

# ウェーブレット変換を用いたユーザの嗜好に基づく楽曲推薦システム

(知能情報システム学) 田仲 雅和

## 1. はじめに

近年、大容量携帯音楽プレーヤが普及し、また、インターネットや携帯電話を利用した音楽のダウンロードが急速な広がりを見せている。その結果、大規模楽曲データベースにアクセスできるようになったことで、アーティストやジャンルなどの情報だけでなく、楽曲の音響的特徴を利用した検索や推薦に関する技術が要求されるようになってきている。

そこで本稿では、ウェーブレット変換を利用した新たな楽曲の特徴量を提示し、ユーザの嗜好に基づく楽曲推薦システムを提案する。ユーザによる楽曲評価から得た嗜好ベクトルとウェーブレット変換を利用した特徴量、楽曲のリズム特性を表す特徴量を利用することで未知の楽曲を含む楽曲データベースから、ユーザの嗜好に基づいて楽曲を推薦することができる。

## 2. 特徴量抽出

初めに推薦システムに必要な特徴量を楽曲から抽出する必要がある。本稿では楽曲の特徴量として、ウェーブレット変換を利用した特徴量とリズム特性を表す特徴量を用いた。以下に各特徴量について詳細を示す。

### 2.1. ウェーブレット変換の特徴量

楽曲を特徴付ける特徴のうち音色や音量などを表すために、音声信号をドベシーの離散ウェーブレット変換によって多重解像度解析した各係数の標準偏差を用いた。

まず、窓幅に区切った音声信号  $f(n)$  を、レベル 0 のスケーリング係数  $s_k^{(0)}$  とみなす。次に、レベル  $j$  のスケーリング係数  $s_k^{(j)}$ 、およびウェーブレット係数  $w_k^{(j)}$  を式 (2.1) (2.2) を使用して、 $s_k^{(0)}$  から、逐次、レベル  $J$  まで求める。図 1 にスケーリング係数が徐々に精度の低いスケーリング係数とウェーブレット係数とに分解される様子を示した。

$$s_k^{(j)} = \sum_n p_{n-2k} s_n^{(j-1)} \quad (2.1)$$

$$w_k^{(j)} = \sum_n q_{n-2k} s_n^{(j-1)} \quad (2.2)$$

ここで  $p_k$  はドベシーのスケーリング関数を表す数列であり、 $q_k$  は式 (2.3) から求められる。

$$q_k = (-1)^k p_{1-k} \quad (2.3)$$

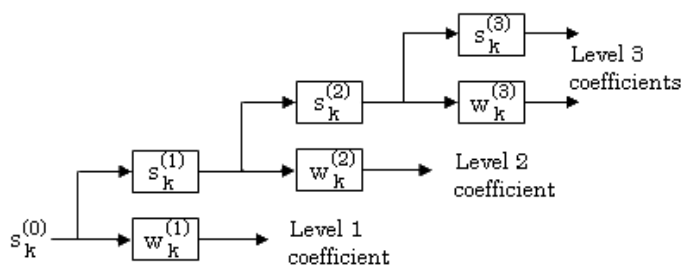


図 1 多重解像度解析

次にレベル 1 から  $J$  の展開係数における標準偏差を求める。ここまでの処理を窓幅ごとにスライドさせながら、楽曲に対して連続して行なう。ここで得られた各展開係数の標準偏差の時系列データを楽曲の特徴量として利用する。

## 2.2. リズム特徴量

楽曲をジャンル分けする論文で提案されたリズム特性を表す特徴量<sup>[1]</sup>を採用した。そして、主のビートの強さ、リズムの規則正しさ、主のビートとサブビートの関係などを特徴ベクトルによって表す。まず、ウェーブレット変換の多重解像度解析を利用して、4個の周波数バンドに分解する。続いて、各周波数バンドに対して、ローパスフィルタなどの前処理を施す。次に、各バンドを逆ウェーブレット変換し、自己相関を計算する。自己相関におけるピークをビートヒストグラムとして音声ファイル全体で蓄積する。図2にロックの楽曲 (RWC-MDB-G-2001 No.7<sup>[2]</sup>) から抽出した 30 s のビートヒストグラムを例示する。ビートヒストグラムでの次の特徴を、リズム特性を表す特徴量とした。

- ・ A0, A1 : 1, 2 番目のピークの相対振幅
- ・ P1, P2 : 1, 2 番目のピークの時間
- ・ RA : 振幅の比率 A1 / A0
- ・ SUM1, SUM2, SUM3 : BPM (Beats Per Minute) が 40~90, 90~140, 140~250 であるヒストグラムの合計

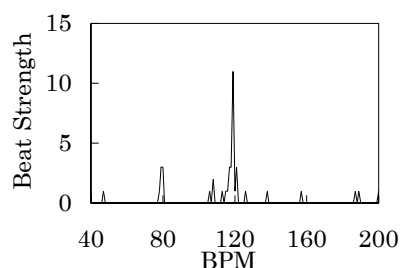


図2 ビートヒストグラム例

## 3. 楽曲推薦システム

本章では2章で述べた特徴量を使った楽曲の推薦方法について説明する。

本稿での楽曲推薦とは、対象ユーザが評価を与えていない楽曲に対して、推薦するかどうかを決定することである。いま、全楽曲のインデックス集合を  $M = \{m | 1, \dots, N_M\}$ 、ユーザが主観評価  $s (1 \leq s \leq 5)$  を与えた楽曲のインデックス集合を  $M_s = \{m_s | 1, \dots, N_{M_s}\} (M_s \subset M)$  とする。ここで  $N_M$  と  $N_{M_s}$  は全楽曲数と評価を与えた楽曲数を表す。

### (1) 一種類の特徴量を用いた方法

全評価楽曲と対象未評価楽曲  $m'$  の合計  $[N_{M_s} + 1]$  曲に対する特徴量を主成分分析し、累積寄与率が 80% 以上になるように  $n$  個の係数を選択し、楽曲  $m$  の特徴ベクトルを  $c_m = (c_{m,1}, \dots, c_{m,n})$  とする。ここで  $m'$  の特徴ベクトル  $c_{m'}$  に対して、ユークリッド距離を用いて、 $M_s$  に対応する特徴ベクトルの集合の各要素との類似度を計算する。計算の結果、類似度が最大となる楽曲を決定し、その評価楽曲のスコア  $s$  を対象未評価楽曲のスコアとする。決定したスコアが 4 または 5 であるとき対象未評価楽曲を推薦する。

### (2) 二種類の特徴量を組み合わせた方法

(1)と同様にして、二種類の特徴量それぞれで未評価楽曲のスコアを決定する。続いて、次のAからCの3種類の手法で推薦するかどうかを決定する。

- A) それぞれの特徴量で求めたスコアが、一方が 4 以上で他方が 3 以上であれば推薦する。
- B) ウェーブレット変換の特徴量で求めたスコアが 4 以上であり、かつウェーブレット変換の特徴量で類似度最大となった楽曲が、リズム特徴量において全評価楽曲の中での類似度の順位が  $[N_{M_s} / 2]$  位以内であった場合推薦する。
- C) ウェーブレット変換の特徴量で求めたスコアが 4 以上であり、かつウェーブレット変換の特徴量で類似度最大となった楽曲が、リズム特徴量において求めた類似度が全評価楽曲の平均より大きければ推薦する。

## 4. 評価実験

### 4.1. 実験条件

今回利用した楽曲は RWC 研究用音楽データベース (音楽ジャンル)<sup>[2]</sup>の 100 曲で、ユーザの音楽嗜好データを収集するため、全ての楽曲について、ユーザ 10 名によって 5 段階の主観評価 (好き : 5 ~ 嫌い : 1) が付与された。

ウェーブレット変換の特徴量を求める際のウェーブレットの窓幅は 524288 サンプル (約 11 秒) とし、楽曲開始から 10s ~ 65s までの信号を利用した。ウェーブレット変換の関数はドベシーの  $N = 2$  とし、レベル 4 で多重解像度解析を行なった。リズム特徴量を求める際は、信号を 22050Hz に変換し、ウェーブレットの窓幅は、65536 サンプル (約 3 秒)、跳躍サイズ (ウェーブレット変換の開始点をずらす幅) を 32768 サンプルとして、楽曲開始から 10s ~ 40s までの信号を利用した。ウェーブレット変換の関数はドベシーの  $N = 4$  とした。

### 4.2. 実験手法

本実験では、3 章記載の楽曲推薦システムの (1) と (2) の A, B, C の 4 種類の方法で、リズム特徴量は予備実験で高い推薦正答率を得ることができた a:[SUM1], b:[A0, A1, SUM1, SUM2, SUM3], c:[A0, A1, P1, P2, RA, SUM1, SUM2, SUM3] の 3 種類の組み合わせで、計 13 種類の推薦正答率を算出し、最良であった推薦方法の結果をシステムの推薦正答率として選択した。また、以下の A, B, C の条件で評価を行った。

条件 A: 評価既知曲 99 曲に対して評価既知曲 1 曲を未知曲として取り扱い、判定する。これを 100 曲に対して行なう。

条件 B: ランダムに  $N$  ( $N = 50$  or  $20$ ) 曲を評価既知曲として選択し、残り評価既知曲  $[100 - N]$  曲を未知曲として取り扱い、1 曲ずつ判定する。ユーザごとに 10 回ずつランダムに  $N$  曲を選択して、実験を実施した。

条件 C: 100 曲の楽曲の内、ポピュラー音楽であるポップス、ダンス、ジャズ、ラテンの 48 曲を対象に、条件 A と同様にして、評価既知曲 47 曲に対して、評価既知曲 1 曲を未知曲として取り扱い、判定する。

### 4.3. 実験結果

各ユーザに対して、条件 A, B, C で、判定が「推薦する」であった曲の内、ユーザの主観評価が 4 または 5 であった曲の割合によってシステムの評価を行なう。

表 1 は推薦正答率とランダムに推薦を行なった場合との差をユーザごとに表示させたものである。対象を全ジャンルの 100 曲とした条件 A, B, C では、ランダムに比べて平均 16.4% 高い正答率で推薦することができた。これは本手法の有効性を示すものである。また、条件 C ではランダムに比べて平均 28.7% 高い正答率で推薦することができた。これは、ジャンルが限定されたことにより、ユーザの嗜好傾向がより安定したことなどが理由であると考えられる。また、全ての条件においてユーザによって大きな差が表れた。これはユーザの音楽嗜好特性の違いによるものであると考えられる。

次に、条件 A で実験を行なった際に最良の結果となった推薦手法と選択した特徴量を表 2 に示した。ユーザにより選択された手法と特徴量は異なり、これもユーザの嗜好特性の違いが表れたものと考えられる。

表1 条件A~Cにおける推薦正答率 (%) とランダムとの差 (%)

| user | 条件 A      |             | 条件 B (N=50) |             | 条件 B (N=20) |             | 条件 C      |             |
|------|-----------|-------------|-------------|-------------|-------------|-------------|-----------|-------------|
|      | 推薦<br>正答率 | ランダム<br>との差 | 推薦<br>正答率   | ランダム<br>との差 | 推薦<br>正答率   | ランダム<br>との差 | 推薦<br>正答率 | ランダム<br>との差 |
| 1    | 51.7      | 11.7        | 49.4        | 9.6         | 46.3        | 5.7         | 76.9      | 37.3        |
| 2    | 37.0      | 8.0         | 48.8        | 19.2        | 46.0        | 17.3        | 66.7      | 31.2        |
| 3    | 50.0      | 16.0        | 43.4        | 9.2         | 42.8        | 9.3         | 68.0      | 24.2        |
| 4    | 47.3      | 5.3         | 49.3        | 10.1        | 47.9        | 5.8         | 68.8      | 10.5        |
| 5    | 58.3      | 29.3        | 46.9        | 16.3        | 33.7        | 4.6         | 50.0      | 12.5        |
| 6    | 42.9      | 15.9        | 50.4        | 24.4        | 36.2        | 11.2        | 33.3      | 12.5        |
| 7    | 56.7      | 22.7        | 48.6        | 16.2        | 48.9        | 15.7        | 53.8      | 33.0        |
| 8    | 62.5      | 27.5        | 53.8        | 20.2        | 48.8        | 14.4        | 73.7      | 40.4        |
| 9    | 50.0      | 10.0        | 47.5        | 10.9        | 50.1        | 10.0        | 43.5      | 3.9         |
| 10   | 40.0      | 30.0        | 52.7        | 43.7        | 50.4        | 41.8        | 100.0     | 81.2        |
| 平均   | 49.6      | 17.6        | 49.1        | 18.0        | 45.1        | 13.6        | 63.5      | 28.7        |

表2 条件Aで選択した手法と特徴量

| user | 手法   | 特徴量      |
|------|------|----------|
| 1    | 手法 A | Rhythm c |
| 2    | 1種類  | Wavelet  |
| 3    | 手法 A | Rhythm b |
| 4    | 手法 A | Rhythm a |
| 5    | 手法 C | Rhythm a |
| 6    | 手法 C | Rhythm a |
| 7    | 手法 C | Rhythm b |
| 8    | 手法 C | Rhythm b |
| 9    | 1種類  | Rhythm c |
| 10   | 手法 B | Rhythm a |

ウェーブレット変換の特徴量とユーザの嗜好の対応を検証するため、ウェーブレット変換特徴量を主成分分析した第1, 第2主成分とユーザの主観評価との対応例を図3, 図4に示した。これらの図よりユーザの評価が高い曲と低い曲が異なる分布となる傾向を見ることができるので、図3, 図4は提案したウェーブレット変換を利用した特徴量の有効性を示唆しているといえる。

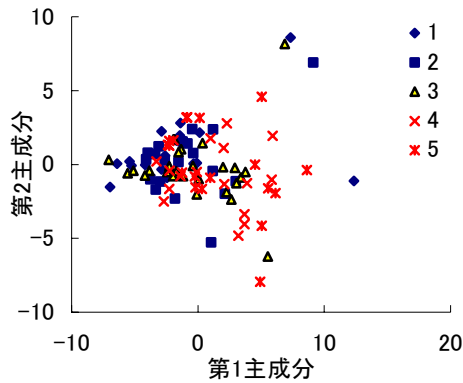


図3 ウェーブレット変換の特徴量と楽曲評価の関係例 (user8: 条件 A)

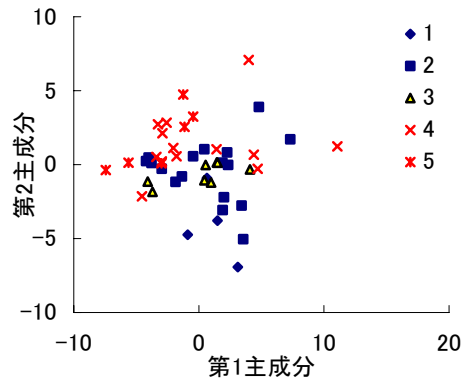


図4 ウェーブレット変換の特徴量と楽曲評価の関係例 (user1: 条件 C)

## 5. まとめ

ウェーブレット変換を利用した新たな楽曲の特徴量を利用した楽曲推薦システムを提案し、実験でその有効性を確認した。今後、ピッチ (メロディー、ベースライン) やコード情報など別の楽曲の特徴量を組み込み、また、推薦順位をつけるなど推薦方法を改善することで、より高い推薦性能を目指していく。また、学習曲が増える場合を想定して、より高い推薦性能を目指し、試聴結果と楽曲の特徴量の関係をニューラルネットワーク、遺伝的アルゴリズムなどで個人ごとに学習する手法を検討している。

## 参考文献

- [1] G. Tzanetakis, P. Cook, "Musical genre classification of audio signals", IEEE Trans. Speech Audio Process, vol.10, pp.293-302, 2002.
- [2] 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一, "RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース", 情報処理学会論文誌, vol.45, no.3, pp.728-738, 2004.