

# 機械学習を用いた楽曲の分類 ～音高に着目したテンポ変化に強い特徴抽出～

白川直人<sup>†</sup> 近山隆<sup>†</sup>

東京大学大学院工学系研究科<sup>†</sup>

## 1. はじめに

今日、私たちは半導体技術の発達により、デジタル化された大量の音楽データを iPod などのデバイスによって携帯し、いつでも音楽を楽しむようになった。近年の音楽配信の急激な成長に表わされるように、音楽はこれからますますデジタルアーカイブ化されていくことは確実である。そして膨大な量の音楽データを収集して構築されたデータベースの中から自動的に音楽検索・推薦を行う技術への需要は高まることが期待される。

人間は音楽を聴いたときに音楽の時系列構造を感覚的に分析して楽曲を分類することができるが、その構造を正確に機械に再現させることは困難である。その問題に取り組んでいるのが、音楽情報処理である。

近年、音楽情報処理分野に関する研究が活発になってきている。例えば、国際的な認識精度コンテスト MIREX [1] が開かれ、ジャンル推定・アーティスト推定・メロディ推定などの認識精度が競われている。

機械がデータベースの中から自動的に音楽検索・推薦を行うには、人間が行っている音楽の分析手順を機械が知る必要がある。これを直接観察することはできないため、MIREX で設定されているような問題に対して有効な特徴量と機械学習の組み合わせを調べることで音楽情報処理の研究は行われている。

人間の音楽認知活動を機械が理解することによって、楽曲の分類・検索から自動伴奏システム・自動作曲まで幅広い分野での応用が考えられる。

## 2. 楽曲のジャンル推定

現在の楽曲のジャンル推定手法は、audio 形式 (wav ファイルなど) の楽曲から MFCC (Mel-Frequency Cepstrum Coefficients) [1] のように音色に着目した特徴量を抽出し、学習器に判別させるといったものが主流である。特徴量を抽出したり、機械学習を行ったりといった作業を簡単に行うことができる Marsyas [2] や MIRtoolbox [3] などの音楽解析ツールが存在するが、これらは音楽の音響を解析することをメインにしている。Marsyas を用いて ISMIR 2004 のデータセットより各 audio データから特徴量として MFCC のみを SVM に学習させただけで 70% 近くの精度が出た。また、MIREX にはさまざまなタスクが存在するが、いずれも audio データを対象とした認識精度を競うものである。このように audio データから音響的な特徴を抽出することで音楽の認識を行う研究が現在は多いが、音響的な解析によって表現できる特徴はリズム、音色であり、音高に関する特徴に着目した音楽認識の研究は少ない。音楽ではこの音高の情報も非常に重要である。例えば、あるロックミュージシャンの曲がピアノで演奏された音楽データがあるとしよう。このデータから音響的な特徴量を抽出しただけでは、その特徴量からいかなる学習をしようともロックではなくクラシックと判別される可能性が高いように思われる。そこで本研究では、音高の変化に着目した楽曲からの特徴量抽出を行う。

Classification of music score by machine learning-Feature Extraction remarked pitch variation-

<sup>†</sup>Naoto Shirakawa · Graduate School of Engineering, The University of Tokyo

<sup>†</sup>Takashi Chikayama · Graduate School of Engineering, The University of Tokyo

### 3. テンポの変化に強い特徴

本研究では、音高の変化に着目した特徴量を提案する。対象とする問題は、楽曲のジャンル推定問題とする。

ここでテンポの変化に強い特徴量を考える。ある楽曲データがあるとすると、仮にそれが時間軸で伸縮したとする。人間がそれを聞いた場合、ある程度までの伸縮であれば人間は伸び縮みする前の状態と同じ曲だと判断する。そこで楽曲から抽出する特徴量も、楽曲の時間軸に対する伸縮にロバストである必要があると考えた。特徴を抽出する音楽の対象をメロディラインと仮定する。曲中のある時刻  $t$  における音高の高さを  $p(t)$  とした。時刻  $t+d$  での音高は  $p(t+d)$  と表現されるが、音高の変化を表す特徴量は

$$f(t, d) = p(t+d) - p(t) \dots (1)$$

と表現される。ここで音の相対的な変化量

$$f(k, t, d) = \{p(t+kd) - p(t)\} / \{p(t+d) - p(t)\} \dots (2)$$

を考える。これはある時間  $d$  後の音高の変化に比べて、その  $k$  倍の時間  $kd$  が経ったときの変化はどのようなものかということを表す変化量である。

さらにこれをテンポの変更に対して不変にするために、この  $f(k, t, d)$  を  $d$  について積分し、さらにこれを  $t$  で積分する。

これによって以下の特徴量が得られる。

$$f(k) = \sum_{t=0, \infty} [\sum_{d=1, \infty} \{f(k, t, d)\}] \dots (3)$$

さらにこれを正規化すると、

$$F(k) = f(k) / \sum_{k=0, \infty} \{f(k)\} \dots (4)$$

という特徴量が得られる。本研究ではこの特徴量を用いた学習を行う。

### 4. 実験

本研究はまだ実験を行えていないため、今後行う予定の実験とその考察を記述する。

#### 4.1 実験環境

Web 上から無作為に集めた MIDI ファイルを学習データおよび評価データとして扱う。各 MIDI データにはジャンル(クラシックやジャズ、ロックなどを予定)のラベルを付けて実験を行う。

#### 4.2 実験条件

本特徴量が有効性を検証するために、別々に作られた原曲が同じ MIDI データを比較する。また、本特徴量を用いたジャンル推定・クラシック作曲家推定による実験・評価を行う予定である。

### 5. おわりに

本研究では音色に着目した特徴量を提案した。音響的な特徴量のみでは音楽の認識精度に限界が来るのは当然予想される。今後は音響的特徴と音高に関する特徴量を組み合わせた音楽認識を行っていきたい。

#### 参考文献

- [1] MIREX2009, [http://www.music-ir.org/mirex/2009/index.php/Main\\_Page](http://www.music-ir.org/mirex/2009/index.php/Main_Page)
- [2] Logan B.: Mel Frequency Cepstral Coefficients for Music Modeling, Proceedings of the International Symposium on Music Information Retrieval (ISMIR' 00), 2000.
- [3] G. Tzanetakis and P. Cook. Marsyas: A framework for audio analysis. Organized Sound, 4(3):169-175, 2000.
- [4] Lartillot, O. and P. Toiviainen. : A Matlab Toolbox for Musical Feature Extraction from Audio, Proceedings of the International Conference on Digital Audio Effects, 2007.