いる．集団学習は精度の高くない弱分類器を複数組み合わせることで精度を上げる手法であり，Random Forest は Breiman[33] が発展させた集団学習法である．これまでの研究で，Random Forest は高次元，低サンプルのデータに強いとの報告がある．

Random Forest は以下の式で表される．

$$\{h(x,\Theta_k), k=1,\ldots,K\} \tag{1}$$

ここで Θ_k はそれぞれが独立かつ理想的に分布するブートストラップサンプルであり，x が入力を表す．また，$h(x,\Theta_k)$ は入力 x から抽出されたブートストラップサンプル Θ_k から導かれた分類器（以下ベース分類器と呼ぶ）を表し，この各々の分類器の出力の平均が Random Forest 全体の出力となる．

Forest を構成する，各々のベース分類器の作成にどのようなアルゴリズムを用いるかについては様々な手法が提案されているが，ここでは，古典的分類木である CART[34] を用いる．ベース分類器として CART を用いる際に，従来の CART と異なる点として，木の剪定を行わない点と，木の分岐の際にランダムに選ばれた変数の中から最適なものを選ぶ点があげられる．この手法の有効性は統計的学習理論の見地から動機づけられる．バイアス–分散分解によって，この分類器の誤り率はそのバイアス要素と分散要素によって決定されるため，これらを小さくすることが分類精度の向上につながる．x を入力ベクトル，T を訓練データ，集団学習器を構成する k 番目のベース分類器を，$f_k(x,T)=f(x,T,\Theta_k)$ で表す．なお，Θ_k は $k,1,\ldots,K$ 回目にランダムに抽出された入力ベクトルであり，理想かつ独立に分布していると仮定する．集団学習器 $F(x,T)=\frac{\sum_k f(x,T,\Theta_k)}{K}$ の分散 $Var(F)$ とバイアス $Bias(F)$ について，

$$Var(F)=E_{X,\Theta,\Theta'}[\rho_T(f(x,T,\Theta),f(x,T,\Theta'))]\times Var_T(f(X,T,\Theta)). \tag{2}$$

$$Bias^2(F)\le E_\Theta Bias^2(f(X,T,\Theta))+E(\varepsilon^2). \tag{3}$$

が成り立つ．なお，ρ_T は，2 つの異なる分類器の相関の平均値，ε は入力のノイズであり不可避である．式 (2) から，分散要素を下げるためには各々のベース分類器ごとの相関を小さくする必要があることが分かる．式 (3) から，全体のバイアスは，1 つの分類木のバイアス要素の 2 乗の期待値と入力のノイズの 2 乗の平均値の和であるから，各々の分類木のバイアスを下げることで小さくできることが分かる．CART を用いた Random Forest においては，各々の分類木を最大まで成長させることで各々の分類木のバイアスを小さくしている．さらに，ブートストラップサンプルを用いかつ木の分岐の際にランダムに選ばれた変数を用いることで，各々の分類木ごとの相関を小さくしている．

ベース分類器として CART を用いた場合の Random Forest の分類アルゴリズムは以下のようになる．

ステップ（1）：

与えられたデータセットから $ntree$ 組のブートストラップサンプル $B_1,B_2,\ldots,B_k,\ldots,B_{ntree}$ を作成する．その際，用いるデータセットの約 3 分の 1 はテスト用として取り除き，残りを学習用とする．テスト用として取り除いたデータを Out-Of-Bag（OOB）データと呼ぶ．

ステップ（2）：

各々のブートストラップサンプル B_i を用いて分類木 T_i を生成し，木の生成に用いていな

い OOB データを用いてテストを行う．その誤り率を OOB 推定値（OOB error rate）と呼ぶ．T_i の構築を行う際の各分岐ノードは，異なる木を多数生成するため，ランダムに $mtry$ 個の変数をサンプリングし，その中から最も分岐が良い変数を用いる．

ステップ（3）：

分類器は，すべてのブートストラップサンプル B_i の OOB 推定値に基づいて多数決をとる．

Random Forest において調整可能なパラメータは，組み合わせるベース分類器の数である $ntree$ と木の分岐の際に用いる変数の数である $mtry$ の 2 つだけである．これらの変数を最適化するアルゴリズムも提案されているが，Random Forest の挙動はこの 2 つの変数にそれほど敏感ではない．提案手法においては，チューニングをしないことを前提としているため，Breiman の推奨する $ntree = 500$，変数の総数を N として $mtry = \sqrt{N}$ とした．

3.3　変数の重要性の推定

本論文では，OOB データを用いた音楽ジャンル分類のメカニズムの分析についても言及する．

OOB データは本来，Random Forest の内部で誤り率を予測するために用いられるが，これを用いることで変数の重要性を推定することができる．具体的には，ある変数のすべての値を OOB データ内でランダムに順序変更することで，次元数を保ったままその変数と目的変数の関係をなくす．その結果 OOB 推定値がどれだけ増加するかを測定し，それにより変数の重要性を推定する．この手法は近年注目を集め，その有効性も検証されている[35]．アルゴリズムは以下のようになる．

ステップ（1）：

k 番目の分類木の OOB 推定値 M_k を計算する．

ステップ（2）：

k 番目の分類木に対応する OOB データの，m 番目の変数のすべての値を OOB データ内でランダムに順序変更する．

ステップ（3）：

ステップ（2）で順序変更した OOB データに k 番目の分類木を適用し，新たな OOB 推定値 M_k' を計算する．

ステップ（4）：

これをすべての $1, \ldots, k, \ldots, ntree$ に対して行い，$|M_k{}' - M_k|$ の平均値を m 番目の変数の重要性とする．

4. 実　　験

本章では，提案手法の妥当性を検証するための実験について述べる．第 1 の実験では，提案手法により従来手法と比較して高い分類精度を得ることができるかを検証する．第 2 の実験では，訓練データのサンプル数が精度に及ぼす影響を他の分類手法と比較して検証する．第 3 の実験では，OOB データを用いた変数の重要性の推定について述べる．

実験においては，現在広く用いられている McKay らによるデータセットを用いた．最近では Cataltepe ら[36] がこのデータセットを用いて音楽ジャンル分類に取り組み，精度を報告しているため，単純に精度を比較することができる．**表 1** にその概要を示す．Root は従来からよく用いられている，Western Classical，Jazz，Popular の 3 つの音楽ジャンルからなるデータセットであり，Leaf はこの Root をさらに細かく，Rap，Country，Baroque，Bebop，Jazz Soul，Modern Classical，Punk，Romantic，Swing の 9 つの音楽ジャンルに分類したデータセットである．

以下の実験はいずれも，Debian GNU/Linux，Intel（R）Xeon（R）CPU X5450，3.00 GHz，メモリ 48 GB のマシンで行った．すべての分類アルゴリズムの実装には R version2.7.0 およびそのパッケージを用い，特に記さない限り，デフォルトパラメータを用いた．特徴量の抽出には jSymbolicFE 12.0.0 [21] を用いた．

4.1　分類精度検証実験

第 1 の実験は，提案手法により，従来手法と比較して高い分類精度が得られるかを検証する実験である．ここでは 10-fold-cross-validation を用いて精度を計算した．

提案手法である 1,022 次元の特徴量および Random Forest（RF）を用いた手法に対する比較対象として，1,022 次元の特徴量と SVM を用いた場合，PCA を用いて次元圧縮を行った 34 次元の特徴量と Random Forest を用いた場合，同様の 34 次元の特徴量と SVM を用いた場合の 4 通りの手法を用意した．

PCA は相関係数行列を対象とし，固有値が 1 以上の主成分を選択した．SVM において

表 1　実験に用いたデータセット概要
Table 1　Dataset used for the experiment.

データセット名	Root	Leaf
総クラス数	3	9
サンプル数	250	250
1 曲の長さの平均（sec）	207.84 s	207.84 s

表 3　Leaf データセットに対する分類結果の混同表
Table 3　Confusion matrix of leaf genre classification.

Leaf	Rap	Country	Baroque	Bebop	Soul	Modern	Punk	Romantic	Swing	class.error
Rap	49	0	0	0	1	0	0	0	0	0.02
Country	0	25	0	0	0	0	0	0	0	0.00
Baroque	0	0	25	0	0	0	0	0	0	0.00
Bebop	0	0	0	23	1	0	0	0	1	0.08
Soul	7	0	0	8	8	0	0	1	1	0.68
Modern	1	0	2	0	0	15	0	7	0	0.40
Punk	1	0	0	0	0	0	24	0	0	0.04
Romantic	0	0	2	0	0	4	0	19	0	0.24
Swing	0	0	0	0	1	0	0	0	24	0.04

表 2　Root データセットに対する分類結果の混同表
Table 2　Confusion matrix of root genre classification.

Root	Jazz	Popular	Western Classical	class.error
Jazz	67	7	1	0.1066
Popular	1	99	0	0.0100
Western Classical	0	0	75	0.0000

は，特徴量の正規化を行い，カーネル関数は実験外データを用いた予備実験の結果をふまえ，Gaussian RBF Kernel を用い，σ の自動推定を行った．

表 2 および，**表 3** は提案手法の OOB データを用いたテストによる分類結果の混同表である．Root データセットに対しては，Jazz の Popular への誤分類が目立つがそれ以外の音楽ジャンルはほとんど正しく分類できている．Leaf データセットに対しては，Jazz Soul（Soul）の誤分類および Modern Classic（Modern）の Romantic に対する誤分類が目立つ．Modern Classic と Romantic の分類は，同じ Classical Music の分類であり，人間においても誤分類が多いと考えられ興味深い．

図 1 および**図 2** は，それぞれ Root および Leaf のデータセットに対する，分類に用いるベース分類器の数と誤り率の関係を示している．ベース分類器の数が増加することで精度が高くなっているが，一方で，この精度の上昇は一定のラインで収束しており，いずれのデータセットに対しても $ntree = 500$ で十分であることが分かる．

次に，従来手法との比較について述べる．**表 4** に 10-fold-cross-validation の結果を示す．まず，Root データセットに対する音楽ジャンル分類結果について述べる．Cataltepe ら[36] の報告している精度 98%に対して提案手法では 97%の精度が得られている．Cataltepe らの

表 4　提案手法と SVM による分類精度の比較
Table 4　Classification accuracy of the proposed method compared to SVM with the same feature set.

特徴量（分類手法）	Root	Leaf
1,022 次元の全特徴量（RF）	0.97	0.84
1,022 次元の全特徴量（SVM）	0.84	0.54
PCA による 34 次元の特徴量（RF）	0.86	0.64
PCA による 34 次元の特徴量（SVM）	0.93	0.74

精度は特徴量の重み付けなどを最適化した結果であるのに対し，提案手法ではすべて自動で決定し，チューニングしていないことも考えれば優れた結果である．次に，Leaf データセットに対する音楽ジャンル分類結果について述べる．Cataltepe ら[36] の報告している精度は 70%程度である．これは，用いた特徴量が粗い音楽ジャンル分類である Root データセットだけにしか対応できていないことを意味している．一方で，提案手法においては 84%と非常に高い精度が得られており，用いる特徴量の次元を拡大することで，粒度の異なるデータセットであってもチューニングなしに高い精度での分類が可能となることを示している．他の研究ではデータセットが異なることから単純な比較はできないが，従来の研究で報告されている精度と同等かそれ以上の結果が得られている．Pedro ら[17] においては，Jazz，Pops，Classical の Root と同じ粒度のデータセットに対して抽出区間や特徴量に関するチューニングを行った結果，最良で 92%の分類精度を得ている．本研究では手動でのチューニングをしていないにもかかわらず，97%の精度を達成している．細かい音楽ジャンルに関しては，Cataltepe ら[36] が MIDI のみを用いた分類で最大 62%の結果を報告しているが，本研究では 84%と細かい音楽ジャンル分類に対しても同様に高い精度を保持していることが分かる．

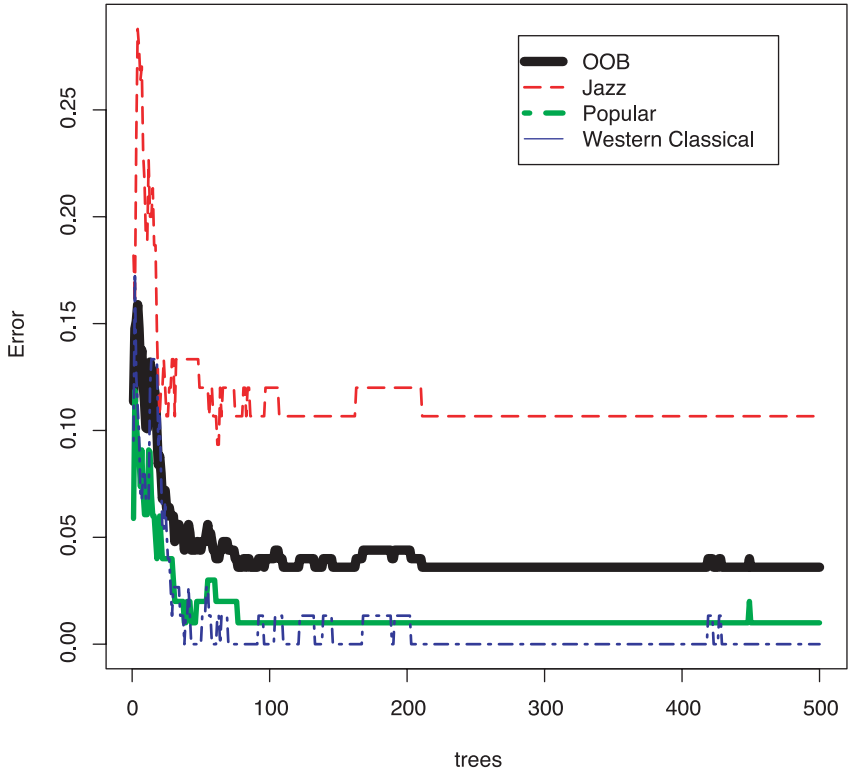

図 1　ベース分類器の数と誤り率の関係（Root）
Fig. 1　Tree size versus error rate (root genre classification).

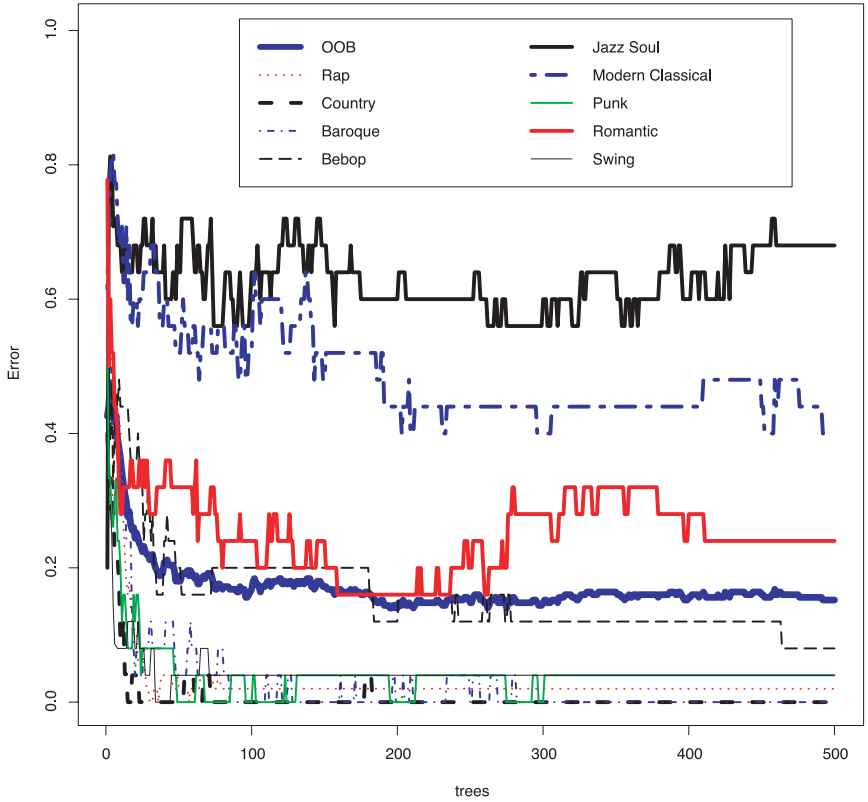

図 2　ベース分類器の数と誤り率の関係（Leaf）
Fig. 2　Tree size versus error rate (leaf genre classification).

SVM を用いた分類はいずれも提案手法より精度が劣っており，特に細かい音楽ジャンルである Leaf データセットについては劣っている．PCA により圧縮された 34 次元の特徴量を用いた場合，SVM では精度が向上しているが，Random Forest を用いた場合は精度が下がっており，全体として，提案手法である 1,022 次元の特徴量および Random Forest を用いた分類が最良の精度となっている．この結果は，次元圧縮にともなう情報喪失により分類精度が劣化する可能性を裏付けている．

図 3 および**図 4** は，それぞれ Root データセットおよび Leaf データセットに対する，分類に用いる特徴量の数と誤り率の関係を示している．なお，標準誤差を SE として点線が $\pm 2SE$ を表す．これらの図から，実際に精度の向上に寄与している特徴量は 100 程度であることが分かる．この研究では，これらの選択を行わずに精度を向上させることを意図しているためチューニングは行わないが，OOB 推定値をもとに特徴量をしぼることで精度をさらに向上させることが可能であると考えられる．

4.2　訓練に用いるサンプル数が精度に与える影響を検証する実験

第 2 の実験は，訓練に用いるサンプル数が精度に与える影響を検証する実験である．ここでは，第 1 の実験で用いたデータセットを用い，訓練に用いるサンプル数と分類精度の関

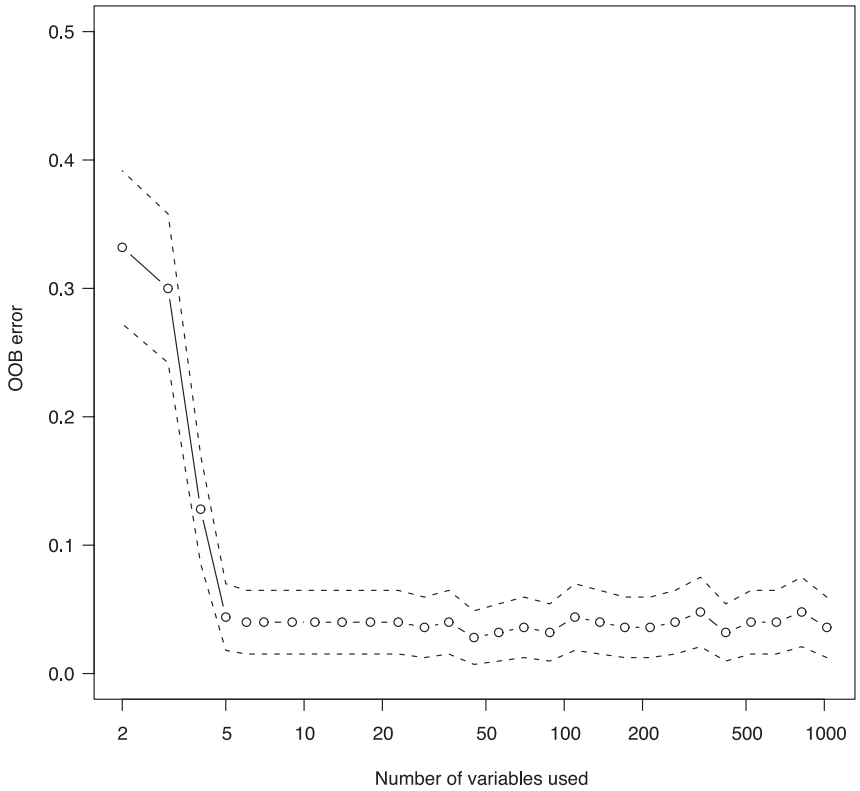

図 3　分類に用いる特徴量の数と誤り率の関係（Root）
Fig. 3　Variable size versus error rate (root genre classification).

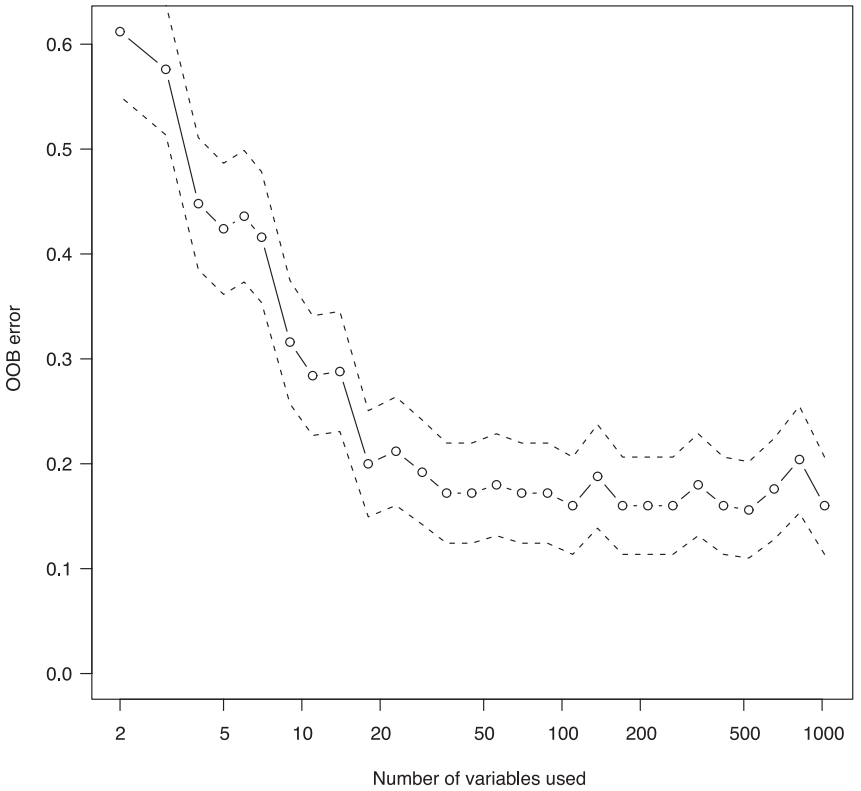

図 4　分類に用いる特徴量の数と誤り率の関係（Leaf）
Fig. 4　Variable size versus error rate (leaf genre classification).

係を検証する．各々のクラスのサンプル数は不均一であるため，均等にサンプル数を減らすとサンプル数が 0 になってしまう．そこで，これを防ぐために，$1, 2, \ldots, i, \ldots, 9$ に対して以下を行う．分類対象であるクラスを $1, 2, \ldots, k, \ldots, K$ として，各々のクラスの総サンプル数を N_k と記す．k 番目のクラスのデータから $N_k \times i/10$ 個のサンプルをランダムで抽出する．これをすべてのクラスに対して行い，それらの和を訓練データとする．すなわち，学習データの全サンプル数を N とすれば，i 番目の訓練データ数は $N \times i/10$ となる（以下，この i をサンプル比率（sample rate）と呼ぶ）．正確さのためこれらを 100 回繰り返し，各々の分類結果の F 値のマクロ平均を評価の基準とした．本研究では Random Forest のほかに，比較対象として，SVM，k 近傍法，単純ベイズ分類器（Bayes），LVQ，Boosting，Bagging を用いる．SVM は，第 1 の実験と同様に特徴量の正規化を行い，カーネル関数は Gaussian RBF Kernel を用い，σ の自動推定を行った．k 近傍法においては同様に実験外データによる予備実験の結果をふまえ，$k = 10$ とした．LVQ においては OLVQ1 を用いてコードブックを作成し，LVQ1 アルゴリズムで学習を行った．Boosting は CART をベース分類器とし，250 回反復した．なお，多クラス分類への拡張である Adaboost.M1 を用いた．

表 5　提案手法による分類結果の F 値のマクロ平均
Table 5　Macro-averaged F-measure of the proposed method.

データセット名	Root		Leaf	
サンプル比率（%）	標準誤差	平均値	標準誤差	平均値
10	0.003031	0.8746	0.006357	0.6529
20	0.002428	0.9089	0.004217	0.7695
30	0.002326	0.9220	0.004095	0.7927
40	0.002455	0.9349	0.003062	0.8064
50	0.002269	0.9454	0.003834	0.8182
60	0.002652	0.9497	0.003950	0.8276
70	0.002330	0.9575	0.005181	0.8368
80	0.002796	0.9541	0.007512	0.8315
90	0.004077	0.9594	0.008269	0.8347

表 6　分類手法ごとの F 値の標準誤差および平均値の全サンプル数に対する平均値（Root）
Table 6　F-measure and its standard deviation of root genre classification averaged for the size of training samples.

分類手法	標準誤差	F 値
RF	0.002707	0.9341
Bagging	0.003560	0.9115
Boosting	0.002927	0.9311
SVM	0.003963	0.8009
K-NN	0.004212	0.7528
LVQ	0.004391	0.7514
Bayes	0.003583	0.8204

Bagging は CART をベース分類器とし，ブートストラップサンプルの数は 500 とした．

表 5 に提案手法に対するサンプル数ごとの F 値の標準誤差および平均値を示す．Root データセットに対しては，訓練データのサンプル比率が 20%以上で 90%以上の F 値となっており，Leaf データセットに対しては，サンプル比率が 40%以上で 80%以上の F 値となっている．**表 6** および**表 7** に，それぞれ Root データセットおよび Leaf データセットに対する各分類手法の F 値の標準誤差および平均値の全サンプル数に対する平均値を示す．どちらのデータセットにおいてもすべての分類手法の中で，Random Forest を用いたものが最良の F 値となっている．また，どちらのデータセットにおいても Random Forest が最も標準誤差が小さく，安定した精度となっていることが分かる．**図 5** および**図 6** は，それぞれ Root データセットおよび Leaf データセットに対する訓練データのサンプル比率と分類結果の F 値の関係を分類手法ごとに示している（なお，エラーバーは 95%信頼区間を表す）．全体として，Random Forest，Boosting，Bagging の集団学習の分類手法が優れた精度となっており，Random Forest が最も高い精度を出している．特に，訓練データに用いるサンプル数の減少に対する精度の減少が Random Forest において最も少なく，低サンプルデータに強いという特徴が現れている．

表 7　分類手法ごとの F 値の標準誤差および平均値の全サンプル数に対する平均値（Leaf）
Table 7　F-measure and its standard deviation of leaf genre classification averaged for the size of training samples.

分類手法	標準誤差	F 値
RF	0.005164	0.7967
Bagging	0.006606	0.7197
Boosting	0.005979	0.7537
SVM	0.007669	0.4295
K-NN	0.006497	0.4338
LVQ	0.009151	0.3615
Bayes	0.006619	0.5978

図 5 においては，サンプル数の非常に少ない状態（サンプル比率が 10%）では Random Forest の方が精度が高いが，それ以外では Boosting の精度が高い．Boosting は Random Forest と異なり特徴量すべてを用いるため，データ量が十分に存在する場合には，Random Forest より高い精度が得られたものと考えられる．図 6 においては，全体を通して Random Forest が最良の精度となっている．これは，細かい音楽ジャンルである Leaf データセットは，Root データセットよりも特徴量の分散が大きく，分類精度が訓練データのサンプル比率に対してより敏感になっているためである．このことから，より詳細な音楽ジャンル分類には Random Forest が適していると考えられる．

4.3　変数の重要性推定を検証する実験

第 3 の実験は，OOB データを用いた変数の重要性に基づき，音楽ジャンル分類のメカニズムを分析する実験である．ここでは 3.3 節で述べた OOB データを用いた変数の重要性推定手法をもとに，Root データセット，Leaf データセットという 2 つの異なる音楽ジャンル分類に固有な特徴量を分析する．

まず，1,022 次元の特徴量の中で特に各々のジャンル分類に寄与しているものだけを選択する．ここでの目的は，分類精度を実質的に変えずに特徴量の数を減らすことであり，そのためにはある程度の誤差を許容する必要がある．そこで，1-SE Rule[34),37)] にならい，標準誤差を基準とし，最良の分類精度の標準誤差の範囲内で，最も特徴量の数が少ない組合せを選択する．アルゴリズムは以下のようになる．

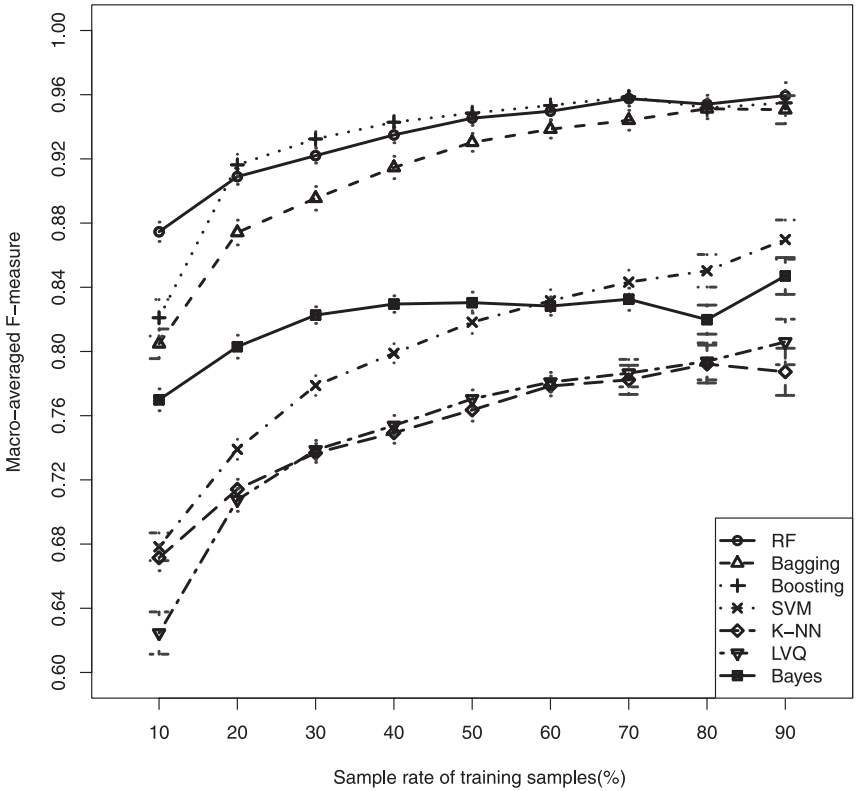

図 5　訓練データのサンプル比率と分類結果の F 値の関係（Root）
Fig. 5　Number of training samples versus macro-averaged F-measure of root genre classification.

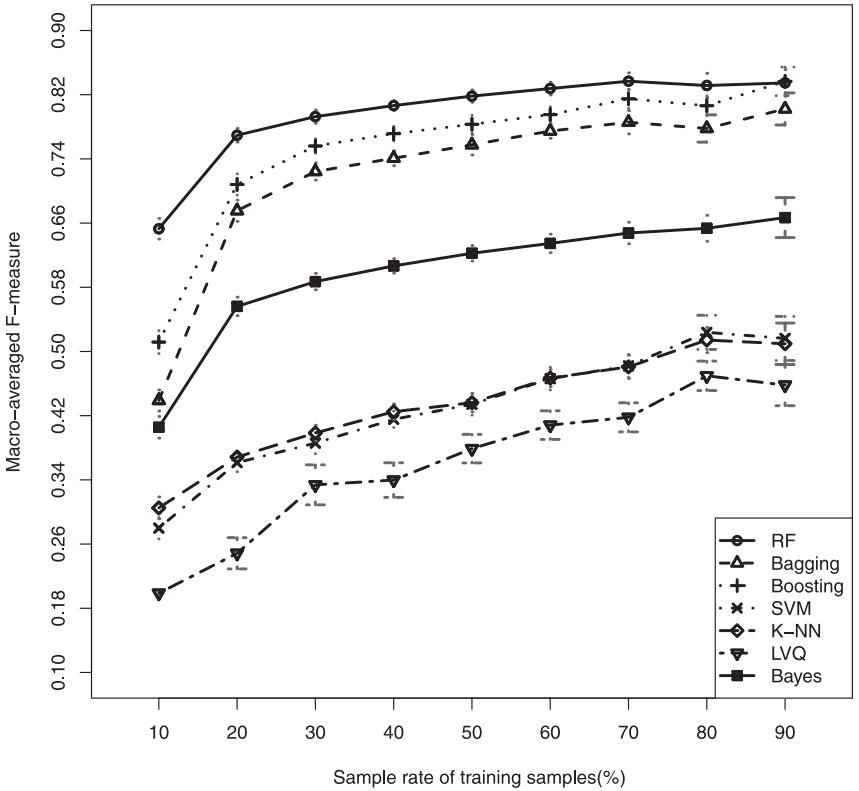

図 6　訓練データのサンプル比率と分類結果の F 値の関係（Leaf）
Fig. 6　Number of training samples versus macro-averaged F-measure of leaf genre classification.

ステップ（1）：

現在の特徴量をすべて用いて Random Forest を構成し，OOB 推定値 err およびその標準誤差 se を計算する．

ステップ（2）：

現在の特徴量の組合せ，err および se を記録し，特徴量の総数が 2 以上ならば 3.3 節で提案した変数の重要性が最も低い特徴量を削除し，ステップ（1）を繰り返す．

ステップ（3）：

ステップ（2）のすべての特徴量の組合せの中で，最良の OOB 推定値およびその標準誤差を err' および se' とし，OOB 推定値が $err' \pm se'$ の範囲に収まる特徴量の組合せの中で最も特徴量の数が少ないものを選択する．

このアルゴリズムにより，Root データセットに対しては 29，Leaf データセットに対しては 36 の特徴量が選択された．これらが分類に有効な特徴量であれば，データスパースネスにより精度が優れなかった SVM の精度は，これらの特徴量のみを使うことで向上するは

表 8　推定された変数の重要性に基づく特徴量選択による精度向上を示す表
Table 8　Improved accuracy of SVM with the selected features compared to that with all features.

データセット名	Root	Leaf
全特徴量（SVM）	0.84	0.54
選択された特徴量のみ（SVM）	0.97	0.81

表 9　Root データセットの分類を特徴づける特徴量
Table 9　Unique features of root genre classification.

番号	特徴量の説明
A-1	ビートヒストグラム
A-2	ビートヒストグラムにおける周波数の最も高い 2 つの Beat Bin の周波数の和
A-3	三全音の旋律進行の割合
A-4	最も頻度の高い音高の割合
A-5	音高の平均値
A-6	ビートヒストグラムにおける 2 番目に周波数の高い Beat Bin の周波数
A-7	Note on イベント間の長さの標準偏差

ずである．**表 8** は OOB 推定値を指標とした変数の重要性の推定アルゴリズムに基づいて，特徴量の数を減らした際の SVM の精度を表す．分類精度は，Root データセットにおいて 13%，Leaf データセットにおいては 27%向上しており，この結果は PCA を用いて次元圧縮を行った場合よりも優れている．OOB データを用いた特徴量の選択は，PCA と異なり分類精度そのものを基準としているためこのような結果となったと考えられる．この結果から，提案手法によって，各音楽ジャンル分類に特に有効な特徴量が選択されていることが分かる．特にデータの分散が大きく，データスパースネスが発生していると考えられる Leaf データセットに対しては非常に大きな効果を発揮している．

提案手法によって選択された特徴量は各々の音楽ジャンル分類に有効な特徴量である．しかし，互いに重複する特徴量は音楽ジャンル分類一般に用いられる特徴量であり，各々に固有の特徴量ではないと考えられる．そこで，これらの差分を調べれば，Root と Leaf という 2 つの音楽ジャンル分類において，それぞれに固有の特徴量が分かると考えられる．この結果を**表 9** および**表 10** に示す．Leaf データセットの分類に固有の特徴量のうち，B-1, B-3，B-5，B-6 はすべてある楽器に属する音の割合であり，楽器に関わる特徴である．一方で，Root データセットに対しては，ビートヒストグラムや音高の平均値など，音高やリズムに関わるものが多いことが分かる．これらの特徴量が，実際に人間が各音楽ジャンル分類を行っている際に注目している特徴量と同一であるかを評価することは難しいが，この結果は音楽ジャンル分類のメカニズム解明に 1 つの足がかりを与えたといえる．

表 10　Leaf データセットの分類を特徴づける特徴量
Table 10　Unique features of leaf genre classification.

番号	特徴量の説明
B-1	アコースティックギターパッチに属する Note on イベントの割合
B-2	Note on イベント間の長さの平均値のチャネルごとの平均値
B-3	ブラスパッチに属する Note on イベントの割合
B-4	曲の長さ
B-5	エレキギターパッチに属する Note on イベントの割合
B-6	サキソフォンパッチに属する Note on イベントの割合

5. おわりに

本論文では，高次元の特徴量を用いた分類手法として CART をベース分類器とする Random Forest を用いた音楽ジャンル分類手法を提案し，その有効性を示すため 3 種類の実験を行った．

第 1 の実験では，Western Classical，Jazz，Popular の 3 つの音楽ジャンルからなる Root データセットおよび，この Root データセットをさらに細かく分類した，Rap，Country，Baroque，Bebop，Jazz Soul，Modern Classical，Punk，Romantic，Swing からなる Leaf データセットを用いて，チューニングをすることなく分類を行い，Root データセットで 97%，Leaf データセットで 84%の精度を達成した．これは，チューニングをともなう従来研究と同等以上の精度である．第 2 の実験では，訓練データのサンプル数が精度に及ぼす影響を検証し，CART をベース分類器とする Random Forest が他の分類手法よりも訓練データのサンプル数の減少に対して優れていることを示した．第 3 の実験では，OOB データを用いることで，音楽ジャンル分類においてどの特徴量が重要であるかを推定できることを確認し，これをもとに 2 つの異なる音楽ジャンル分類においてそれぞれに固有な特徴量を分析できることを示した．

OOB データを用いた特徴量の重要性推定は音楽学などへの応用が期待できるが，その音楽学的意義については音楽学者との議論をともなったさらなる実証研究が必要である．また，より細かい音楽ジャンル分類や音響ファイルに対する提案手法の有効性の検証も今後の課題である．

参 考 文 献

1) Brown, M. and Dempster, D.J.: The scientific image of music theory, *Journal of Music Theory*, Vol.33, No.1, pp.65–106 (1989).
2) Nicholas, C.: *A guide to musical analysis*, Oxford: Oxford University Press (1987).
3) Huron, D.: The melodic arch in Western folksongs, *Computing in Musicology*, Vol.10, pp.3–23 (1996).
4) Conklin, D. and Witten, I.: Multiple viewpoint systems for music prediction, *Journal of New Music Research*, Vol.24, No.1, pp.51–73 (1995).
5) Conklin, D. and Anagnostopoulou, C.: Segmental pattern discovery in music, *INFORMS Journal on Computing*, Vol.18, No.3, pp.285–293 (2006).
6) Hamanaka, M., Hirata, K. and Tojo, S.: Implementing "A generative theory of tonal music", *Journal of New Music Research*, Vol.35, No.4, pp.249–277 (2006).
7) Huber, D.M.: *The MIDI Manual, 3rd Edition: A Practical Guide to MIDI in the Project Studio*, London: Focal Press (2007).
8) Tzanetakis, G. and Cook, P.: Music genre classification of audio signals, *IEEE Trans. Speech and Audio Processing*, Vol.10, No.5, pp.293–302 (2002).
9) Grimaldi, M., Kokaram, A. and Cunningham, P.: Classifying music by genre using the wavelet packet transform and a round-robin ensemble, Technical Report, Dublin, Ireland: Trinity College (2002).
10) Deshpande, H., Nam, U. and Singh, R.: Classification of music signals in the visual domain, *Proc. Digital Audio Effects Workshop* (2001).
11) McKinney, M.F. and Breebaart, J.: Features for audio and music classification, *Proc. International Symposium on Music Information Retrieval*, pp.151–158 (2003).
12) Xu, C., Maddage, N.C., Shao, X., Cao, F. and Tian, Q.: Musical genre classification using support vector machines, *Proc. International Conference on Acoustics, Speech and Signal Processing*, pp.429–432 (2003).
13) Jin, X. and Bie, R.: Random Forest and PCA for Self-Organizing Maps based Automatic Music Genre Discrimination, *DMIN*, Crone, S.F., Lessmann, S. and Stahlbock, R. (Eds.), pp.414–417, CSREA Press (2006).
14) Lei, W., Shen, H., Shijin, W., Jiaen, L. and Bo, X.: Music Genre Classification Based on Multiple Classifier Fusion, *International Conference on Natural Computation*, Vol.5, pp.580–583 (2008).
15) Chai, W. and Vercoe, B.: Folk music classification using hidden Markov models, *Proc. International Conference on Artificial Intelligence* (2001).
16) Ponce de León, P.J. and José, M.I.: Musical style identification using selforganising maps, *Proc. International Conference on Web Delivery of Music*, pp.82–89 (2002).
17) Ponce de León, P.J. and José, M.I.: Pattern recognition approach for music style identification using shallow statistical descriptors, *IEEE Trans. Systems, Man and Cybernetics. Part C, Applications and Reviews*, Vol.37, No.2, pp.248–256 (2007).
18) Conklin, D.: Melodic analysis with segment classes, *Machine Learning*, Vol.65, No.2-3, pp.349–360 (2006).
19) Rizo, D., Ponce de León, P.J., Perez-Sancho, C., Pertusa, A. and Iñesta, J.M.: A Pattern Recognition Approach for Melody Track Selection, *Proc. International Conference on Music Information Retrieval* (2006).
20) McKay, C. and Fujinaga, I.: Style-independent computer-assisted exploratory analysis of large music collections, *Journal of Interdisciplinary Music Studies*, Vol.1, pp.63–85 (2007).
21) McKay, C. and Fujinaga, I.: jSymbolic: A feature extractor for MIDI files, *Proc. International Computer Music Conference*, pp.302–305 (2006).
22) LaRue, J.: *Guidelines for style analysis*, Warren, MI: Harmonie Park Press (1992).
23) Cooper, G. and Meyer, L.B.: *The rhythmic structure of music*, Chicago: University of Chicago Press (1960).
24) Reti, R.: *The thematic process in music*, New York: Macmillan (1951).
25) Tarasti, E.: *Signs of music: A guide to musical semiotics*, New York: Mouton de Gruyter (2002).
26) Temperley, D.: *The cognition of basic musical structures*, Cambridge, MA: MIT Press (2001).
27) Lomax, A.: *Folk song style and culture*, Washington: American Association for the Advancement of Science (1968).
28) Cumming, J.E.: *The motet in the age of Du Fay*, Cambridge: Cambridge University Press (1999).
29) Tagg, P.: Analysing popular music: Theory, method and practice, *Popular Music*, Vol.2, pp.37–67 (1982).
30) Tzanetakis, G., Essl, G. and Cook, P.: Automatic Musical Genre Classification of Audio Signals, *Proc. International Conference on Music Information Retrieval* (2001).
31) Bergstra, J., Casagrande, N., Erhan, D., Eck, D. and Kégl, B.: Aggregate features and ADABOOST for music classification, *Machine Learning*, Vol.65, No.2-3, pp.473–484 (2006).
32) Pickens, J.: A Survey of Feature Selection Techniques for Music Information Retrieval, Technical Report, Center for Intelligent Information Retrieval, Department of Computer Science, University of Massachusetts (2001).
33) Breiman, L.: Random Forests, *Machine Learning*, Vol.45, pp.5–32 (2001).
34) Breiman, L., Friedman, J., Stone, H.J. and Olshen, R.: *Classification and Regression Trees*, London: Chapman & Hall CRC (1984).

35) Archer, K.J. and Kimes, R.V.: Empirical characterization of random forest variable importance measures, *Computational Statistics & Data Analysis*, Vol.52, No.4, pp.2249–2260 (2008).
36) Cataltepe, Z., Yaslan, Y. and Sonmez, A.: Music genre classification using MIDI and audio features, *EURASIP Journal on Advances in Signal Processing*, Vol.2007, No.1, pp.150–158 (2007).
37) Ripley, B.D.: *Pattern recognition and neural networks*, Cambridge: Cambridge University Press (1996).

(平成 21 年 3 月 8 日受付)
(平成 21 年 9 月 11 日採録)

新妻　雅弘

2009 年慶應義塾大学大学院理工学研究科修士課程修了．現在，英国クイーンズ大学博士課程（School of Music and Sonic Arts）在籍中．音楽学および音楽情報処理の研究に従事．

斎藤　博昭（正会員）

1991 年慶應義塾大学大学院理工学研究科修了．工学博士．現在，慶應義塾大学理工学部情報工学科准教授．自然言語処理および音楽情報処理の研究に従事．言語処理学会，電子情報通信学会各会員．