

## 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

A Music Recommendation System based on User Preferences

Using the Discrete Wavelet Transform

田中 雅和†\*

Masakazu Tanaka

吉富 康成†

Yasunari Yoshitomi

## 1. まえがき

近年、大容量携帯音楽プレーヤが普及し、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。そして、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別の検索だけではデータベースの効率的な利用には十分でなく、楽曲推薦が必要になってきた。楽曲推薦を実現する方策として、協調フィルタリングを用いた手法と楽曲の属性に基づく手法が提案されている。

協調フィルタリングを用いた手法では、他のユーザの楽曲評価を利用して対象ユーザに楽曲を推薦している[1]。例えば、対象ユーザが、楽曲A, Bを好きで、他の多くのユーザが楽曲A, B, Cを好きな場合には、対象ユーザに楽曲Cを推薦する。この手法では、どのユーザも評価していない楽曲を推薦対象に含めることができない。

一方、コンテンツベースの手法[2,3,4]では、自動的に楽曲から抽出した音色やリズムなどの情報(特徴量)、あるいはジャンルや歌詞などの楽曲のメタ情報と、個人の嗜好の関係を基に楽曲を推薦する。自動的に楽曲の特徴量を抽出する手法では、どのユーザも評価していない楽曲も推薦対象に含めることができる。近年、楽曲の特徴量の抽出にウェーブレット変換を用いた研究が行われてきた[5,6]。

本研究では、ウェーブレット変換を用いて得られた楽曲の特徴量を利用した、ユーザの嗜好に基づく楽曲推薦システムを提案し、その性能評価を行った。

## 2. 特徴量抽出

初めに推薦に必要な特徴量を楽曲から抽出する必要がある。本報では楽曲の特徴量として、ウェーブレット変換を用いて得られた特徴量とリズム特性を表す特徴量の2種類を用いた。以下に各特徴量について詳細に説明する。

## 2.1 ウェーブレット変換の特徴量

楽曲に離散ウェーブレット変換を施し、その多重解像度表現部のウェーブレット係数のヒストグラムをとると0を中心とした分布となる[7]。著者らは、楽曲及びその開始からの時間帯により、この標準偏差がかなり異なる場合があることを見出した。そこで、楽曲の音色や音量を表す特徴量として、音声信号を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のスケール係数 $s_k^{(0)}$ とみなす。次に、レベル $j$ のスケール係数 $s_k^{(j)}$ 、およびウェーブレット係数 $w_k^{(j)}$ を式(1)、(2)を使用して、 $s_k^{(0)}$ から、逐次、レベル $J$ まで求める。

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

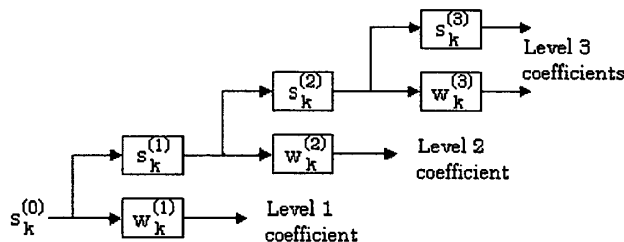


図1 多重解像度解析

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

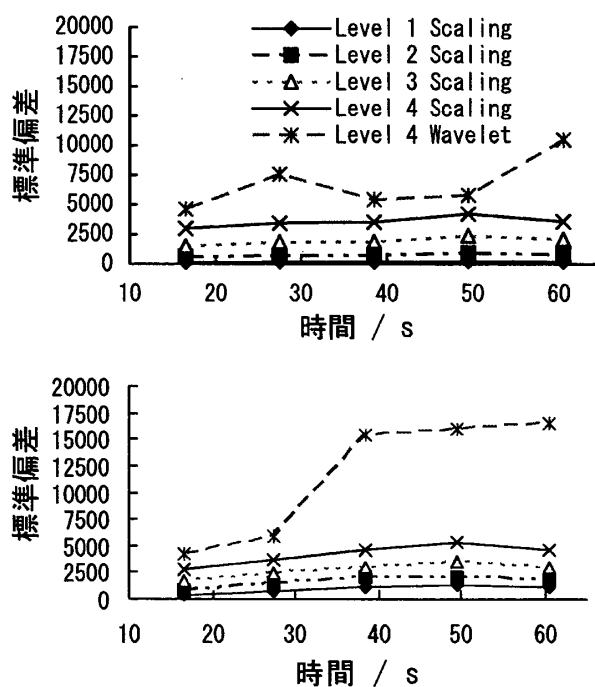
図1にスケール係数が次々にスケール係数とウェーブレット係数とに分解される様子を示した。

本研究では、 $p_k$ としてドベシーのスケール関数を表す数列を用いた。 $q_k$ は式(3)から求められる。

$$q_k = (-1)^k p_{1-k} \quad (3)$$

次にレベル1から $J$ まで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、楽曲に対して順次行う。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ベクトルの各要素とする。

図2はクラシック(RWC-MDB-G-2001 No.49[8])とポップス(RWC-MDB-G-2001 No.1[8])における計算結果例である(計算条件は、4章1節に記載)。

図2 ウェーブレット変換の特徴量の計算結果例  
上図：クラシック, 下図：ポップス

† 京都府立大学大学院, Graduate School of Kyoto Prefectural University

\* 株式会社 NTT データ, NTT DATA Corp. (現在)

## 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[9]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3に30秒間の2ジャンルの音声(クラシック:RWC-MDB-G-2001 No.49,ロック:RWC-MDB-G-2001 No.7)[8]におけるビートヒストグラムを例示する。ビートヒストグラムにおける次の量を、リズム特性を表す特徴量として用いた。

- 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

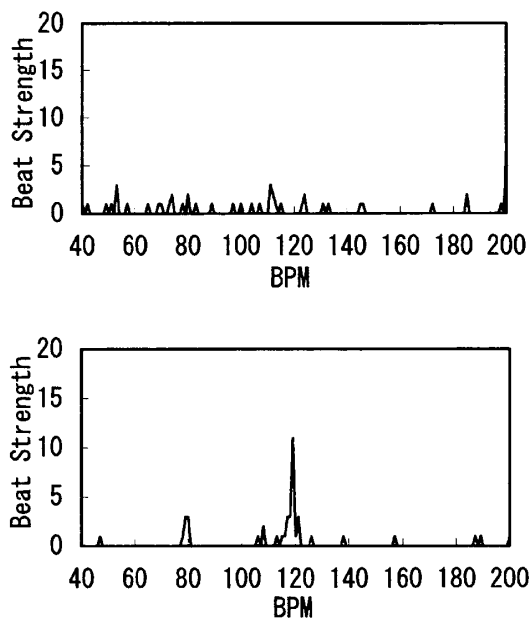


図3 ビートヒストグラム例  
上図:クラシック, 下図:ロック

## 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述べる。本報での楽曲推薦とは、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  $s (1 \leq s \leq 5)$  (好き: 5 ~ 嫌い: 1)を与えた楽曲のインデックス集合を  $M_s = \{m_s | 1, \dots, N_{M_s}\} (M_s \subseteq M)$  とする。ここで  $N_M$  と  $N_{M_s}$  は全楽曲数と評価済み楽曲数を表す。

## 3.1 1種類の特徴量を用いた方法

全評価済み楽曲と対象未評価楽曲  $m'$  の合計  $[N_{M_s} + 1]$  曲の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  $n$  個の成分を選択する。そして、評価済み楽曲  $m$  の特徴ベクトル  $c_m = (c_{m,1}, \dots, c_{m,n})$  と  $m'$  の特徴ベクトル  $c_{m'}$  との類似度(=ユークリッド距離の逆数)を計算する。その結果、類似度が最大となる評価済み楽曲  $m^*$  を  $N_{M_s}$  曲の中から決定し、その評価済み楽曲のスコア  $s$  を  $m'$  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  $m'$  を推薦する。

## 3.2 2種類の特徴量を組み合わせた方法

3.1章と同様にして、2種類の特徴量それぞれで対象未評価楽曲  $m'$  のスコアを決定する。続いて、次のA~Eの5種類の手法で推薦するかどうかを決定する。これらの5つの手法は、「1つの特徴ベクトルでスコアが4以上となり、かつ、もう1つの特徴ベクトルで、平均以上の場合に「推薦」」、という考え方で設定した。「平均」の基準を以下のように3種類(スコア3, 順位平均, 類似度平均)設けた。

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、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済み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、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済み楽曲の平均より大きければ推薦する。

D: リズム特徴量で求めたスコアが4以上で、かつリズム特徴量で類似度最大となった楽曲が、ウェーブレット変換の特徴量において全評価済み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。

E: リズム特徴量で求めたスコアが4以上で、かつリズム特徴量で類似度最大となった楽曲が、ウェーブレット変換の特徴量において求めた類似度が全評価済み楽曲の平均より大きければ推薦する。

## 4. 評価実験

### 4.1 条件

まず、RWC 研究用音楽データベース(音楽ジャンル)[8]の100曲に対して、10名のユーザごとの5段階の主観評価を付与した。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を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。楽曲開始直後は、無音の時間が入る場合があるので、楽曲開始10sから、窓幅の5倍にあたる65sまでサンプリングを行った。上記の窓幅および録音時間については実験的に決定した。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、文献[9]と同様に信号を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 評価方法

3章記載のシステム(3.1章と3.2章のA~Eの方法)で楽曲推薦を行う。リズム特徴量について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種類を用いて、計19通りの推薦正答率を算出し、そのうち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た(実用上は、最良であった方法だけを使うので、本報では、その結果だけを示す)。そして、以下の3つの条件でシステムの評価を行った。

条件1: 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、推薦するかを判定する。これをすべての組み合わせ(100通り)について行う。

条件2: ランダムに N(N=50 or 20)曲を評価既知曲とし、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曲ずつ未知曲を推薦するかを判定する。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3: 楽曲100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1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、推薦するかを判定する。これをすべての組み合わせ(48通り)について行う。

4.3 結果と考察

各ユーザに対して、条件1~3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であった曲の割合である推薦正答率を用いてシステムの評価を行う。

推薦正答率とランダムに推薦を行なった場合との差を表1に示した。対象を全ジャンルの100曲とした条件1,2では、各々ランダムに比べて平均21.0%(条件1),18.4%(条件2,N=50),14.3%(条件2,N=20)だけ高い正答率で推薦できた。条件1,2の結果から、評価既知曲の数が多いほど推薦正答率が高い傾向がみられた。また、条件3ではランダムに比べて平均29.5%だけ高い正答率で推薦することができた。

条件3では、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明確になったために本システムの効果が大きかったものと考えられる。これらのことから、表1は本システムの有効性を示していると考えられる。なお、ユーザの主観評価が4または5であった割合(表2)と推薦正答率(表1)との間に明確な関係はみられなかった。

今回は、すべてのユーザが10~42%の曲に4または5の主観評価を与えた。評価既知曲の中に4または5の主観評価の曲がなければ、未知曲に対して「推薦する」という判定はでない。

次に、各条件で最良の結果となった推薦手法と選択したRhythm特徴量を表3に示した。1種類の特徴量を用いた方法が選択されたのは2例のみであったことから、2種類の特徴量を組み合わせることが有効であることがわかった。選択された手法とRhythm特徴量はユーザおよび条件により異なった。

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の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4に示した。図4よりユーザの評価が高い曲と低い曲が異なる

分布となる傾向をみることで、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考えられる。同様に、リズム特性を表す特徴量を、第1,第2主成分で2次的に図示したところ、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がある程度みられた(図5)。

本システムを実際に用いる際には、評価既知曲のうち1曲を未知曲として取り扱い、条件1と同様にして、各ユーザにとって最良の推薦方法を、本システムが自動的に19種類の方法の中から1つ選択する。そして、その方法で実際の未知楽曲を推薦するかを判定し、推薦した場合にはその楽曲の評価をユーザにってもらうことで評価既知曲を1曲増やし、本システムが最良の推薦方法を自動的に再選択する。このようにして、推薦曲の評価をシステムの性能改善に逐次利用することができる。

表1 条件1~3における推薦正答率(%)とランダムとの差(%)

user	条件1		条件2(N=50)	
	推薦正答率	ランダムとの差	推薦正答率	ランダムとの差
1	55.6	15.6	49.7	9.9
2	37.5	8.5	48.8	19.2
3	64.3	30.3	43.4	9.2
4	55.2	13.2	49.3	10.1
5	58.3	29.3	46.9	16.3
6	42.9	15.9	50.4	24.4
7	56.7	22.7	48.6	16.2
8	62.5	27.5	53.8	20.2
9	54.2	14.2	51.4	14.8
10	42.9	32.9	52.7	43.7
平均	53.0	21.0	49.5	18.4

user	条件2(N=20)		条件3	
	推薦正答率	ランダムとの差	推薦正答率	ランダムとの差
1	49.6	9.0	76.9	37.3
2	46.0	17.3	66.7	31.2
3	44.1	10.6	68.0	24.2
4	47.9	5.8	68.8	10.5
5	33.7	4.6	50.0	12.5
6	37.2	12.2	37.2	16.4
7	50.1	16.9	53.8	33.0
8	48.8	14.4	73.7	40.4
9	50.1	10.0	47.5	7.9
10	50.4	41.8	100.0	81.2
平均	45.8	14.3	64.3	29.5

表2 ユーザの主観評価が4または5であった曲の割合

user	1	2	3	4	5	6	7	8	9	10
評価4,5の割合(%)	40	29	34	42	29	27	34	35	40	10

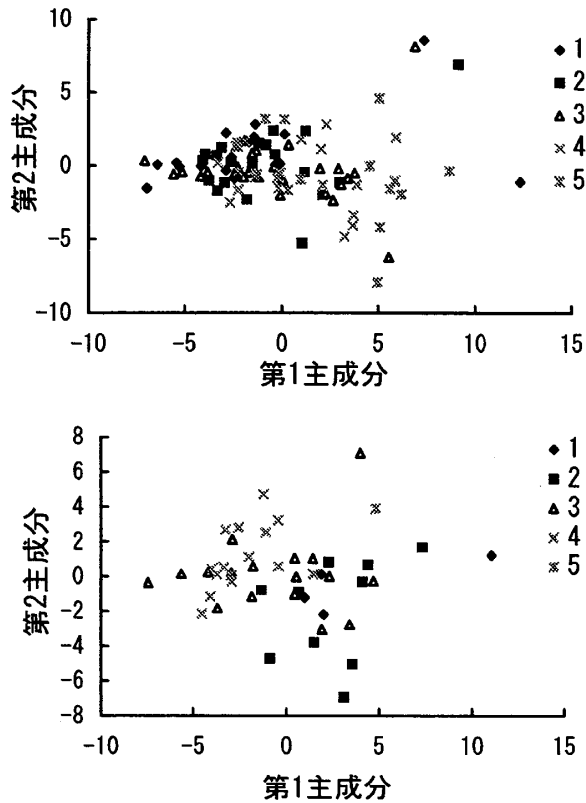


図4 ウェーブレット変換の特徴量と楽曲評価の関係例  
上図: user8, 条件1, 下図: user2, 条件3

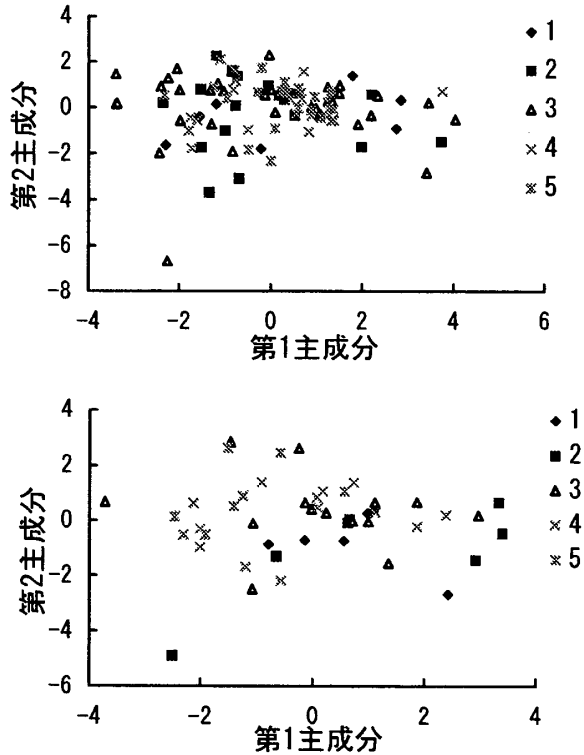


図5 リズム特性を表す特徴量(Rhythm c)と楽曲評価の関係例; 上図: user9, 条件1, 下図: user3, 条件3

表3 各条件で選択した手法と Rhythm 特徴量

user	条件1		条件2				条件3	
	手法	Rhythm 特徴量	N=50		N=20		手法	Rhythm 特徴量
			手法	Rhythm 特徴量	手法	Rhythm 特徴量		
1	D	c	E	b	E	a	C	b
2	E	c	B	b	C	b	C	c
3	D	b	B	a	D	b	A	c
4	D	a	B	a	C	a	B	a
5	C	a	C	a	B	c	A	a
6	C	a	C	a	D	a	D	a
7	C	b	C	b	D	c	1種類	c
8	D	a	B	b	B	b	C	b
9	D	a	E	a	1種類	b	D	b
10	D,E	b	B	a	B	c	B	a

5. まとめ

本報ではウェーブレット変換を利用して抽出した楽曲の特徴量を用いた楽曲推薦システムを提案し、その有効性を示した。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、より高性能な推薦システムを目指していく。また、評価既知曲が大量となる場合に備えて、より高い推薦正答率を目指して、ユーザーごとに試聴結果と楽曲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学習する手法を検討している。

参考文献

- [1] J. Breese, D. Heckerman, C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering". Proc. UBI, pp. 43-52, 1998.
- [2] B. Logan, "Music recommendation from song sets", Proc. ISMIR, pp.425-428, 2004.
- [3] 帆足 啓一郎, 井ノ上 直己, "ユーザの音楽嗜好に基づく音楽情報検索手法", 情報処理学会研究報告, MUS[音楽情報科学研究会報告], vol.49, pp.79-84, 2003.
- [4] 奥乃 博, 北原 鉄朗, 吉井 和佳, "楽曲の特徴量抽出と検索技術", 電気学会誌, vol.127, no.7, pp.417-420, 2007.
- [5] G. Li, A. Khokhar, "Content-based indexing and retrieval of audio data using wavelets". Proc. of ICME 2000, vol.2, pp. 885-888, 2000.
- [6] T. Li, M. Ogiwara, Q. Li, "A comparative study on content-based music genre classification". Proc of ACM SIGIR 2003, pp.282-289, 2003.
- [7] 村田 真一, 吉富 康成, 石井 博昭, "ウェーブレット変換を用いた音声電子透かし埋め込み位置の最適化", 日本オペレーションズ・リサーチ学会秋季研究発表会アブストラクト集, pp.210-211, 2007.
- [8] 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.
- [9] G. Tzanetakis, P. Cook, "Musical genre classification of audio signals", IEEE Trans. Speech Audio Process, vol.10, no.5, pp.293-302, 2002.