

打楽器・低音旋律パターン特徴とクロマベクトルの相互相関 特徴量を用いた自動ムード認識*

角尾 衣未留, 赤瀬太一, 小野順貴, 嵯峨山茂樹 (東大・情報理工)

1 はじめに

近年のインターネット上の音楽ファイル増加に伴い、音楽情報検索の研究が盛んに行われるようになったが、関連する研究の中でも、ムード分類に代表される自動楽曲分類は非常に重要である。このタスクにおいては、小節単位の打楽器パターンや低音旋律に関する特徴量は非常に有力であると考えられる。本研究ではそのようなムードを代表する小節単位のパターンをクラスタリングし、抽出する事によって自動楽曲分類の精度向上を目的とする。

従来の、自動楽曲分類の研究ではジャンル分類において音色特徴量、ピッチ情報特徴量、リズム情報特徴量がそれぞれ抽出されるが [1]、音色特徴量以外は十分に有効であるとは言い難く、音色特徴量のみが広く用いられるようになった。これを音楽ムード分類に利用した研究もあり [2]、ムードの定義の曖昧性から音楽ムード分類の難しさが示された。その曖昧性に対応するため機械学習に人手を加えたアクティブラーニングについても発表された [3]。しかしこれらの研究は統計的な特徴のみを扱っており、時間情報が効率的に用いられていなかった。

本研究では、入力音響信号の小節境界を推定し、小節単位の打楽器パターンと低音旋律パターンを抽出する手法を提案する。また、それらの情報を従来の統計的な音色情報や音楽のスケール情報と共に用いることにより音楽ムード分類を行い、その有効性を検証する。

2 小節単位のパターン情報の抽出

2.1 小節境界推定

小節境界の推定問題は非常に難しい問題として未だよい解決策は提案されていない。しかし、特に近年の楽曲では打楽器が小節単位に繰り返している為、打楽器のパターンに着目することでそれを推定することができると考えられる。しかし、一般に楽曲は打楽器音以外にも旋律楽器によって構成される場合があり、また、テンポが変動したり打楽器のパターンが複数種類存在する可能性があり、それらは当初は未知である。

これらの問題を解決するために動的計画法の一種である One-pass DP 法と k -means クラスタリング法を組み合わせたアルゴリズムが提案されてきた [4]。この手法では調波音・打楽器音分離手法 [5] を前処理として打楽器音が強調され、小節単位の打楽器パターンとその小節への最適なセグメンテーションが同時に推定される。初期の複数の打楽器パターンを基に動的計画法でアラインメントを計算し、 k -means クラスタリングに基づいてテンプレートパターンを平均計算により更新する。この二つのステップを繰り返

すことにより収束後は、最適なテンプレートとセグメンテーションが得られ、ここで得られるセグメンテーション情報を用いる事で小節境界を与える事ができる。

2.2 打楽器パターンの抽出

打楽器パターンはムードを特徴付ける特徴の一つであると考えられ、ムードに共通の打楽器パターンを推定することによってこれらの自動楽曲分類に有効であると考えられる。

上述の小節境界推定手法は打楽器音に着目しており、小節境界と同時に打楽器パターンも推定するため、このアルゴリズムを同じムードの楽曲のコレクションに適用することによってムードに共通な打楽器パターンを抽出することができる [6]。 k -means アルゴリズムによるクラスタのセントロイドの計算の際には、複数のテンプレートから One-pass DP 法によって全ての楽曲に対して得られたアラインメントに基づき、全ての楽曲の対応する箇所全てを平均化することによって実現する。

2.3 低音旋律パターンの抽出

低音旋律も同様にムードを特徴付けると考えられ、その単位パターンは上の打楽器パターン抽出と同様な枠組で実現することができる。唯一異なる点は同じ低音旋律パターンでも和声によってその音高がシフトするというピッチシフト問題があることである。

ここでは、 k -means クラスタリングの枠組でピッチシフト不変となるような距離尺度を定義した新しいクラスタリング手法 [7] を利用することが考えられる。複数のテンプレートをピッチ方向にシフトさせながら距離を計算し、それが最小となるテンプレートとシフト量に基づいて平均化しテンプレートを更新するやり方である。この距離計算と平均化によるテンプレート更新を繰り返す事により、収束後に複数の低音旋律パターンテンプレートが得られる。

3 特徴量抽出

3.1 小節パターン特徴量

上のようなアルゴリズムを用いて学習されたムード内で共通のパターンがムード間で有意に異なる事を期待した場合、理想的にはクラス未知の楽曲に対し、いずれのパターンが含まれているかを計算しクラス属性を解析する事によってムード分類が可能である。しかし、実際にはムードによっては明確にパターンが異なる事は稀であったり、確かにそのムードのパターンは含まれているが、異なる時刻には他のムードのパターンも含まれているなどの場合が考えられる。

そのため、そのような事象による認識誤差を減ら

* Audio mood classification using percussive and bass-line pattern features and chroma cross-correlation features. By TSUNOO, Emiru, AKASE Taichi, ONO Nobutaka, and SAGAYAMA Shigeki (University of Tokyo)

すため、抽出されたパターンに基づいて特徴量ベクトルとし、その特徴量空間でのベクトルを教師あり学習分類を用いることによりムードを認識することが考えられる。その特徴量抽出法は、楽曲を構成するパターンの属するクラスのヒストグラムやそれらパターンと全パターンとの距離などがありうる。

3.2 統計的なクロマベクトルの相互相関特徴量

音楽スケールも音楽ムードを特徴付けると考えられる。例えばドの音が鳴っている時にミの音が鳴れば和音は明るくなるし、ミフラットが鳴れば暗くなる。このような情報を抽出する方法の一つにクロマベクトルの相関の計算がある。

クロマベクトルは12次元のベクトルで各要素がオクターブの12半音のエネルギーの重ね合わせである。時刻 t のクロマベクトルの i 番目の要素を $c(t, i)$ とすると、正規化されたベクトルの相互相関は

$$S(t, \tau) = \sum_{i=1}^{12} \frac{\left(\frac{e(t)}{12} - c(t, i) \right) \left(\frac{e(t)}{12} - c(t, i + \tau) \right)}{e(t)^2} \quad (1)$$

となる。ただし $e(t) = \sum_{i=1}^{12} c(t, i)$ である。これは $\tau = 0$ の時は単にエネルギーであり、この相関値は対称である。そのため、6つの係数のみが有益な情報である。この平均と標準偏差が各楽曲に対して計算され特徴量として用いることが考えられる。

4 評価実験

4.1 データセット

ムード分類にはCAL500[8]を用いて評価実験を行った。このデータセットはAggressive, Fun, Passionate, Sadの4ムードでそれぞれ180曲である。データセットは22.05Hz、1ch信号にダウンサンプリングしたものをを用いた。低音旋律の帯域を82.4Hzから330.0Hzとしてローパスフィルタ処理を行った。

4.2 テンプレートの学習と特徴量抽出

まずは前述のアルゴリズムに基づいて各ムードに共通の打楽器・低音旋律パターンを学習した。各ムードのデータを半分に分けた90曲でクラスごとに打楽器パターンは10個のテンプレート学習を行い、低音旋律パターンは20個のテンプレートを学習した。これらを用いて別の半分のデータから打楽器パターン特徴量：ヒストグラム(4次元)と、低音旋律パターン特徴量：各テンプレートとの距離の曲での平均(80次元)を抽出した。次節での認識率計算の際は二つの半分のデータの結果を平均化した。

4.3 特徴量による分類とその結果

特徴量空間における分類器の学習については、マシンラーニングのツールキット、“Weka”を用いた[9]。識別器として線形SVMを用い、10フォールドの交差確認を行った。比較対象としてはTzanetakisらの特徴量を用いた。これはMIREX[10]のムード認識のタスクで高い認識率であったシステムで、音色に関する特徴量、つまりMFCC等のスペクトログラムに関する情報の統計量を用いたもの(68次元)であった。この特徴量を用いた認識率と提案する打楽器・低音旋律パターン、クロマベクトル相互相関特徴量とを合

Table 1 ムード分類における従来の音色特徴量と提案特徴量と従来特徴量を組み合わせた場合の認識率の平均

Features	Accuracy
Existing (timbre, 68 dim.)	53.5%
Timbre + Chroma (80 dim.)	51.6%
Timbre + Rhythm/Bass (152 dim.)	55.9%
All Merged (164 dim.)	58.8 %

わせた場合での結果を表1に示す。クロマベクトルによる音楽スケール特徴量と音色特徴量を組み合わせただけでは従来の認識率を少し下回ったが、それに打楽器・低音旋律パターンを組み合わせた場合認識率に向上が見られるため、打楽器・低音旋律パターン特徴量の有効性を確認した。

5 おわりに

本研究では、音楽音響信号からの自動ムード分類を目的とした特徴量抽出として、ムード内で共通の小節単位の打楽器・低音旋律パターンを抽出するアルゴリズムを提案した。さらにクロマベクトルの相互相関を計算することで音楽スケールの特徴を抽出する手法を提案した。また、それらを用いた特徴量抽出を実際の楽曲に対し行い、打楽器・低音旋律パターン、クロマベクトルの相互相関特徴量の自動ムード分類における有効性を確認した。

今後の課題としては、和声進行情報を利用した自動ムード分類の精度向上や音楽情報検索への応用が考えられる。

謝辞 本研究の一部はCrestMuse Projectの支援を受けて行なわれた。

参考文献

- [1] G. Tzanetakis and P. Cook, “Musical genre classification of audio signals,” *IEEE Trans. on Speech and Audio Processing*, pp. 293–302, 2002.
- [2] T. Li and M. Ogihara, “Detecting emotion in music,” in *Proc. of ISMIR2003*, 2003, pp. 239–240.
- [3] M. Mandel, G. Poliner, and D. Ellis, “Support vector machine active learning for music retrieval,” *ACM Multimedia Systems*, pp. 3–13, 2006.
- [4] E. Tsunoo, N. Ono, and S. Sagayama, “Rhythm map: extraction of unit rhythmic patterns and analysis of rhythmic structure from music acoustic signals,” in *Proc. of ICASSP2008*, 2009, pp. 185–188.
- [5] N. Ono, K. Miyamoto, H. Kameoka, and S. Sagayama, “A real-time equalizer of harmonic and percussive components in music signals,” in *Proc. of ISMIR2008*, 2008, pp. 139–144.
- [6] E. Tsunoo, G. Tzanetakis, N. Ono, and S. Sagayama, “Audio genre classification using percussive pattern combined with timbral features,” in *Proc. of ICME2009*, 2009, pp. 382–385.
- [7] E. Tsunoo, N. Ono, and S. Sagayama, “Musical bass-line pattern clustering and its application to genre classification,” in *Proc. of ISMIR2009*, 2009, pp. 219–224.
- [8] D. Turnbull, L. Barrington, D. Torres, and G. Lanckriet, “Towards musical query-by-semantic-description using the cal500 data set,” in *Proc. of ACM SIGIR*, 2007, pp. 439–446.
- [9] I. Witten and E. Frank, “Data mining: practical machine learning tools and techniques,” *Morgan Kaufmann*, 2005.
- [10] “Mirex 2008,” <http://www.music-ir.org/mirex/2008>.