低音旋律の潜在意味解析による音楽ジャンル分類*

○上田雄 (東大・工), 角尾衣未留, 小野順貴, 嵯峨山茂樹 (東大・情報理工)

1 はじめに

個人がハードディスク上に音楽をデータベースとして保持したり、インターネットなどを通じて新しい音楽が次々と生産・提供されるような時代となった現在、膨大な量の音楽の整理、検索などの目的のための自動音楽ジャンル分類が注目されており、多くの研究がなされている。

各音楽ジャンルには特に低音旋律において音高推移・リズムが共通しているものが多い。例えば、ロックではルート音が連続するのに対し, ジャズではウォーキングベースと呼ばれる経過音を多く含むような音型が多く用いられるように、低音旋律が一定のパターンを持っていると考えられる。そのため、それらを抽出することが可能であれば、音楽ジャンル分類を実現できるであろう。

低音の特徴量を用いた従来研究では、低音旋律の音高推移を用いた土橋ら [1] の研究があり、ジャンル認識における低音旋律の有用性が示されている。

本稿では低音旋律に着目し、その音楽的遷移の潜在意味解析を用いて自動音楽ジャンル分類を行い、性能を評価する。

2 低音旋律の楽譜情報に基づく音楽ジャンル分類

2.1 潜在意味解析

低音旋律の音高推移、リズムの特徴は音楽ジャンルにおいて非常に有益であると考えられる。このような情報は楽譜にシンボリックな形で与えられており、それを解析する問題は自然言語における文章中の単語を解析する問題と似ている。そのため、低音旋律の特徴パターンを単語、楽曲を文書とみなして自然言語処理等で用いられる手法である潜在意味解析 (Latent Semantic Analysis:LSA) [2] を適用することができると考えられる。

LSA では、各楽曲に含まれる低音旋律の特徴パターンの出現回数を表した特徴パターン-楽曲行列が用いられる。その行列における列は特徴パターンに対応し、行が楽曲に対応する。特徴パターン-楽曲行列 W の (i,j) 成分 $w_{i,j}$ を重みづけし、多くの楽曲で使われる特徴パターンの影響を抑え、少ない楽曲にしか含まれない特徴パターンを強調することにより、概念の離れた楽曲をより遠くする効果がある。特徴パターンが M 種類、楽曲が N 曲である $M \times N$ の特徴パターン-楽曲行列 W において $c_{i,j}$ を楽曲 j 中で特徴パターン i が現れた回数、n_j を楽曲 j 中の特徴パターンの数、N を楽曲数 N_i を特徴パターン i を含む楽曲数として行列の (i,j) 成分は

$$w_{i,j} = (1-\epsilon_i)\frac{c_{i,j}}{n_j} \tag{1}$$

$$\epsilon_i = -\frac{1}{\log T}\sum_{j=1}^{N}\frac{c_{i,j}}{t_i}\log\frac{c_{i,j}}{t_i} \tag{2}$$

と表される。多くの楽曲に含まれる特徴パターン i における ϵ_i の値は 1 に近づき、少ない楽曲にしか含まれない場合は ϵ_i の値が 0 近づく。そのため多くの楽曲で使われる特徴パターンの影響が抑えられ、少ない楽曲にしか含まれない特徴パターンが強調される。

Table 1　評価に用いた RWC 音楽データベースのそれぞれのジャンルにおける曲数

Genre	Pop	Rock	Dance	Jazz	Latin
Number	59	48	113	63	84

Table 2　特徴パターン-楽曲行列の重みづけによるジャンル分類認識率と、特異値分解による次元を低減した場合の ジャンル分類認識率

Features	Accuracy
Baseline	30.8%
Without SVD	40.1%
With SVD	49.1%

2.2 特異値分解

得られた行列 W は非常に次元が大きくスパースになることが多いため、一般に特異値分解による次元の低減が行われる。

$$W \approx \hat{W} = USV^T \tag{3}$$

ただし、U は $M\times Q$ 行列、S は $Q\times Q$ 行列、V は $N\times Q$ 行列で、Q は W の階数である。S は対角行列で、対角成分が $s_1 > s_2 > s_3 ... > s_Q$ と並ぶ。$\{s_i\}_{Q_0+1}^{Q} = 0$ として特異値を減らすことで $\hat{W}$ の階数は Q_0 となり、$\hat{W}$ は階数が Q_0 となる行列の内 W の最良の近似となっている。この操作は次元を低減させるだけでなく、データに含まれるノイズを除去する効果もある。

次元低減後の階数 Q_0 を決定する際、特異値の寄与を考慮すると

$$\frac{\sum_{i=1}^{Q_0} s_i^2}{\sum_{i=1}^{Q} s_i^2} \leq \tau \tag{4}$$

における閾値 τ を決定することで、妥当な Q_0 の値を決定することができるであろう。

2.3 特徴パターン選択

低音旋律の音楽的特徴は音高推移、リズムにより表現できると考えられる。音高推移は隣り合う音高の差、つまり音程を順に追うことで観測することができる。また、リズムは隣り合うオンセットのタイミングの組合せによって構成される。そのため、コンテキスト解析に広く用いられる n-gram モデルに基づき、音高推移は音程、リズムはオンセット間隔 (Inter-onset Interval) の比を特徴パターンとして捉えることができる。

2.4 MIDI データを用いた評価実験

楽譜情報が正確に取得できている場合について、評価実験を行った。用いたデータは RWC 研究用音楽データベース [3] のうち、Pop, Rock, Dance, Jazz, Latin の 5 ジャンルの MIDI データを 30 秒ごとに分割したものを用いた。それぞれにおける曲数は Table 1 であった。分類は機械学習ツールキット weka[4] を利用し、分類器は線形のサポートベクトルマシーンを用いた。

特徴パターンは半音毎の上下を含めた音程 120 通り、2 や 0.5 などに量子化した IOI 比 81 通りとし、それぞれの uni-gram、bi-gram を用いた。それらが

*Music Genre Classification by Latent Semantic Analysis of Bass Line . by UEDA, Yushi, TSUNOO, Emiru, ONO, Nobutaka, SAGAYAMA, Shigeki (The University of Tokyo)

楽曲に現れた階数をカウントし、重みづけされた特徴パターン-楽曲行列を計算した。音程に関する特徴量は 14520 次元、IOI 比に関する特徴量は 6642 次元で、合計 21162 次元の特徴量が並んだ行列であった。

重みづけのみをした場合と、$\tau = 0.8$ で特異値分解をし次元を低減した場合について分類を行った結果は Table 2 であった。表に示されているように特異値分解により次元が低減されたため認識率が向上し、ランダム分類器によるベースライン性能を上回り、低音旋律はジャンル分類に有用であることが確認できた。

3 音響信号への拡張

3.1 音響信号からの特徴抽出における問題点

前節では予め楽譜情報が与えられていたが、音響信号から同様の特徴パターンを取得するには以下の 3 つの問題が考えられる。

- 打楽器音の存在によるオンセット位置の推定誤りの可能性
- 他の楽器の存在による基本周波数・オンセット位置の推定誤りの可能性
- 楽器の倍音構造による基本周波数の推定誤りの可能性

次の節でこれらの問題を解決する手法を提案する。

3.2 音響信号からの楽譜情報の抽出

上に挙げた一つ目の、楽曲が打楽器を含むという問題に対しては、宮本ら [5] による調波音・打楽器音の分離を利用することが考えられる。これは調波音と打楽器音のスペクトログラム上での特徴の違いに着目し、マスクを設計することによりそれぞれを分離する手法である。また、二つ目の問題の他の楽器の存在は、低音旋律楽器が最も低い周波数帯域にあるとするとローパスフィルタによって抑圧することができるであろう。最後の問題は低音旋律は単音であることを仮定すると、上で述べた二つの手法により低音旋律の単音ピッチ推定の問題と捉えられるため、Yegnanarayana ら [6] のアルゴリズムにより基本周波数の推定、Klapuri[7] のアルゴリズムによりオンセット位置を推定することで特徴パターンを抽出できると考えられる。

3.3 低音旋律のピッチ推定

低音旋律のピッチ推定は以下のように行われる。対象とする音楽音響信号から分離された調波音にローパスフィルタをかけた信号を $s[n]$ とする。その差 $x[n] = s[n] - s[n-1]$ を零周波数フィルタに 2 回通し、周辺 10ms 程度の平均を引くことにより基本周波数正弦波に近い信号 $y[n]$ が得られる。

$$y_1[n] = -\sum_{k=1}^{2} a_k y_1[n-k] + x[n] \tag{5}$$

$$y_2[n] = -\sum_{k=1}^{2} a_k y_2[n-k] + y_1[n] \tag{6}$$

ただし $a_1 = -2, a_2 = 1$ とする。$y[n]$ は

$$y[n] = y_2[n] - \frac{1}{2N+1} \sum_{m=-N}^{N} y_2[n+m] \tag{7}$$

である。$y[n]$ の positive zero-crossing のインターバルからその時刻の周波数を推定することができる。

3.4 低音旋律のオンセット位置推定

オンセットを推定するために、Klapuri の手法では、$s[n]$ をバンドパスフィルタを用いて周波数帯域に分け、それぞれでオンセットを検知した後、全体を統合する。各々の帯域でのオンセットは、100ms の half-Hanning 窓を畳み込むことによって振幅エンベロープ $A(t)$ を計算し、その対数の微分

$$W(t) = \frac{d}{dt}(\log(A(t))) \tag{8}$$

の単純なピークピッキングにより推定される。最終的なオンセットは各帯域でのオンセット候補の時間的に近いものをまとめ、音量の閾値により推定される。

Table 3　特徴パターン-楽曲行列の重みづけによるジャンル分類認識率と、特異値分解による次元を低減した場合の ジャンル分類認識率

Features	Accuracy
Baseline	10.0%
Without SVD	29.7%
With SVD	25.2%

3.5 評価実験

音響信号からの楽譜情報を取得した場合の評価実験を行った。用いたデータは GTZAN データセット [8] の 10 ジャンル、Blues、Classical、Country、Disco、Hiphop、Jazz、Metal、Pop、Reggae、Rock の各ジャンル 100 曲、合計 1000 曲の WAV ファイルを用いた。全ての楽曲は 30 秒で 22.05kHz サンプリング、1ch 信号であった。

特徴パターンは半音毎の上下を含めた音程 68 通り、IOI 比は 81 通りとり、それぞれの uni-gram、bi-gram を用いた。音程に関する特徴量が 4492 次元、IOI

に関する特徴量が 6642 次元で、合計 11134 次元であった。

2.4 節の評価実験と同様に線形 SVM を用いて音楽ジャンル分類を行った。特異値分解の際は、閾値 $\tau = 0.8$ とした。特異値分解を行わない場合、行う場合における認識率を Table 3 に示した。ベースライン性能を有意に上回ったものの、特異値分解によって認識率を下げる結果となった。原因としては、音響信号からの楽譜情報の推定精度が十分でなかったことが考えられる。

4 まとめ

本研究では低音旋律の音高推移、リズム特徴の潜在意味解析により音楽ジャンル分類を行った。低音旋律のパターン情報を得るために、音程・IOI 比に基づく特徴量を提案した。実験により楽譜情報を適切に抽出できた場合、LSA によって分類性能を向上させ、低音旋律特徴量がジャンル分類に対し有用であることを確認した。

今後の課題としては、本稿で提案した音楽音響信号からの特徴抽出の精度を向上させることが考えられる。

参考文献

[1] 土橋佑亮他, 情報処理学会研究報告, 2008-MUS-74-38, pp. 217–224, 2 月, 2008.

[2] S. Deerwester, S. Dumais, G. Furnas, T. Landauer and R. Harshman, J. Am. Soc. Info. Sci. vol.41, no.6, pp. 391–407, 1990.

[3] 後藤真孝他, 音講論 (春), pp. 843–844, 3 月, 2003.

[4] I. Witten and E. Frank: Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 2005.

[5] 宮本賢一他: 音講論 (春), pp. 903–904, 3 月, 2008.

[6] B. Yegnanarayana, K. Sri Rama Murty and S. Rajendran, Proc. ITRW, Aalborg, Denmark, 2008.

[7] A. Klapuri, Pro. ICASSP, pp. 3089–3092, 1999.

[8] G. Tzanetakis and P. Cook, IEEE Trans. SAP, vol. 10, no. 5, pp. 293–302, 2002.